The suggested citation for this software is
StataCorp. 2013. *Stata: Release 13*. Statistical Software. College Station, TX: StataCorp LP.
Contents

intro .............................................. Introduction to longitudinal-data/panel-data manual 1
xt .......................................................... Introduction to xt commands 2
quadchk .............................................. Check sensitivity of quadrature approximation 10
vce_options ........................................ Variance estimators 20
xtabond ............................................. Arellano–Bond linear dynamic panel-data estimation 24
xtabond postestimation ................................. Postestimation tools for xtabond 42
xtcloglog ........................................... Random-effects and population-averaged cloglog models 45
xtcloglog postestimation ............................... Postestimation tools for xtcloglog 60
xtdata .................................................... Faster specification searches with xt data 63
xtdescribe ............................................... Describe pattern of xt data 70
xtdpd ..................................................... Linear dynamic panel-data estimation 74
xtdpd postestimation ................................ Postestimation tools for xtdpd 93
xtdpdsys .......... Arellano–Bover/Blundell–Bond linear dynamic panel-data estimation 98
xtdpdsys postestimation ............................... Postestimation tools for xtdpdsys 108
xtfrontier ............................................... Stochastic frontier models for panel data 112
xtfrontier postestimation ............................... Postestimation tools for xtfrontier 124
xtgee ..................................................... Fit population-averaged panel-data models by using GEE 127
xtgee postestimation ................................ Postestimation tools for xtgee 146
xtgls ..................................................... Fit panel-data models by using GLS 155
xtgls postestimation ................................ Postestimation tools for xtgls 165
xthtaylor .............................................. Hausman–Taylor estimator for error-components models 167
xthtaylor postestimation ............................... Postestimation tools for xthtaylor 180
xtintreg ............................................... Random-effects interval-data regression models 184
xtintreg postestimation ............................... Postestimation tools for xtintreg 193
xtivreg ................................................ Instrumental variables and two-stage least squares for panel-data models 197
xtivreg postestimation ............................... Postestimation tools for xtivreg 219
xtline .................................................... Panel-data line plots 221
xtlogit ................................................. Fixed-effects, random-effects, and population-averaged logit models 225
xtlogit postestimation ............................... Postestimation tools for xtlogit 243
xtnbreg . . . . . Fixed-effects, random-effects, & population-averaged negative binomial models 247
<table>
<thead>
<tr>
<th>Command</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>xtbreg</code></td>
<td>Postestimation tools for <code>xtbreg</code></td>
<td>260</td>
</tr>
<tr>
<td><code>xtologit</code></td>
<td>Random-effects ordered logistic models</td>
<td>263</td>
</tr>
<tr>
<td><code>xtologit</code></td>
<td>Postestimation tools for <code>xtologit</code></td>
<td>272</td>
</tr>
<tr>
<td><code>xtoprob</code></td>
<td>Random-effects ordered probit models</td>
<td>275</td>
</tr>
<tr>
<td><code>xtoprob</code></td>
<td>Postestimation tools for <code>xtoprob</code></td>
<td>284</td>
</tr>
<tr>
<td><code>xtcse</code></td>
<td>Linear regression with panel-corrected standard errors</td>
<td>287</td>
</tr>
<tr>
<td><code>xtcse</code></td>
<td>Postestimation tools for <code>xtcse</code></td>
<td>298</td>
</tr>
<tr>
<td><code>xtpoisson</code></td>
<td>Fixed-effects, random-effects, and population-averaged Poisson models</td>
<td>300</td>
</tr>
<tr>
<td><code>xtpoisson</code></td>
<td>Postestimation tools for <code>xtpoisson</code></td>
<td>323</td>
</tr>
<tr>
<td><code>xtprobit</code></td>
<td>Random-effects and population-averaged probit models</td>
<td>327</td>
</tr>
<tr>
<td><code>xtprobit</code></td>
<td>Postestimation tools for <code>xtprobit</code></td>
<td>347</td>
</tr>
<tr>
<td><code>xtc</code></td>
<td>Random-coefficients model</td>
<td>350</td>
</tr>
<tr>
<td><code>xtc</code></td>
<td>Postestimation tools for <code>xtc</code></td>
<td>357</td>
</tr>
<tr>
<td><code>xtreg</code></td>
<td>Fixed-, between-, and random-effects and population-averaged linear models</td>
<td>359</td>
</tr>
<tr>
<td><code>xtreg</code></td>
<td>Postestimation tools for <code>xtreg</code></td>
<td>388</td>
</tr>
<tr>
<td><code>xtregar</code></td>
<td>Fixed- and random-effects linear models with an AR(1) disturbance</td>
<td>395</td>
</tr>
<tr>
<td><code>xtregar</code></td>
<td>Postestimation tools for <code>xtregar</code></td>
<td>410</td>
</tr>
<tr>
<td><code>xtset</code></td>
<td>Declare data to be panel data</td>
<td>412</td>
</tr>
<tr>
<td><code>xtsum</code></td>
<td>Summarize xt data</td>
<td>425</td>
</tr>
<tr>
<td><code>xttab</code></td>
<td>Tabulate xt data</td>
<td>427</td>
</tr>
<tr>
<td><code>xttobit</code></td>
<td>Random-effects tobit models</td>
<td>430</td>
</tr>
<tr>
<td><code>xttobit</code></td>
<td>Postestimation tools for <code>xttobit</code></td>
<td>438</td>
</tr>
<tr>
<td><code>xtunitroot</code></td>
<td>Panel-data unit-root tests</td>
<td>440</td>
</tr>
<tr>
<td><strong>Glossary</strong></td>
<td></td>
<td>471</td>
</tr>
<tr>
<td><strong>Subject and author index</strong></td>
<td></td>
<td>475</td>
</tr>
</tbody>
</table>
Cross-referencing the documentation

When reading this manual, you will find references to other Stata manuals. For example,

[U] 26 Overview of Stata estimation commands
[R] regress
[D] reshape

The first example is a reference to chapter 26, Overview of Stata estimation commands, in the User’s Guide; the second is a reference to the regress entry in the Base Reference Manual; and the third is a reference to the reshape entry in the Data Management Reference Manual.

All the manuals in the Stata Documentation have a shorthand notation:

[GSM] Getting Started with Stata for Mac
[GSU] Getting Started with Stata for Unix
[GSW] Getting Started with Stata for Windows
[U] Stata User’s Guide
[R] Stata Base Reference Manual
[XT] Stata Longitudinal-Data/Panel-Data Reference Manual
[M] Stata Multiple-Imputation Reference Manual
[PSS] Stata Power and Sample-Size Reference Manual
[SVY] Stata Survey Data Reference Manual
[I] Stata Glossary and Index
This entry describes this manual and what has changed since Stata 12.

Following this entry, [XT] xt provides an overview of the xt commands. The other parts of this manual are arranged alphabetically. If you are new to Stata’s xt commands, we recommend that you read the following sections first:

[XT] xt Introduction to xt commands
[XT] xtset Declare a dataset to be panel data
[XT] xtreg Fixed-, between-, and random-effects, and population-averaged linear models

Stata is continually being updated, and Stata users are always writing new commands. To find out about the latest cross-sectional time-series features, type `search panel data` after installing the latest official updates; see [R] update.

For a complete list of all the new features in Stata 13, see [U] 1.3 What’s new.

Also see

[U] 1.3 What’s new
[R] intro — Introduction to base reference manual
The xt series of commands provides tools for analyzing panel data (also known as longitudinal data or in some disciplines as cross-sectional time series when there is an explicit time component). Panel datasets have the form $x_{it}$, where $x_{it}$ is a vector of observations for unit $i$ and time $t$. The particular commands (such as xtdescribe, xtsum, and xtreg) are documented in alphabetical order in the entries that follow this entry. If you do not know the name of the command you need, try browsing the second part of this description section, which organizes the xt commands by topic. The next section, Remarks and examples, describes concepts that are common across commands.

The xtset command sets the panel variable and the time variable; see [XT] xtset. Most xt commands require that the panel variable be specified, and some require that the time variable also be specified. Once you xtset your data, you need not do it again. The xtset information is stored with your data.

If you have previously tsset your data by using both a panel and a time variable, these settings will be recognized by xtset, and you need not xtset your data.


Setup

xtset Declare data to be panel data

Data management and exploration tools

xtdescribe Describe pattern of xt data
xtsum Summarize xt data
xttab Tabulate xt data
xtdata Faster specification searches with xt data
xtline Panel-data line plots
Introduction to xt commands

Linear regression estimators

- `xtreg`  Fixed-, between-, and random-effects, and population-averaged linear models
- `xtregar`  Fixed- and random-effects linear models with an AR(1) disturbance
- `xtgls`  Panel-data models by using GLS
- `xtpcs`  Linear regression with panel-corrected standard errors
- `xthtaylor`  Hausman–Taylor estimator for error-components models
- `xtfrontier`  Stochastic frontier models for panel data
- `xtc`  Random-coefficients regression
- `xtivreg`  Instrumental variables and two-stage least squares for panel-data models

Unit-root tests

- `xtunitroot`  Panel-data unit-root tests

Dynamic panel-data estimators

- `xtabond`  Arellano–Bond linear dynamic panel-data estimation
- `xtdp`  Linear dynamic panel-data estimation
- `xtdpdsys`  Arellano–Bover/Blundell–Bond linear dynamic panel-data estimation

Censored-outcome estimators

- `xttobit`  Random-effects tobit models
- `xtintreg`  Random-effects interval-data regression models

Binary-outcome estimators

- `xtlogit`  Fixed-effects, random-effects, and population-averaged logit models
- `xtprobit`  Random-effects and population-averaged probit models
- `xtcloglog`  Random-effects and population-averaged cloglog models

Ordinal-outcome estimators

- `xtologit`  Random-effects ordered logistic models
- `xtoprobit`  Random-effects ordered probit models

Count-data estimators

- `xtpoisson`  Fixed-effects, random-effects, and population-averaged Poisson models
- `xtnbreg`  Fixed-effects, random-effects, & population-averaged negative binomial models

Generalized estimating equations estimator

- `xtgee`  Population-averaged panel-data models by using GEE

Utility

- `quadchk`  Check sensitivity of quadrature approximation
## Remarks and examples

Consider having data on $n$ units—individuals, firms, countries, or whatever—over $T$ periods. The data might be income and other characteristics of $n$ persons surveyed each of $T$ years, the output and costs of $n$ firms collected over $T$ months, or the health and behavioral characteristics of $n$ patients collected over $T$ years. In panel datasets, we write $x_{it}$ for the value of $x$ for unit $i$ at time $t$. The xt commands assume that such datasets are stored as a sequence of observations on $(i, t, x)$.

For a discussion of panel-data models, see Baltagi (2013), Greene (2012, chap. 11), Hsiao (2003), and Wooldridge (2010). Cameron and Trivedi (2010) illustrate many of Stata’s panel-data estimators.

### Example 1

If we had data on pulmonary function (measured by forced expiratory volume, or FEV) along with smoking behavior, age, sex, and height, a piece of the data might be

```
. list in 1/6, separator(0) divider

    +----------+-----------------+-------+-------+-------+---------+---------+---------+
    | pid | yr_visit | fev   | age  | sex | height | smokes |
    +----------+-----------------+-------+-------+-------+---------+---------+---------+
    | 1.  | 1071   | 1991  | 1.21 | 25  | 1      | 69      | 0       |
    | 2.  | 1071   | 1992  | 1.52 | 26  | 1      | 69      | 0       |
    | 3.  | 1071   | 1993  | 1.32 | 28  | 1      | 68      | 0       |
    | 4.  | 1072   | 1991  | 1.33 | 18  | 1      | 71      | 1       |
    | 5.  | 1072   | 1992  | 1.18 | 20  | 1      | 71      | 1       |
    | 6.  | 1072   | 1993  | 1.19 | 21  | 1      | 71      | 0       |
    +----------+-----------------+-------+-------+-------+---------+---------+---------+
```

The xt commands need to know the identity of the variable identifying patient, and some of the xt commands also need to know the identity of the variable identifying time. With these data, we would type

```
. xtset pid yr_visit
```

If we resaved the data, we need not respecify `xtset`.

### Technical note

Panel data stored as shown above are said to be in the long form. Perhaps the data are in the wide form with 1 observation per unit and multiple variables for the value in each year. For instance, a piece of the pulmonary function data might be

```
    +----------+-------+-------+-------+-------+-------+-------+-------+-------+-------+-------+
    | pid | sex | fev91 | fev92 | fev93 | age91 | age92 | age93 |
    +----------+-------+-------+-------+-------+-------+-------+-------+-------+-------+-------+
    | 1071  | 1    | 1.21  | 1.52  | 1.32  | 25    | 26    | 28    |
    | 1072  | 1    | 1.33  | 1.18  | 1.19  | 18    | 20    | 21    |
    +----------+-------+-------+-------+-------+-------+-------+-------+-------+-------+-------+
```

Data in this form can be converted to the long form by using `reshape`; see `[D] reshape`.

### Example 2

Data for some of the periods might be missing. That is, we have panel data on $i = 1, \ldots, n$ and $t = 1, \ldots, T$, but only $T_i$ of those observations are defined. With such missing periods—called unbalanced data—a piece of our pulmonary function data might be

```
    +----------+-------+-------+-------+-------+-------+-------+-------+-------+-------+
    | pid | sex | fev91 | fev92 | fev93 | age91 | age92 | age93 |
    +----------+-------+-------+-------+-------+-------+-------+-------+-------+-------+
    | 1071  | 1    | 1.21  | 1.52  | 1.32  | 25    | 26    | 28    |
    | 1072  | 1    | 1.33  | 1.18  | 1.19  | 18    | 20    | 21    |
    +----------+-------+-------+-------+-------+-------+-------+-------+-------+-------+
```
. list in 1/6, separator(0) divider

<table>
<thead>
<tr>
<th>pid</th>
<th>yr_visit</th>
<th>fev</th>
<th>age</th>
<th>sex</th>
<th>height</th>
<th>smokes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>1071</td>
<td>1991</td>
<td>1.21</td>
<td>25</td>
<td>69</td>
<td>0</td>
</tr>
<tr>
<td>2.</td>
<td>1071</td>
<td>1992</td>
<td>1.52</td>
<td>26</td>
<td>69</td>
<td>0</td>
</tr>
<tr>
<td>3.</td>
<td>1071</td>
<td>1993</td>
<td>1.32</td>
<td>28</td>
<td>68</td>
<td>0</td>
</tr>
<tr>
<td>4.</td>
<td>1072</td>
<td>1991</td>
<td>1.33</td>
<td>18</td>
<td>71</td>
<td>1</td>
</tr>
<tr>
<td>5.</td>
<td>1072</td>
<td>1993</td>
<td>1.19</td>
<td>21</td>
<td>71</td>
<td>0</td>
</tr>
<tr>
<td>6.</td>
<td>1073</td>
<td>1991</td>
<td>1.47</td>
<td>24</td>
<td>64</td>
<td>0</td>
</tr>
</tbody>
</table>

Patient ID 1072 is not observed in 1992. The xt commands are robust to this problem.

Technical note

In many of the entries in [XT], we will use data from a subsample of the NLSY data (Center for Human Resource Research 1989) on young women aged 14–26 years in 1968. Women were surveyed in each of the 21 years 1968–1988, except for the six years 1974, 1976, 1979, 1981, 1984, and 1986. We use two different subsets: nlswork.dta and union.dta.

For nlswork.dta, our subsample is of 4,711 women in years when employed, not enrolled in school and evidently having completed their education, and with wages in excess of $1/hour but less than $700/hour.
. use http://www.stata-press.com/data/r13/nlswork
(National Longitudinal Survey. Young Women 14-26 years of age in 1968)
. describe

Contains data from http://www.stata-press.com/data/r13/nlswork.dta

| obs: 28,534 | National Longitudinal Survey. Young Women 14-26 years of age in 1968 |
| size: 941,622 |

<table>
<thead>
<tr>
<th>variable name</th>
<th>type</th>
<th>format</th>
<th>label</th>
</tr>
</thead>
<tbody>
<tr>
<td>idcode</td>
<td>int</td>
<td>%8.0g</td>
<td>NLS ID</td>
</tr>
<tr>
<td>year</td>
<td>byte</td>
<td>%8.0g</td>
<td>interview year</td>
</tr>
<tr>
<td>birth_yr</td>
<td>byte</td>
<td>%8.0g</td>
<td>birth year</td>
</tr>
<tr>
<td>age</td>
<td>byte</td>
<td>%8.0g</td>
<td>age in current year</td>
</tr>
<tr>
<td>race</td>
<td>byte</td>
<td>%8.0g</td>
<td>racelbl race</td>
</tr>
<tr>
<td>msp</td>
<td>byte</td>
<td>%8.0g</td>
<td>1 if married, spouse present</td>
</tr>
<tr>
<td>nev_mar</td>
<td>byte</td>
<td>%8.0g</td>
<td>1 if never married</td>
</tr>
<tr>
<td>grade</td>
<td>byte</td>
<td>%8.0g</td>
<td>current grade completed</td>
</tr>
<tr>
<td>collgrad</td>
<td>byte</td>
<td>%8.0g</td>
<td>1 if college graduate</td>
</tr>
<tr>
<td>not_smsa</td>
<td>byte</td>
<td>%8.0g</td>
<td>1 if not SMSA</td>
</tr>
<tr>
<td>c_city</td>
<td>byte</td>
<td>%8.0g</td>
<td>1 if central city</td>
</tr>
<tr>
<td>south</td>
<td>byte</td>
<td>%8.0g</td>
<td>1 if south</td>
</tr>
<tr>
<td>ind_code</td>
<td>byte</td>
<td>%8.0g</td>
<td>industry of employment</td>
</tr>
<tr>
<td>occ_code</td>
<td>byte</td>
<td>%8.0g</td>
<td>occupation</td>
</tr>
<tr>
<td>union</td>
<td>byte</td>
<td>%8.0g</td>
<td>1 if union</td>
</tr>
<tr>
<td>wks_ue</td>
<td>byte</td>
<td>%8.0g</td>
<td>weeks unemployed last year</td>
</tr>
<tr>
<td>ttl_exp</td>
<td>float</td>
<td>%9.0g</td>
<td>total work experience</td>
</tr>
<tr>
<td>tenure</td>
<td>float</td>
<td>%9.0g</td>
<td>job tenure, in years</td>
</tr>
<tr>
<td>hours</td>
<td>int</td>
<td>%8.0g</td>
<td>usual hours worked</td>
</tr>
<tr>
<td>wks_work</td>
<td>int</td>
<td>%8.0g</td>
<td>weeks worked last year</td>
</tr>
<tr>
<td>ln_wage</td>
<td>float</td>
<td>%9.0g</td>
<td>ln(wage/GNP deflator)</td>
</tr>
</tbody>
</table>

Sorted by: idcode year
Many of the variables in the nlswork dataset are indicator variables, so we have used factor variables (see [U] 11.4.3 Factor variables) in many of the examples in this manual. You will see terms like c.age#c.age or 2.race in estimation commands. c.age#c.age is just age interacted with age, or age-squared, and 2.race is just an indicator variable for black (race = 2).

Instead of using factor variables, you could type

```
. generate age2 = age*age
. generate black = (race==2)
```

and substitute age2 and black in your estimation command for c.age#c.age and 2.race, respectively.

There are advantages, however, to using factor variables. First, you do not actually have to create new variables, so the number of variables in your dataset is less.

Second, by using factor variables, we are able to take better advantage of postestimation commands. For example, if we specify the simple model

```
. xtreg ln_wage age age2, fe
```

then age and age2 are completely separate variables. Stata has no idea that they are related—that one is the square of the other. Consequently, if we compute the average marginal effect of age on the log of wages,

```
. margins, dydx(age)
```

then the reported marginal effect is with respect to the age variable alone and not with respect to the true effect of age, which involves the coefficients on both age and age2.

If instead we fit our model using an interaction of age with itself for the square of age,

```
. xtreg ln_wage age c.age#c.age, fe
```
then Stata has a deep understanding that the coefficients `age` and `c.age#c.age` are related. After fitting this model, the marginal effect reported by `margins` includes the full effect of age on the log of income, including the contribution of both coefficients.

`. margins, dydx(age)`

There are other reasons for preferring factor variables; see [R] `margins` for examples.

For `union.dta`, our subset was sampled only from those with union membership information from 1970 to 1988. Our subsample is of 4,434 women. The important variables are `age` (16–46), `grade` (years of schooling completed, ranging from 0 to 18), `not_smsa` (28% of the person-time was spent living outside a standard metropolitan statistical area (SMSA)), and `south` (41% of the person-time was in the South). The dataset also has variable `union`. Overall, 22% of the person-time is marked as time under union membership, and 44% of these women have belonged to a union.

`. use http://www.stata-press.com/data/r13/union
(NLS Women 14–24 in 1968)`

`. describe
Contains data from http://www.stata-press.com/data/r13/union.dta
obs: 26,200
vars: 8
size: 235,800

variable name storage display value variable label
--- ----- ---------- ------- -------
idcode int %8.0g NLS ID
year byte %8.0g interview year
age byte %8.0g age in current year
grade byte %8.0g current grade completed
not_smsa byte %8.0g 1 if not SMSA
south byte %8.0g 1 if south
union byte %8.0g 1 if union
black byte %8.0g race black

Sorted by: idcode year

`. summarize
Variable | Obs     Mean   Std. Dev. | Min  | Max
---------|----------|----------|------|------
idcode | 26200   2611.582 1484.994 | 1    | 5159
year    | 26200   79.47137 5.965499 | 70   | 88
age     | 26200   30.43221 6.489056 | 16   | 46
grade   | 26200   12.76145 2.411715 | 0    | 18
not_smsa| 26200   .2837023 .4508027 | 0    | 1
south   | 26200   .4130153 .4923849 | 0    | 1
union   | 26200   .2217939 .4154611 | 0    | 1
black   | 26200   .274542  .4462917 | 0    | 1

In many of the examples where the `union` dataset is used, we also include an interaction between the `year` variable and the `south` variable—`south#c.year`. This interaction is created using factor-variables notation; see [U] 11.4.3 Factor variables.

With both datasets, we have typed

`. xtset idcode year

8 xt — Introduction to xt commands
Technical note

The `xtset` command sets the $t$ and $i$ index for xt data by declaring them as characteristics of the data; see [P] char. The panel variable is stored in `_dta[iis]` and the time variable is stored in `_dta[tis]`.

Technical note

Throughout the entries in [XT], when random-effects models are fit, a likelihood-ratio test that the variance of the random effects is zero is included. These tests occur on the boundary of the parameter space, invalidating the usual theory associated with such tests. However, these likelihood-ratio tests have been modified to be valid on the boundary. In particular, the null distribution of the likelihood-ratio test statistic is not the usual $\chi^2_1$ but is rather a 50:50 mixture of a $\chi^2_0$ (point mass at zero) and a $\chi^2_1$, denoted as $\chi^2_{01}$. See Gutierrez, Carter, and Drukker (2001) for a full discussion.

References


Also see

[XT] `xtset` — Declare data to be panel data
quadchk — Check sensitivity of quadrature approximation

Syntax

    quadchk [ #1 #2 ] [ , nooutput nofrom ]

Menu

    Statistics > Longitudinal/panel data > Setup and utilities > Check sensitivity of quadrature approximation

Description

    quadchk checks the quadrature approximation used in the random-effects estimators of the following commands:
    xtcloglog
    xtitreg
    xtlogit
    xtologit
    xtoprobit
    xtpoisson, re with the normal option
    xtpoisson
    xttobit

    quadchk refits the model for different numbers of quadrature points and then compares the different solutions.

    #1 and #2 specify the number of quadrature points to use in the comparison runs of the previous model. The default is to use (roughly) $2n_q/3$ and $4n_q/3$ points, where $n_q$ is the number of quadrature points used in the original estimation.

    Most options supplied to the original model are respected by quadchk, but some are not. These are or, vce(), and the maximize_options.

Options

    nooutput suppresses the iteration log and output of the refitted models.

    nofrom forces the refitted models to start from scratch rather than starting from the previous estimation results. Specifying the nofrom option can level the playing field in testing estimation results.

Remarks and examples

    Remarks are presented under the following headings:
    
    What makes a good random-effects model fit?
    How do I know whether I have a good quadrature approximation?
    What can I do to improve my results?
What makes a good random-effects model fit?

Some random-effects estimators in Stata use adaptive or nonadaptive Gauss–Hermite quadrature to compute the log likelihood and its derivatives. As a rule, adaptive quadrature, which is the default integration method, is much more accurate. The quadchk command provides a means to look at the numerical accuracy of either quadrature approximation. A good random-effects model fit depends on both the goodness of the quadrature approximation and the goodness of the data.

The accuracy of the quadrature approximation depends on three factors. The first and second are how many quadrature points are used and where the quadrature points fall. These two factors directly influence the accuracy of the quadrature approximation. The number of quadrature points may be specified with the \texttt{intpoints()} option. However, once the number of points is specified, their abscissas (locations) and corresponding weights are completely determined. Increasing the number of points expands the range of the abscissas and, to a lesser extent, increases the density of the abscissas. For this reason, a function that undulates between the abscissas can be difficult to approximate.

Third, the smoothness of the function being approximated influences the accuracy of the quadrature approximation. Gauss–Hermite quadrature estimates integrals of the type

$$\int_{-\infty}^{\infty} e^{-x^2} f(x) \, dx$$

and the approximation is exact if \(f(x)\) is a polynomial of degree less than the number of integration points. Therefore, \(f(x)\) that are well approximated by polynomials of a given degree have integrals that are well approximated by Gauss–Hermite quadrature with that given number of integration points. Both large panel sizes and high \(\rho\) can reduce the accuracy of the quadrature approximation.

A final factor affects the goodness of the random-effects model: the data themselves. For high \(\rho\), for example, there is high intrapanel correlation, and panels look like observations. The model becomes unidentified. Here, even with exact quadrature, fitting the model would be difficult.

How do I know whether I have a good quadrature approximation?

\texttt{quadchk} is intended as a tool to help you know whether you have a good quadrature approximation. As a rule of thumb, if the coefficients do not change by more than a relative difference of \(10^{-4}\) (0.01%), the choice of quadrature points does not significantly affect the outcome, and the results may be confidently interpreted. However, if the results do change appreciably—greater than a relative difference of \(10^{-2}\) (1%)—then quadrature is not reliably approximating the likelihood.

What can I do to improve my results?

If the \texttt{quadchk} command indicates that the estimation results are sensitive to the number of quadrature points, there are several things you can do. First, if you are not using adaptive quadrature, switch to adaptive quadrature.

Adaptive quadrature can improve the approximation by transforming the integrand so that the abscissas and weights sample the function on a more suitable range. Details of this transformation are in \textit{Methods and formulas} for the given commands; for example, see \texttt{[XT] xtprobit}.

If the model still shows sensitivity to the number of quadrature points, increase the number of quadrature points with the \texttt{intpoints()} option. This option will increase the range and density of the sampling used for the quadrature approximation.

If neither of these works, you may then want to consider an alternative model, such as a fixed-effects, pooled, or population-averaged model. Alternatively, a different random-effects model whose likelihood is not approximated via quadrature (for example, \texttt{xtpoisson, re}) may be a better choice.
Example 1

Here we synthesize data according to the model

\[ E(y) = 0.05 x_1 + 0.08 x_2 + 0.08 x_3 + 0.1 x_4 + 0.1 x_5 + 0.1 x_6 + 0.1 \varepsilon \]

\[ z = \begin{cases} 
1 & \text{if } y \geq 0 \\
0 & \text{if } y < 0 
\end{cases} \]

where the intrapanel correlation is 0.5 and the \( x_1 \) variable is constant within panels. We first fit a random-effects probit model, and then we check the stability of the quadrature calculation:

```
. use http://www.stata-press.com/data/r13/quad1
. xtset id
   panel variable:  id (balanced)
. xtprobit z x1-x6
   (output omitted)
```

```
Random-effects probit regression
Number of obs  =  6000
Group variable:  id
Number of groups  =  300
Random effects u_i ~ Gaussian
Obs per group:  min =  20
                avg =  20.0
                max =  20
Integration method:  mvaghermite
Integration points =  12
Wald chi2(6)  =  29.24
Log likelihood =  -3347.1097  Prob > chi2 =  0.0001
```

```
z   Coef.     Std. Err.     z  P>|z|     [95% Conf. Interval]
----------|-------------------|---------|---------|-------------------|
x1  .0043068  .0607058  0.07  0.943  -.1146743  .1232879
x2  .1000742  .066331  1.51  0.131  -.0299323  .2300806
x3  .1503539  .0662503  2.27  0.023  .0205057  .2802021
x4  .123015  .0377089  3.26  0.001  .0491069  .196923
x5  .1342988  .0657222  2.04  0.041  .0054856  .263112
x6  .0879933  .0455573  1.93  0.054  -.0013325  .1773192
_cons  .0757067  .060359  1.25  0.210  -.0425948  .1940083
/lnsig2u  -.0329916  .1026847  -0.32  0.745  -.23425  .1682667
sigma_u  .9836395  .0505024  19.59  0.000  .889474  1.087774
rho  .4917528  .0256642  19.08  0.000  .4417038  .5419677
```

Likelihood-ratio test of rho=0:  chibar2(01) = 1582.67 Prob >= chibar2 = 0.000
We see that the largest difference is in the $x_1$ variable with a relative difference of 0.03% between the model with 12 integration points and 16. This example is somewhat rare in that the differences between eight quadrature points and 12 are smaller than those between 12 and 16. Usually the opposite occurs: the model results converge as you add quadrature points. Here we have an indication that perhaps some minor feature of the model was missed with eight points and 12 but seen with 16. Because all differences are very small, we could accept this model as is. We would like to have a largest relative difference of about 0.01%, and this is close. The differences and relative differences are small, indicating that refitting the random-effects probit model with a few more integration points will yield a satisfactory result. Indeed, refitting the model with the `intpoints(20)` option yields completely satisfactory results when checked with `quadchk`.

Nonadaptive Gauss–Hermite quadrature does not yield such robust results.
. xtprobit z x1-x6, intmethod(ghermite) nolog

Random-effects probit regression
Number of obs = 6000
Group variable: id
Number of groups = 300
Random effects u_i ~ Gaussian
Obs per group: min = 20
avg = 20.0
max = 20
Integration method: ghermite
Integration points = 12
Wald chi2(6) = 36.15
Log likelihood = -3349.6926 Prob > chi2 = 0.0000

|        | Coef.  | Std. Err. | z     | P>|z|   | [95% Conf. Interval] |
|--------|--------|-----------|-------|-------|----------------------|
| x1     |  .1156  | .0555     | 2.08  | 0.037 | .0069131 .2244396   |
| x2     |  .1006  | .0662     | 1.52  | 0.129 | -.0292469 .230358   |
| x3     |  .1542  | .0661     | 2.33  | 0.020 | .0246941 .2837433   |
| x4     |  .1258  | .0376     | 3.35  | 0.001 | .0521108 .1994123   |
| x5     |  .1366  | .0655     | 2.09  | 0.037 | .0082823 .2649182   |
| x6     |  .0870  | .0453     | 1.92  | 0.055 | -.0018497 .1759147  |
| _cons  |  .1098  | .0501     | 2.19  | 0.028 | .0117404 .2079382   |
| /lnsig2u|-.0791  | .0971     |       |       | -.2695071 .1111428  |

likelihood-ratio test of rho=0: chibar2(01) = 1577.50 Prob >= chibar2 = 0.000

sigma_u  .9611824  .0466685   .8739313  1.057145
rho     .4802148  .0242386   .4330281  .5277571
. quadchk, nooutput
Refitting model intpoints() = 8
Refitting model intpoints() = 16

<table>
<thead>
<tr>
<th>Quadrature check</th>
<th>Fitted quadrature</th>
<th>Comparison quadrature</th>
<th>Comparison quadrature</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>12 points</td>
<td>8 points</td>
<td>16 points</td>
</tr>
<tr>
<td>Log likelihood</td>
<td>-3349.6926</td>
<td>-3354.6372</td>
<td>-3348.3881</td>
</tr>
<tr>
<td></td>
<td>-4.9446636</td>
<td>1.3045063</td>
<td>.00147615</td>
</tr>
<tr>
<td></td>
<td>.00147615</td>
<td>-.00038944</td>
<td></td>
</tr>
<tr>
<td>z:</td>
<td>.11567633</td>
<td>.16153998</td>
<td>.07007833</td>
</tr>
<tr>
<td>x1</td>
<td>.04586365</td>
<td>-.045598</td>
<td>.39648262</td>
</tr>
<tr>
<td></td>
<td>.39648262</td>
<td>-.39418608</td>
<td></td>
</tr>
<tr>
<td>z:</td>
<td>.10055552</td>
<td>.10317831</td>
<td>.09937417</td>
</tr>
<tr>
<td>x2</td>
<td>.002602279</td>
<td>-.00118135</td>
<td>.02608297</td>
</tr>
<tr>
<td></td>
<td>.02608297</td>
<td>-.1174825</td>
<td></td>
</tr>
<tr>
<td>z:</td>
<td>.1542187</td>
<td>.15465369</td>
<td>.15150516</td>
</tr>
<tr>
<td>x3</td>
<td>.00043499</td>
<td>-.00271354</td>
<td>.00282062</td>
</tr>
<tr>
<td></td>
<td>.00282062</td>
<td>-.0175954</td>
<td></td>
</tr>
<tr>
<td>z:</td>
<td>.12576159</td>
<td>.12880254</td>
<td>.1243974</td>
</tr>
<tr>
<td>x4</td>
<td>.00304096</td>
<td>-.00136418</td>
<td>.02418032</td>
</tr>
<tr>
<td></td>
<td>.02418032</td>
<td>-.01084739</td>
<td></td>
</tr>
<tr>
<td>z:</td>
<td>.13660028</td>
<td>.13475211</td>
<td>.13707075</td>
</tr>
<tr>
<td>x5</td>
<td>-.00184817</td>
<td>.00047047</td>
<td>-.01352978</td>
</tr>
<tr>
<td></td>
<td>-.01352978</td>
<td>.00344411</td>
<td></td>
</tr>
<tr>
<td>z:</td>
<td>.08703252</td>
<td>.08568342</td>
<td>.08738135</td>
</tr>
<tr>
<td>x6</td>
<td>-.0013491</td>
<td>.00034883</td>
<td>-.0155011</td>
</tr>
<tr>
<td></td>
<td>-.0155011</td>
<td>.00400809</td>
<td></td>
</tr>
<tr>
<td>z:</td>
<td>.10983928</td>
<td>.11031299</td>
<td>.09654975</td>
</tr>
<tr>
<td>_cons</td>
<td>.00047371</td>
<td>-.01328953</td>
<td>.00431274</td>
</tr>
<tr>
<td></td>
<td>.00431274</td>
<td>-.12099067</td>
<td></td>
</tr>
<tr>
<td>ln.sig2u:</td>
<td>-.07918212</td>
<td>-.18133821</td>
<td>-.05815644</td>
</tr>
<tr>
<td>_cons</td>
<td>-.10215609</td>
<td>.02102568</td>
<td>1.2901408</td>
</tr>
<tr>
<td></td>
<td>1.2901408</td>
<td>-.26553572</td>
<td></td>
</tr>
</tbody>
</table>

Here we see that the x1 variable (the one that was constant within panel) changed with a relative difference of nearly 40%! This example clearly demonstrates the benefit of adaptive quadrature methods.
Example 2

Here we rerun the previous nonadaptive quadrature model, but using the `intpoints(120)` option to increase the number of integration points to 120. We get results close to those from adaptive quadrature and an acceptable `quadchk`. This example demonstrates the efficacy of increasing the number of integration points to improve the quadrature approximation.

```
.xtprobit z x1-x6, intmethod(ghermite) intpoints(120) nolog
Random-effects probit regression
Number of obs = 6000
Group variable: id
Number of groups = 300
Random effects u_i ~ Gaussian
Obs per group: min = 20
avg = 20.0
max = 20
Integration method: ghermite
Integration points = 120
Wald chi2(6) = 29.24
Log likelihood = -3347.1099
Prob > chi2 = 0.0001

|    | Coef.  | Std. Err. | z     | P>|z|  | [95% Conf. Interval] |
|----|--------|-----------|-------|------|----------------------|
| z  |        |           |       |      |                      |
|x1  | .0043059 | .0607087  | 0.07  | 0.943| -.114681 .1232929    |
|x2  | .1000743 | .0663311  | 1.51  | 0.131| -.0299322 .2300808   |
|x3  | .1503541 | .0662503  | 2.27  | 0.023| -.0205058 .2802023   |
|x4  | .1230151 | .0677089  | 3.26  | 0.001| .049107 .1969232     |
|x5  | .134299  | .0657223  | 2.04  | 0.041| .0054865 .2631123    |
|x6  | .0879935 | .0455753  | 1.93  | 0.054| -.0013325 .1773194   |
|_cons| .0757054 | .0603621  | 1.25  | 0.210| -.0426021 .1940128   |
_/lnsig2u| -.0329832| .1026863  |       |      | -.2342446 .1682783   |
|sigma_u| .9836437 | .0505034  |       |      | .8894764 1.08778     |
|rho  | .491755  | .0256646  |       |      | .4417052 .5419706    |
```

Likelihood-ratio test of rho=0: chibar2(01) = 1582.67 Prob > chibar2 = 0.000
. quadchk, nooutput
Refitting model intpoints() = 80
Refitting model intpoints() = 160

Quadrature check

<table>
<thead>
<tr>
<th>Fitted quadrature</th>
<th>Comparison quadrature</th>
<th>Comparison quadrature</th>
</tr>
</thead>
<tbody>
<tr>
<td>120 points</td>
<td>80 points</td>
<td>160 points</td>
</tr>
<tr>
<td>Log likelihood</td>
<td>-3347.1099</td>
<td>-3347.1099</td>
</tr>
<tr>
<td></td>
<td>-.00007138</td>
<td>2.440e-07</td>
</tr>
<tr>
<td></td>
<td>2.133e-08</td>
<td>-7.289e-11</td>
</tr>
<tr>
<td>Difference</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Relative difference</td>
<td></td>
<td></td>
</tr>
<tr>
<td>z: .00430592</td>
<td>.00431318</td>
<td>.00430553</td>
</tr>
<tr>
<td>x1</td>
<td>7.259e-06</td>
<td>-3.871e-07</td>
</tr>
<tr>
<td></td>
<td>.00168592</td>
<td>-.00008991</td>
</tr>
<tr>
<td>Difference</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Relative difference</td>
<td></td>
<td></td>
</tr>
<tr>
<td>z: .10007431</td>
<td>.10007415</td>
<td>.10007431</td>
</tr>
<tr>
<td>x2</td>
<td>-1.519e-07</td>
<td>5.585e-09</td>
</tr>
<tr>
<td></td>
<td>-1.517e-06</td>
<td>5.580e-08</td>
</tr>
<tr>
<td>Difference</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Relative difference</td>
<td></td>
<td></td>
</tr>
<tr>
<td>z: .15035406</td>
<td>.15035407</td>
<td>.15035406</td>
</tr>
<tr>
<td>x3</td>
<td>1.699e-08</td>
<td>7.636e-09</td>
</tr>
<tr>
<td></td>
<td>1.130e-07</td>
<td>5.078e-08</td>
</tr>
<tr>
<td>Difference</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Relative difference</td>
<td></td>
<td></td>
</tr>
<tr>
<td>z: .12301506</td>
<td>.12301512</td>
<td>.12301506</td>
</tr>
<tr>
<td>x4</td>
<td>6.036e-08</td>
<td>5.353e-09</td>
</tr>
<tr>
<td></td>
<td>4.907e-07</td>
<td>4.352e-08</td>
</tr>
<tr>
<td>Difference</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Relative difference</td>
<td></td>
<td></td>
</tr>
<tr>
<td>z: .13429895</td>
<td>.13429962</td>
<td>.13429896</td>
</tr>
<tr>
<td>x5</td>
<td>6.646e-07</td>
<td>4.785e-09</td>
</tr>
<tr>
<td></td>
<td>4.949e-06</td>
<td>3.563e-08</td>
</tr>
<tr>
<td>Difference</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Relative difference</td>
<td></td>
<td></td>
</tr>
<tr>
<td>z: .08799345</td>
<td>.08799334</td>
<td>.08799346</td>
</tr>
<tr>
<td>x6</td>
<td>-1.123e-07</td>
<td>3.049e-09</td>
</tr>
<tr>
<td></td>
<td>-1.276e-06</td>
<td>3.465e-08</td>
</tr>
<tr>
<td>Difference</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Relative difference</td>
<td></td>
<td></td>
</tr>
<tr>
<td>z: .07570536</td>
<td>.07570205</td>
<td>.07570442</td>
</tr>
<tr>
<td>_cons</td>
<td>-3.305e-06</td>
<td>-9.405e-07</td>
</tr>
<tr>
<td></td>
<td>-.00004365</td>
<td>-.00001242</td>
</tr>
<tr>
<td>Difference</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Relative difference</td>
<td></td>
<td></td>
</tr>
<tr>
<td>lnsig2u: -.03298317</td>
<td>-.03298909</td>
<td>-.03298186</td>
</tr>
<tr>
<td>_cons</td>
<td>-5.915e-06</td>
<td>1.304e-06</td>
</tr>
<tr>
<td></td>
<td>.00017945</td>
<td>-.00003952</td>
</tr>
<tr>
<td>Difference</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Relative difference</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Example 3

Here we synthesize data the same way as in the previous example, but we make the intrapanel correlation equal to 0.1 instead of 0.5. We again fit a random-effects probit model and check the quadrature:
. use http://www.stata-press.com/data/r13/quad2
. xtset id
    panel variable:  id (balanced)
. xtprobit z x1-x6

Fitting comparison model:
Iteration 0:  log likelihood = -4142.2915
Iteration 1:  log likelihood = -4120.4109
Iteration 2:  log likelihood = -4120.4099
Iteration 3:  log likelihood = -4120.4099

Fitting full model:
rho = 0.0  log likelihood = -4120.4099
rho = 0.1  log likelihood = -4065.7986
rho = 0.2  log likelihood = -4087.7703
Iteration 0:  log likelihood = -4065.7986
Iteration 1:  log likelihood = -4065.3157
Iteration 2:  log likelihood = -4065.3144
Iteration 3:  log likelihood = -4065.3144

Random-effects probit regression
Group variable:  id  Number of obs   =   6000
Random effects u_i ~ Gaussian  Number of groups =   300
Obs per group:  min =      20
                avg =     20.0
                max =      20
Integration method: mvaghermite  Integration points =      12
Wald chi2(6)    =     39.43
Log likelihood = -4065.3144

                     z           Coef. Std. Err.      z    P>|z|     [95% Conf. Interval]

x1            .0246943   .0251120     0.98  0.325    -.0245243    .0739129
x2            .1300123   .0587906     2.21  0.027     .0147847    .2452398
x3            .1190409   .0579539     2.05  0.040    .0054533    .2326284
x4            .1391970   .0331817     4.19  0.000     .0741621    .2042319
x5            .0773640   .0578454     1.34  0.181    -.0360111    .1907389
x6            .0862028   .0401185     2.15  0.032     .0075720    .1648336
    _cons        .0922653   .0244392     3.78  0.000     .0443653    .1401652

/lnsig2u    -2.343939   .1575275   -2.652687   -2.038191

sigma_u    .3097563   .0243976     26.54461    .3614631
rho     .0875487   .0125839     .0658236    .1155574

Likelihood-ratio test of rho=0:  chibar2(01) =  110.19  Prob >= chibar2 = 0.000
Here we see that the quadrature approximation is stable. With this result, we can confidently interpret the results. Satisfactory results are also obtained in this case with nonadaptive quadrature.
This entry describes the `vce_options`, which are common to most xt estimation commands. Not all the options documented below work with all xt estimation commands; see the documentation for the particular estimation command. If an option is listed there, it is applicable.

The `vce()` option specifies how to estimate the variance–covariance matrix (VCE) corresponding to the parameter estimates. The standard errors reported in the table of parameter estimates are the square root of the variances (diagonal elements) of the VCE.

### Options

<table>
<thead>
<tr>
<th><code>vce()</code> option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>vce(oim)</code></td>
<td>observed information matrix (OIM)</td>
</tr>
<tr>
<td><code>vce(opg)</code></td>
<td>outer product of the gradient (OPG) vectors</td>
</tr>
<tr>
<td><code>vce(robust)</code></td>
<td>Huber/White/sandwich estimator</td>
</tr>
<tr>
<td><code>vce(cluster clustvar)</code></td>
<td>clustered sandwich estimator</td>
</tr>
<tr>
<td><code>vce(bootstrap[, bootstrap_options])</code></td>
<td>bootstrap estimation</td>
</tr>
<tr>
<td><code>vce(jackknife[, jackknife_options])</code></td>
<td>jackknife estimation</td>
</tr>
</tbody>
</table>

- `vce(oim)` is usually the default for models fit using maximum likelihood. `vce(oim)` uses the observed information matrix (OIM); see [R] `ml`.
- `vce(opg)` uses the sum of the outer product of the gradient (OPG) vectors; see [R] `ml`. This is the default VCE when the `technique(bhhh)` option is specified; see [R] `maximize`.
- `vce(robust)` uses the robust or sandwich estimator of variance. This estimator is robust to some types of misspecification so long as the observations are independent; see [U] 20.21 Obtaining robust variance estimates.

If the command allows `pweights` and you specify them, `vce(robust)` is implied; see [U] 20.23.3 Sampling weights.
**vce_options** — Variance estimators

**vce(cluster clustvar)** specifies that the standard errors allow for intragroup correlation, relaxing the usual requirement that the observations be independent. That is to say, the observations are independent across groups (clusters) but not necessarily within groups. **clustvar** specifies to which group each observation belongs, for example, **vce(cluster personid)** in data with repeated observations on individuals. **vce(cluster clustvar)** affects the standard errors and variance–covariance matrix of the estimators but not the estimated coefficients; see **[U] 20.21 Obtaining robust variance estimates.**

**vce(bootstrap [ , bootstrap_options ])** uses a nonparametric bootstrap; see **[R] bootstrap.** After estimation with **vce(bootstrap),** see **[R] bootstrap postestimation** to obtain percentile-based or bias-corrected confidence intervals.

**vce(jackknife [ , jackknife_options ])** uses the delete-one jackknife; see **[R] jackknife.**

**nmp** specifies that the divisor $N - P$ be used instead of the default $N$, where $N$ is the total number of observations and $P$ is the number of coefficients estimated.

**scale(x2 | dev | phi | #)** overrides the default scale parameter. By default, **scale(1)** is assumed for the discrete distributions (binomial, negative binomial, and Poisson), and **scale(x2)** is assumed for the continuous distributions (gamma, Gaussian, and inverse Gaussian).

**scale(x2)** specifies that the scale parameter be set to the Pearson chi-squared (or generalized chi-squared) statistic divided by the residual degrees of freedom, which is recommended by McCullagh and Nelder (1989) as a good general choice for continuous distributions.

**scale(dev)** sets the scale parameter to the deviance divided by the residual degrees of freedom. This option provides an alternative to **scale(x2)** for continuous distributions and for over- or underdispersed discrete distributions.

**scale(phi)** specifies that the scale parameter be estimated from the data. xtgee’s default scaling makes results agree with other estimators and has been recommended by McCullagh and Nelder (1989) in the context of GLM. When comparing results with calculations made by other software, you may find that the other packages do not offer this feature. In such cases, specifying **scale(phi)** should match their results.

**scale(#)** sets the scale parameter to #. For example, using **scale(1)** in **family(gamma)** models results in exponential-errors regression (if you assume independent correlation structure).

**Remarks and examples**

When you are working with panel-data models, we strongly encourage you to use the **vce(bootstrap) or vce(jackknife)** option instead of the corresponding prefix command. For example, to obtain jackknife standard errors with **xtlogit**, type
. use http://www.stata-press.com/data/r13/clogitid
. xtlogit y x1 x2, fe vce(jackknife)
    (running xtlogit on estimation sample)
Jackknife replications (66)
.................................................. 50
...........................
Conditional fixed-effects logistic regression
    Number of obs = 369
    Group variable: id
    Number of groups = 66
    Obs per group: min = 2
                   avg = 5.6
                   max = 10
    F( 2, 65) = 4.58
    Log likelihood = -123.41386
    Prob > F = 0.0137
    (Replications based on 66 clusters in id)

| y     | Coef. | Std. Err. | t     | P>|t|  | [95% Conf. Interval] |
|-------|-------|-----------|-------|------|----------------------|
| x1    | .653363 | .3010608 | 2.17  | 0.034 | .0623717 1.2454623 |
| x2    | .0659169 | .0487858 | 1.35  | 0.181 | -.0315151 .1633489 |

If you wish to specify more options to the bootstrap or jackknife estimation, you can include them within the `vce()` option. Below we refit our model requesting bootstrap standard errors based on 300 replications, we set the random-number seed so that our results can be reproduced, and we suppress the display of the replication dots.

. xtlogit y x1 x2, fe vce(bootstrap, reps(300) seed(123) nodots)
Conditional fixed-effects logistic regression
    Number of obs = 369
    Group variable: id
    Number of groups = 66
    Obs per group: min = 2
                   avg = 5.6
                   max = 10
    Wald chi2(2) = 8.52
    Log likelihood = -123.41386
    Prob > chi2 = 0.0141
    (Replications based on 66 clusters in id)

| y     | Observed Coef. | Bootstrap Std. Err. | z         | P>|z|  | [95% Conf. Interval] |
|-------|----------------|---------------------|-----------|------|----------------------|
| x1    | .653363        | .3015317            | 2.17      | 0.030 | .0623717 1.244354   |
| x2    | .0659169       | .0512331            | 1.29      | 0.198 | -.0344981 .1663319  |

Technical note

To perform jackknife estimation on panel data, you must omit entire panels rather than individual observations. To replicate the output above using the `jackknife` prefix command, you would have to type

. jackknife, cluster(id): xtlogit y x1 x2, fe
(output omitted)

Similarly, bootstrap estimation on panel data requires you to resample entire panels rather than individual observations. The `vce(bootstrap)` and `vce(jackknife)` options handle this for you automatically.
Methods and formulas

By default, Stata’s maximum likelihood estimators display standard errors based on variance estimates given by the inverse of the negative Hessian (second derivative) matrix. If \texttt{vce(robust)}, \texttt{vce(cluster clustvar)}, or \texttt{pweights} are specified, standard errors are based on the robust variance estimator (see [U] 20.21 Obtaining robust variance estimates); likelihood-ratio tests are not appropriate here (see [SVY] survey), and the model $\chi^2$ is from a Wald test. If \texttt{vce(opg)} is specified, the standard errors are based on the outer product of the gradients; this option has no effect on likelihood-ratio tests, though it does affect Wald tests.

If \texttt{vce(bootstrap)} or \texttt{vce(jackknife)} is specified, the standard errors are based on the chosen replication method; here the model $\chi^2$ or $F$ statistic is from a Wald test using the respective replication-based covariance matrix. The $t$ distribution is used in the coefficient table when the \texttt{vce(jackknife)} option is specified. \texttt{vce(bootstrap)} and \texttt{vce(jackknife)} are also available with some commands that are not maximum likelihood estimators.

Reference


Also see

[R] bootstrap — Bootstrap sampling and estimation
[R] jackknife — Jackknife estimation
[R] ml — Maximum likelihood estimation
[U] 20 Estimation and postestimation commands
Syntax

\texttt{xtabond \ depvar \ indepvars} \ [ \ if \ ] \ [ \ in \ ] \ [ , \ options ]

\begin{tabular}{ll}
\textbf{options} & \textbf{Description} \\
\hline
\texttt{noconstant} & suppress constant term \\
\texttt{diffvars(varlist)} & already-differenced exogenous variables \\
\texttt{inst(varlist)} & additional instrument variables \\
\texttt{lags(#)} & use \# lags of dependent variable as covariates; default is \texttt{lags(1)} \\
\texttt{maxldep(#)} & maximum lags of dependent variable for use as instruments \\
\texttt{maxlags(#)} & maximum lags of predetermined and endogenous variables for use as instruments \\
\texttt{twostep} & compute the two-step estimator instead of the one-step estimator \\
\texttt{pre(varlist[...])} & predetermined variables; can be specified more than once \\
\texttt{endogenous(varlist[...])} & endogenous variables; can be specified more than once \\
\\hline
\texttt{vce(vcetype)} & \texttt{vcetype} may be \texttt{gmm} or \texttt{robust} \\
\texttt{level(#)} & set confidence level; default is \texttt{level(95)} \\
\texttt{artests(#)} & use \# as maximum order for AR tests; default is \texttt{artests(2)} \\
\texttt{display_options} & control spacing and line width \\
\texttt{coeflegend} & display legend instead of statistics \\
\end{tabular}

A panel variable and a time variable must be specified; use \texttt{xtset}; see \cite{xtset}. \\indepvars\ and all \varlists, except \texttt{pre(varlist[...])} and \texttt{endogenous(varlist[...])}, may contain time-series operators; see \cite{Time-series varlists}. The specification of \texttt{depvar}, however, may not contain time-series operators. \texttt{by}, \texttt{statsby}, and \texttt{xi} are allowed; see \cite{Prefix commands}. \texttt{coeflegend} does not appear in the dialog box. \texttt{See \cite{Estimation and postestimation commands}} for more capabilities of estimation commands.
**Description**

Linear dynamic panel-data models include $p$ lags of the dependent variable as covariates and contain unobserved panel-level effects, fixed or random. By construction, the unobserved panel-level effects are correlated with the lagged dependent variables, making standard estimators inconsistent. Arellano and Bond (1991) derived a consistent generalized method of moments (GMM) estimator for the parameters of this model; xtabond implements this estimator.

This estimator is designed for datasets with many panels and few periods, and it requires that there be no autocorrelation in the idiosyncratic errors. For a related estimator that uses additional moment conditions, but still requires no autocorrelation in the idiosyncratic errors, see [XT] xtdpdsys. For estimators that allow for some autocorrelation in the idiosyncratic errors, at the cost of a more complicated syntax, see [XT] xtdpd.

**Options**

**Model**

- **noconstant;** see [R] estimation options.
- **diffvars(varlist)** specifies a set of variables that already have been differenced to be included as strictly exogenous covariates.
- **inst(varlist)** specifies a set of variables to be used as additional instruments. These instruments are not differenced by xtabond before including them in the instrument matrix.
- **lags(#)** sets $p$, the number of lags of the dependent variable to be included in the model. The default is $p = 1$.
- **maxldep(#)** sets the maximum number of lags of the dependent variable that can be used as instruments. The default is to use all $T_i - p - 2$ lags.
- **maxlags(#)** sets the maximum number of lags of the predetermined and endogenous variables that can be used as instruments. For predetermined variables, the default is to use all $T_i - p - 1$ lags. For endogenous variables, the default is to use all $T_i - p - 2$ lags.
- **twostep** specifies that the two-step estimator be calculated.

**Predetermined**

- **pre(varlist, lagstruct(prelags, premaxlags))** specifies that a set of predetermined variables be included in the model. Optionally, you may specify that prelags lags of the specified variables also be included. The default for prelags is 0. Specifying premaxlags sets the maximum number of further lags of the predetermined variables that can be used as instruments. The default is to include $T_i - p - 1$ lagged levels as instruments for predetermined variables. You may specify as many sets of predetermined variables as you need within the standard Stata limits on matrix size. Each set of predetermined variables may have its own number of prelags and premaxlags.

**Endogenous**

- **endogenous(varlist, lagstruct(endlags, endmaxlags))** specifies that a set of endogenous variables be included in the model. Optionally, you may specify that endlags lags of the specified variables also be included. The default for endlags is 0. Specifying endmaxlags sets the maximum number of further lags of the endogenous variables that can be used as instruments. The default is to include $T_i - p - 2$ lagged levels as instruments for endogenous variables. You may specify as many sets of endogenous variables as you need within the standard Stata limits on matrix size. Each set of endogenous variables may have its own number of endlags and endmaxlags.
SE/Robust

vce(vcetype) specifies the type of standard error reported, which includes types that are derived from asymptotic theory and that are robust to some kinds of misspecification; see Remarks and examples below.

vce(gmm), the default, uses the conventionally derived variance estimator for generalized method of moments estimation.

vce(robust) uses the robust estimator. After one-step estimation, this is the Arellano–Bond robust VCE estimator. After two-step estimation, this is the Windmeijer (2005) WC-robust estimator.

level(#) specifies the maximum order of the autocorrelation test to be calculated. The tests are reported by estat abond; see [XT] xtabond postestimation. Specifying the order of the highest test at estimation time is more efficient than specifying it to estat abond, because estat abond must refit the model to obtain the test statistics. The maximum order must be less than or equal to the number of periods in the longest panel. The default is artests(2).

display_options: vsquish and nolstretch; see [R] estimation options.

The following option is available with xtabond but is not shown in the dialog box:

coeflegend; see [R] estimation options.

Remarks and examples

Anderson and Hsiao (1981, 1982) propose using further lags of the level or the difference of the dependent variable to instrument the lagged dependent variables that are included in a dynamic panel-data model after the panel-level effects have been removed by first-differencing. A version of this estimator can be obtained from xtitreg (see [XT] xtitreg). Arellano and Bond (1991) build upon this idea by noting that, in general, there are many more instruments available. Building on Holtz-Eakin, Newey, and Rosen (1988) and using the GMM framework developed by Hansen (1982), they identify how many lags of the dependent variable, the predetermined variables, and the endogenous variables are valid instruments and how to combine these lagged levels with first differences of the strictly exogenous variables into a potentially large instrument matrix. Using this instrument matrix, Arellano and Bond (1991) derive the corresponding one-step and two-step GMM estimators, as well as the robust VCE estimator for the one-step model. They also found that the robust two-step VCE was seriously biased. Windmeijer (2005) worked out a bias-corrected (WC) robust estimator for VCEs of two-step GMM estimators, which is implemented in xtabond. The test of autocorrelation of order \( m \) and the Sargan test of overidentifying restrictions derived by Arellano and Bond (1991) can be obtained with estat abond and estat sargan, respectively; see [XT] xtabond postestimation.

Example 1: One-step estimator

Arellano and Bond (1991) apply their new estimators and test statistics to a model of dynamic labor demand that had previously been considered by Layard and Nickell (1986) using data from an unbalanced panel of firms from the United Kingdom. All variables are indexed over the firm \( i \) and time \( t \). In this dataset, \( n_{it} \) is the log of employment in firm \( i \) at time \( t \), \( w_{it} \) is the natural log of the real product wage, \( k_{it} \) is the natural log of the gross capital stock, and \( y_{st} \) is the natural log of industry output. The model also includes time dummies yr1980, yr1981, yr1982, yr1983, and yr1984. In table 4 of Arellano and Bond (1991), the authors present the results they obtained from several specifications.
In column a1 of table 4, Arellano and Bond report the coefficients and their standard errors from the robust one-step estimators of a dynamic model of labor demand in which \( n_{it} \) is the dependent variable and its first two lags are included as regressors. To clarify some important issues, we will begin with the homoskedastic one-step version of this model and then consider the robust case. Here is the command using \texttt{xtabond} and the subsequent output for the homoskedastic case:

```stata
use http://www.stata-press.com/data/r13/abdata
xtabond n l(0/1).w l(0/2).(k ys) yr1980-yr1984 year, lags(2) noconstant
```

|               | Coef. | Std. Err. | z     | P>|z| | [95% Conf. Interval] |
|---------------|-------|-----------|-------|------|----------------------|
| **n**         |       |           |       |      |                      |
| L1.           | .6862261 | .1486163  | 4.62  | 0.000 | .3949435 .9775088    |
| L2.           | -.0853582 | .0444365  | -1.92 | 0.055 | -.1724523 .0017358   |
| **w**         |       |           |       |      |                      |
| --.           | -.6078208 | .0657694  | -9.24 | 0.000 | -.7367265 -.4789151  |
| L1.           | .3926237 | .1092374  | 3.59  | 0.000 | .1785222 .6067251    |
| **k**         |       |           |       |      |                      |
| --.           | .3568456 | .0370314  | 9.64  | 0.000 | .2842653 .4294259    |
| L1.           | -.0580012 | .0583051  | -0.99 | 0.320 | -.172277 .0562747    |
| L2.           | -.0199475 | .0416274  | -0.48 | 0.632 | -.1015357 .0616408   |
| **ys**        |       |           |       |      |                      |
| --.           | .6085073 | .1345412  | 4.52  | 0.000 | .3448115 .8722031    |
| L1.           | -.7111651 | .1844599  | -3.86 | 0.000 | -.1.0727 -.3496304   |
| L2.           | .1057969 | .1428568  | 0.74  | 0.459 | -.1741974 .3857912   |
| **yr1980**    |       |           |       |      |                      |
|               | .0029062 | .0212705  | 0.14  | 0.891 | -.0387832 .0445957   |
| **yr1981**    |       |           |       |      |                      |
|               | -.0404378 | .0354707  | -1.14 | 0.254 | -.1099591 .0290836   |
| **yr1982**    |       |           |       |      |                      |
|               | -.0652767 | .048209   | -1.35 | 0.176 | -.1.597646 .0292111  |
| **yr1983**    |       |           |       |      |                      |
|               | -.0690928 | .0627354  | -1.10 | 0.271 | -.1.920521 .0538664  |
| **yr1984**    |       |           |       |      |                      |
|               | -.0650302 | .0781322  | -0.83 | 0.405 | -.2181665 .0881061   |
| **year**      |       |           |       |      |                      |
|               | .0095545 | .0142073  | 0.67  | 0.501 | -.0182912 .0374002   |

The coefficients are identical to those reported in column a1 of table 4, as they should be. Of course, the standard errors are different because we are considering the homoskedastic case. Although the moment conditions use first-differenced errors, \texttt{xtabond} estimates the coefficients of the level model and reports them accordingly.

The footer in the output reports the instruments used. The first line indicates that \texttt{xtabond} used lags from 2 on back to create the GMM-type instruments described in Arellano and Bond (1991) and Holtz-Eakin, Newey, and Rosen (1988); also see \textit{Methods and formulas} in \texttt{[XT] xtdpd}. The second and third lines indicate that the first difference of all the exogenous variables were used as standard instruments. GMM-type instruments use the lags of a variable to contribute multiple columns to the
instrument matrix, whereas each standard instrument contributes one column to the instrument matrix. The notation \(L(2/)\cdot n\) indicates that GMM-type instruments were created using lag 2 of \(n\) from on back. \((L(2/4)\cdot n\) would indicate that GMM-type instruments were created using only lags 2, 3, and 4 of \(n\).

After \texttt{xtabond}, \texttt{estat sargan} reports the Sargan test of overidentifying restrictions.

\begin{verbatim}
. estat sargan
Sargan test of overidentifying restrictions
  H0: overidentifying restrictions are valid
  chi2(25)     =  65.81806
  Prob > chi2  =  0.0000
\end{verbatim}

Only for a homoskedastic error term does the Sargan test have an asymptotic chi-squared distribution. In fact, Arellano and Bond (1991) show that the one-step Sargan test overrejects in the presence of heteroskedasticity. Because its asymptotic distribution is not known under the assumptions of the \texttt{vce(robust)} model, \texttt{xtabond} does not compute it when \texttt{vce(robust)} is specified. The Sargan test, reported by Arellano and Bond (1991, table 4, column a1), comes from the one-step homoskedastic estimator and is the same as the one reported here. The output above presents strong evidence against the null hypothesis that the overidentifying restrictions are valid. Rejecting this null hypothesis implies that we need to reconsider our model or our instruments, unless we attribute the rejection to heteroskedasticity in the data-generating process. Although performing the Sargan test after the two-step estimator is an alternative, Arellano and Bond (1991) found a tendency for this test to underreject in the presence of heteroskedasticity. (See \texttt{XT xtdpd} for an example indicating that this rejection may be due to misspecification.)

By default, \texttt{xtabond} calculates the Arellano–Bond test for first- and second-order autocorrelation in the first-differenced errors. (Use \texttt{artests()} to compute tests for higher orders.) There are versions of this test for both the homoskedastic and the robust cases, although their values are different. Use \texttt{estat abond} to report the test results.

\begin{verbatim}
. estat abond
Arellano-Bond test for zero autocorrelation in first-differenced errors

<table>
<thead>
<tr>
<th>Order</th>
<th>z</th>
<th>Prob &gt; z</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-3.9394</td>
<td>0.0001</td>
</tr>
<tr>
<td>2</td>
<td>-0.54239</td>
<td>0.5876</td>
</tr>
</tbody>
</table>

H0: no autocorrelation
\end{verbatim}

When the idiosyncratic errors are independently and identically distributed (i.i.d.), the first-differenced errors are first-order serially correlated. So, as expected, the output above presents strong evidence against the null hypothesis of zero autocorrelation in the first-differenced errors at order 1. Serial correlation in the first-differenced errors at an order higher than 1 implies that the moment conditions used by \texttt{xtabond} are not valid; see \texttt{XT xtdpd} for an example of an alternative estimation method. The output above presents no significant evidence of serial correlation in the first-differenced errors at order 2.
Example 2: A one-step estimator with robust VCE

Consider the output from the one-step robust estimator of the same model:

```
xtabond n l(0/1).w l(0/2).(k ys) yr1980-yr1984 year, lags(2) vce(robust) 
> noconstant
```

Arellano-Bond dynamic panel-data estimation

<table>
<thead>
<tr>
<th>Number of obs</th>
<th>611</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group variable: id</td>
<td>140</td>
</tr>
<tr>
<td>Time variable: year</td>
<td></td>
</tr>
<tr>
<td>Obs per group:</td>
<td>4</td>
</tr>
<tr>
<td>avg = 4.364286</td>
<td></td>
</tr>
<tr>
<td>max = 6</td>
<td></td>
</tr>
<tr>
<td>Number of instruments = 41</td>
<td></td>
</tr>
<tr>
<td>Wald chi2(16) = 1727.45</td>
<td></td>
</tr>
<tr>
<td>Prob &gt; chi2 = 0.0000</td>
<td></td>
</tr>
</tbody>
</table>

One-step results

(Std. Err. adjusted for clustering on id)

| Coef. | Std. Err. | z | P>|z| | [95% Conf. Interval] |
|-------|-----------|---|------|---------------------|
| n L1. | .6862261  | .1445943  | 4.75 | 0.000 | .4028266 .9696257 |
| L2. | -.0853582 | .0560155  | -1.52 | 0.128 | -.1951467 .0244302 |
| w --. | -.6078208 | .1782055  | -3.41 | 0.001 | -.9570972 -.2585445 |
| L1. | .3926237  | .1679931  | 2.34 | 0.019 | .0633632 .7218842 |
| k --. | .3568456  | .0590203  | 6.05 | 0.000 | .241168 .4725233 |
| L1. | -.0580012 | .0731797  | -.79 | 0.428 | -.2014308 .0854284 |
| L2. | -.0199475 | .0590203  | -0.61 | 0.542 | -.0840631 .0441681 |
| ys --. | .6085073  | .1725313  | 3.53 | 0.000 | .2703522 .9466624 |
| L1. | -.7111651 | .2317163  | -3.07 | 0.002 | -1.165321 -.2570095 |
| L2. | .1057969  | .1412021  | 0.75 | 0.454 | -.1709542 .382548 |
| yr1980 | .0029062 | .0158028  | 0.18 | 0.854 | -.0280667 .0338791 |
| yr1981 | -.0404378 | .0280582  | -1.44 | 0.150 | -.0954307 .0145552 |
| yr1982 | -.0652767 | .0365451  | -1.79 | 0.074 | -.1369038 .0063503 |
| yr1983 | -.0690928 | .047413   | -1.46 | 0.145 | -.1620205 .0238348 |
| yr1984 | -.0650302 | .0576305  | -1.13 | 0.259 | -.1779839 .0479235 |
| year | .0095545  | .0102896  | 0.93 | 0.353 | -.0106127 .0297217 |

Instruments for differenced equation

GMM-type: L(2/).n

The coefficients are the same, but now the standard errors match that reported in Arellano and Bond (1991, table 4, column a1). Most of the robust standard errors are higher than those that assume a homoskedastic error term.
The Sargan statistic cannot be calculated after requesting a robust VCE, but robust tests for serial correlation are available.

```
. estat abond
Arellano-Bond test for zero autocorrelation in first-differenced errors
```

<table>
<thead>
<tr>
<th>Order</th>
<th>z</th>
<th>Prob &gt; z</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-3.5996</td>
<td>0.0003</td>
</tr>
<tr>
<td>2</td>
<td>-0.51603</td>
<td>0.6058</td>
</tr>
</tbody>
</table>

H0: no autocorrelation

The value of the test for second-order autocorrelation matches those reported in Arellano and Bond (1991, table 4, column a1) and presents no evidence of model misspecification.

Example 3: The Wald model test

xtabond reports the Wald statistic of the null hypothesis that all the coefficients except the constant are zero. Here the null hypothesis is that all the coefficients are zero, because there is no constant in the model. In our previous example, the null hypothesis is soundly rejected. In column a1 of table 4, Arellano and Bond report a chi-squared test of the null hypothesis that all the coefficients are zero, except the time trend and the time dummies. Here is this test in Stata:

```
. test l.n l2.n w l.w k l.k l2.k ys l.ys l2.ys
   ( 1) L.n = 0
   ( 2) L2.n = 0
   ( 3) w = 0
   ( 4) L.w = 0
   ( 5) k = 0
   ( 6) L.k = 0
   ( 7) L2.k = 0
   ( 8) ys = 0
   ( 9) L.ys = 0
  (10) L2.ys = 0
    chi2( 10) = 408.29
    Prob > chi2 = 0.0000
```

Example 4: A two-step estimator with Windmeijer bias-corrected robust VCE

The two-step estimator with the Windmeijer bias-corrected robust VCE of the same model produces the following output:
xtabond  —  Arellano–Bond linear dynamic panel-data estimation

```
    . xtabond n l(0/1).w l(0/2).(k ys) yr1980-yr1984 year, lags(2) twostep
    > vce(robust) noconstant

Arellano-Bond dynamic panel-data estimation

Number of obs = 611
Group variable: id
Number of groups = 140
Time variable: year
Obs per group:  min = 4
               avg = 4.364286
               max = 6
Number of instruments = 41
Wald chi2(16) = 1104.72
Prob > chi2 = 0.0000

Two-step results

(Std. Err. adjusted for clustering on id)

<table>
<thead>
<tr>
<th></th>
<th>WC-Robust</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coef. Std. Err.</td>
<td>z</td>
<td>P&gt;</td>
<td>z</td>
</tr>
<tr>
<td>n</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L1.</td>
<td>.6287089 .1934138</td>
<td>3.25</td>
<td>0.001</td>
<td>.2496248 1.007793</td>
</tr>
<tr>
<td>L2.</td>
<td>-.0651882 .0450501</td>
<td>-1.45</td>
<td>0.148</td>
<td>-.1534847 .0231084</td>
</tr>
<tr>
<td>w</td>
<td>-.5257597 .1546107</td>
<td>-3.40</td>
<td>0.001</td>
<td>-.828791 -.2227284</td>
</tr>
<tr>
<td>L1.</td>
<td>.3112899 .2030006</td>
<td>1.53</td>
<td>0.125</td>
<td>-.086584 .7091638</td>
</tr>
<tr>
<td>k</td>
<td>.2783619 .0728019</td>
<td>3.82</td>
<td>0.000</td>
<td>.1356728 .421051</td>
</tr>
<tr>
<td>L1.</td>
<td>.0140994 .0924575</td>
<td>0.15</td>
<td>0.879</td>
<td>-.167114 .1954129</td>
</tr>
<tr>
<td>L2.</td>
<td>-.0402484 .0432745</td>
<td>-0.93</td>
<td>0.352</td>
<td>-.1250649 .0445681</td>
</tr>
<tr>
<td>ys</td>
<td>-.5919243 .1730916</td>
<td>3.42</td>
<td>0.001</td>
<td>.252671 .9311776</td>
</tr>
<tr>
<td>L1.</td>
<td>-.5659863 .2611008</td>
<td>-2.17</td>
<td>0.030</td>
<td>-.1077734 -.0842381</td>
</tr>
<tr>
<td>L2.</td>
<td>.1005433 .1610987</td>
<td>0.62</td>
<td>0.533</td>
<td>-.2152043 .4162908</td>
</tr>
<tr>
<td>yr1980</td>
<td>.0006378 .0168042</td>
<td>0.04</td>
<td>0.970</td>
<td>-.0322978 .0335734</td>
</tr>
<tr>
<td>yr1981</td>
<td>-.0550044 .0313389</td>
<td>-1.76</td>
<td>0.079</td>
<td>-.1164275 .0064187</td>
</tr>
<tr>
<td>yr1982</td>
<td>-.075978 .0419276</td>
<td>-1.81</td>
<td>0.070</td>
<td>-.1581545 .0061986</td>
</tr>
<tr>
<td>yr1983</td>
<td>-.0740708 .0528381</td>
<td>-1.40</td>
<td>0.161</td>
<td>-.1776315 .02949</td>
</tr>
<tr>
<td>yr1984</td>
<td>-.0906606 .0642615</td>
<td>-1.41</td>
<td>0.158</td>
<td>-.2166108 .0352896</td>
</tr>
<tr>
<td>year</td>
<td>.0112155 .0116783</td>
<td>0.96</td>
<td>0.337</td>
<td>-.0116735 .0341045</td>
</tr>
</tbody>
</table>

Instruments for differenced equation
GMM-type: L(2/.).n
Standard: D.w LD.w D.k LD.k L2D.k D.ys LD.ys L2D.ys D.yr1980
```

Arellano and Bond recommend against using the two-step nonrobust results for inference on the coefficients because the standard errors tend to be biased downward (see Arellano and Bond 1991 for details). The output above uses the Windmeijer bias-corrected (WC) robust VCE, which Windmeijer (2005) showed to work well. The magnitudes of several of the coefficient estimates have changed, and one even switched its sign.
The test for autocorrelation presents no evidence of model misspecification:

```
estat abond
Arellano-Bond test for zero autocorrelation in first-differenced errors

<table>
<thead>
<tr>
<th>Order</th>
<th>z</th>
<th>Prob &gt; z</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-2.1255</td>
<td>0.0335</td>
</tr>
<tr>
<td>2</td>
<td>-0.35166</td>
<td>0.7251</td>
</tr>
</tbody>
</table>
```

H0: no autocorrelation

Manuel Arellano (1957– ) was born in Elda in Alicante, Spain. He earned degrees in economics from the University of Barcelona and the London School of Economics. After various posts in Oxford and London, he returned to Spain as professor of econometrics at Madrid in 1991. He is a leading expert on panel-data econometrics.

Stephen Roy Bond (1963– ) earned degrees in economics from Cambridge and Oxford. Following various posts at Oxford, he now works mainly at the Institute for Fiscal Studies in London. His research interests include company taxation, dividends, and the links between financial markets, corporate control, and investment.

Example 5: Including an estimator for the constant

Thus far we have been specifying the `noconstant` option to keep to the standard Arellano–Bond estimator, which uses instruments only for the differenced equation. The constant estimated by `xtabond` is a constant in the level equation, and it is estimated from the level errors. The output below illustrates that including a constant in the model does not affect the other parameter estimates.
. xtabond n l(0/1).w l(0/2).k ys yr1980-yr1984 year, lags(2) twostep vce(robust)

Arellano-Bond dynamic panel-data estimation

Number of obs = 611
Group variable: id
Number of groups = 140
Time variable: year

Obs per group:
  min = 4
  avg = 4.364286
  max = 6

Number of instruments = 42
Wald chi2(16) = 1104.72
Prob > chi2 = 0.0000

Two-step results

(Std. Err. adjusted for clustering on id)

<table>
<thead>
<tr>
<th></th>
<th>WC-Robust</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coef.</td>
<td>Std. Err.</td>
<td>z</td>
<td>P&gt;</td>
</tr>
<tr>
<td>----</td>
<td>-------</td>
<td>------------</td>
<td>------</td>
<td>------</td>
</tr>
<tr>
<td>n</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L1.</td>
<td>.6287089</td>
<td>.1934138</td>
<td>3.25</td>
<td>0.001</td>
</tr>
<tr>
<td>L2.</td>
<td>-.0651882</td>
<td>.0450501</td>
<td>-1.45</td>
<td>0.148</td>
</tr>
<tr>
<td>w</td>
<td>-.5257597</td>
<td>.1546107</td>
<td>-3.40</td>
<td>0.001</td>
</tr>
<tr>
<td>L1.</td>
<td>.3112899</td>
<td>.2030006</td>
<td>1.53</td>
<td>0.125</td>
</tr>
<tr>
<td>k</td>
<td>.2783619</td>
<td>.0728019</td>
<td>3.82</td>
<td>0.000</td>
</tr>
<tr>
<td>L1.</td>
<td>.0140994</td>
<td>.0924575</td>
<td>0.15</td>
<td>0.879</td>
</tr>
<tr>
<td>L2.</td>
<td>-.0402484</td>
<td>.0432745</td>
<td>-0.93</td>
<td>0.352</td>
</tr>
<tr>
<td>ys</td>
<td>.5919243</td>
<td>.1730916</td>
<td>3.42</td>
<td>0.001</td>
</tr>
<tr>
<td>L1.</td>
<td>-.5659863</td>
<td>.2611008</td>
<td>-2.17</td>
<td>0.030</td>
</tr>
<tr>
<td>L2.</td>
<td>.1005433</td>
<td>.1610987</td>
<td>0.62</td>
<td>0.533</td>
</tr>
<tr>
<td>yr1980</td>
<td>.0006378</td>
<td>.0168042</td>
<td>0.04</td>
<td>0.970</td>
</tr>
<tr>
<td>yr1981</td>
<td>-.0550044</td>
<td>.0313389</td>
<td>-1.76</td>
<td>0.079</td>
</tr>
<tr>
<td>yr1982</td>
<td>-.075978</td>
<td>.0419276</td>
<td>-1.81</td>
<td>0.070</td>
</tr>
<tr>
<td>yr1983</td>
<td>-.0740708</td>
<td>.0528381</td>
<td>-1.40</td>
<td>0.161</td>
</tr>
<tr>
<td>yr1984</td>
<td>-.0906606</td>
<td>.0642615</td>
<td>-1.41</td>
<td>0.158</td>
</tr>
<tr>
<td>year</td>
<td>.0112155</td>
<td>.016783</td>
<td>0.66</td>
<td>0.507</td>
</tr>
<tr>
<td>_cons</td>
<td>-21.53725</td>
<td>23.23138</td>
<td>-0.93</td>
<td>0.354</td>
</tr>
</tbody>
</table>

Instruments for differenced equation

GMM-type: L(2/).n

Instruments for level equation

Standard: _cons

Including the constant does not affect the other parameter estimates because it is identified only by the level errors; see [XT] xtdpd for details.

Example 6: Including predetermined covariates

Sometimes we cannot assume strict exogeneity. Recall that a variable, \( x_{it} \), is said to be strictly exogenous if \( E[x_{it}\epsilon_s] = 0 \) for all \( t \) and \( s \). If \( E[x_{it}\epsilon_s] \neq 0 \) for \( s < t \) but \( E[x_{it}\epsilon_s] = 0 \) for all \( s \geq t \), the variable is said to be predetermined. Intuitively, if the error term at time \( t \) has some feedback on the subsequent realizations of \( x_{it} \), \( x_{it} \) is a predetermined variable. Because unforecastable errors today might affect future changes in the real wage and in the capital stock, we might suspect that the log of the real product wage and the log of the gross capital stock are predetermined instead of strictly exogenous. Here we treat \( w \) and \( k \) as predetermined and use lagged levels as instruments.
. xtabond n 1(0/1).ys yr1980-yr1984 year, lags(2) twostep pre(w, lag(1,.))
> pre(k, lag(2, .)) noconstant vce(robust)

Arellano-Bond dynamic panel-data estimation  Number of obs =  611
Group variable: id  Number of groups = 140
Time variable: year
Obs per group:  min =  4
avg =  4.364286
max =  6
Number of instruments = 83  Wald chi2(15) = 958.30
Prob > chi2 =  0.0000

Two-step results
(Std. Err. adjusted for clustering on id)

<table>
<thead>
<tr>
<th></th>
<th>WC-Robust</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coef.</td>
<td>Std. Err.</td>
<td>z</td>
<td>P&gt;</td>
</tr>
<tr>
<td>n</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L1.</td>
<td>0.8580958</td>
<td>0.1265515</td>
<td>6.78</td>
<td>0.000</td>
</tr>
<tr>
<td>L2.</td>
<td>-0.081207</td>
<td>0.0760703</td>
<td>-1.07</td>
<td>0.286</td>
</tr>
<tr>
<td>w</td>
<td>-0.6910855</td>
<td>0.1387684</td>
<td>-4.98</td>
<td>0.000</td>
</tr>
<tr>
<td>L1.</td>
<td>0.5961712</td>
<td>0.1497338</td>
<td>3.98</td>
<td>0.000</td>
</tr>
<tr>
<td>k</td>
<td>-0.4140654</td>
<td>0.1382788</td>
<td>2.99</td>
<td>0.003</td>
</tr>
<tr>
<td>L1.</td>
<td>-0.1537048</td>
<td>0.1220244</td>
<td>-1.26</td>
<td>0.208</td>
</tr>
<tr>
<td>L2.</td>
<td>-0.1025833</td>
<td>0.0710886</td>
<td>-1.44</td>
<td>0.149</td>
</tr>
<tr>
<td>ys</td>
<td>-0.6936392</td>
<td>0.1728623</td>
<td>4.01</td>
<td>0.000</td>
</tr>
<tr>
<td>L1.</td>
<td>-0.8773678</td>
<td>0.2183085</td>
<td>-4.02</td>
<td>0.000</td>
</tr>
<tr>
<td>yr1980</td>
<td>-0.0072451</td>
<td>0.0171633</td>
<td>-0.42</td>
<td>0.673</td>
</tr>
<tr>
<td>yr1981</td>
<td>-0.0609608</td>
<td>0.0302072</td>
<td>-2.02</td>
<td>0.044</td>
</tr>
<tr>
<td>yr1982</td>
<td>-0.1130369</td>
<td>0.0454826</td>
<td>-2.49</td>
<td>0.013</td>
</tr>
<tr>
<td>yr1983</td>
<td>-0.1335249</td>
<td>0.0600213</td>
<td>-2.22</td>
<td>0.026</td>
</tr>
<tr>
<td>yr1984</td>
<td>-0.1623177</td>
<td>0.0725434</td>
<td>-2.24</td>
<td>0.025</td>
</tr>
</tbody>
</table>

The footer informs us that we are now including GMM-type instruments from the first lag of \L.w on back and from the first lag of \L2.k on back.

Technical note

The above example illustrates that xtabond understands `pre(w, lag(1, .))` to mean that \L.w is a predetermined variable and `pre(k, lag(2, .))` to mean that \L2.k is a predetermined variable. This is a stricter definition than the alternative that `pre(w, lag(1, .))` means only that \(w\) is predetermined but includes a lag of \(w\) in the model and that `pre(k, lag(2, .))` means only that \(k\) is predetermined but includes first and second lags of \(k\) in the model. If you prefer the weaker definition, xtabond still gives you consistent estimates, but it is not using all possible instruments; see [XT] xtdpd for an example of how to include all possible instruments.
Example 7: Including endogenous covariates

We might instead suspect that $w$ and $k$ are endogenous in that $E[x_{it}\epsilon_{is}] \neq 0$ for $s \leq t$ but $E[x_{it}\epsilon_{is}] = 0$ for all $s > t$. By this definition, endogenous variables differ from predetermined variables only in that the former allow for correlation between the $x_{it}$ and the $\epsilon_{it}$ at time $t$, whereas the latter do not. Endogenous variables are treated similarly to the lagged dependent variable. Levels of the endogenous variables lagged two or more periods can serve as instruments. In this example, we treat $w$ and $k$ as endogenous variables.

`. xtabond n l(0/1).ys yr1980-yr1984 year, lags(2) twostep
> endogenous(w, lag(1,.)) endogenous(k, lag(2,.)) noconstant vce(robust)
Arellano-Bond dynamic panel-data estimation Number of obs = 611
Group variable: id Number of groups = 140
Time variable: year
Obs per group: min = 4
avg = 4.364286
max = 6
Number of instruments = 71 Wald chi2(15) = 967.61
Prob > chi2 = 0.0000
Two-step results
(Std. Err. adjusted for clustering on id)

<table>
<thead>
<tr>
<th></th>
<th>WC-Robust</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coef. Std. Err. z P&gt;</td>
<td>z</td>
<td>[95% Conf. Interval]</td>
<td></td>
</tr>
<tr>
<td>n</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L1.</td>
<td>.6640937 .1278908 5.19 0.000 .4134323 .914755</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L2.</td>
<td>-.041283 .081801 -0.50 0.614 -.2016101 .1190441</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>w</td>
<td>-.7143942 .13083 -5.46 0.000 -.9708162 -.4579721</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>k</td>
<td>-.3644198 .184758 1.97 0.049 .0023008 .7265388</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ys</td>
<td>-.5028874 .1205419 4.17 0.000 .2666296 .7391452</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>yr1980</td>
<td>-.2160842 .0972855 -2.22 0.026 -.4067603 -.025408</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>yr1981</td>
<td>-.0549654 .0793673 -0.69 0.489 -.2105225 .1005917</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>yr1982</td>
<td>-.5989356 .1779731 3.37 0.001 .2501148 .9477564</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>yr1983</td>
<td>-.6770367 .1961166 -3.45 0.001 -.1061418 -.2926553</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>yr1984</td>
<td>-.0061122 .0155287 -0.39 0.694 -.0365478 .0243235</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>yr1980</td>
<td>-.04715 .0298348 -1.58 0.114 -.1056252 .0113251</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>yr1981</td>
<td>-.0817646 .0486049 -1.68 0.093 -.1770285 .0134993</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>yr1982</td>
<td>-.0939251 .0675804 -1.39 0.165 -.2263802 .0085299</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>yr1983</td>
<td>-.117228 .0804716 -1.46 0.145 -.2749493 .0404934</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>yr1984</td>
<td>.0208857 .0103485 2.02 0.044 .0006031 .0411684</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Instruments for differenced equation
GMM-type: L(2/.) n L(2/.) L.w L(2/.) L2.k

Although some estimated coefficients changed in magnitude, none changed in sign, and these results are similar to those obtained by treating $w$ and $k$ as predetermined.

The Arellano–Bond estimator is for datasets with many panels and few periods. (Technically, the large-sample properties are derived with the number of panels going to infinity and the number of
The number of instruments increases quadratically in the number of periods. If your dataset is better described by a framework in which both the number of panels and the number of periods is large, then you should consider other estimators such as those in [XT] xtivreg or xtreg, fe in [XT] xtreg; see Alvarez and Arellano (2003) for a discussion of this case.

### Example 8: Restricting the number of instruments

Treating variables as predetermined or endogenous quickly increases the size of the instrument matrix. (See Methods and formulas in [XT] xtdpd for a discussion of how this matrix is created and what determines its size.) GMM estimators with too many overidentifying restrictions may perform poorly in small samples. (See Kiviet 1995 for a discussion of the dynamic panel-data case.)

To handle these problems, you can set a maximum number of lagged levels to be included as instruments for lagged-dependent or the predetermined variables. Here is an example in which a maximum of three lagged levels of the predetermined variables are included as instruments:

```
.xtabond n l(0/1).ys yr1980-yr1984 year, lags(2) twostep
> pre(w, lag(1,3)) pre(k, lag(2,3)) noconstant vce(robust)
```

```
Arellano-Bond dynamic panel-data estimation
Number of obs = 611
Group variable: id Number of groups = 140
Time variable: year
Obs per group: min = 4
avg = 4.364286
max = 6
Number of instruments = 67
Wald chi2(15) = 1116.89
Prob > chi2 = 0.0000
Two-step results
(Std. Err. adjusted for clustering on id)

<table>
<thead>
<tr>
<th></th>
<th>WC-Robust</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coef.</td>
<td>Std. Err.</td>
<td>z</td>
<td>P&gt;</td>
</tr>
<tr>
<td>-------</td>
<td>-------------------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
</tr>
<tr>
<td>n</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L1.</td>
<td>.931121</td>
<td>.1456964</td>
<td>6.39</td>
<td>0.000</td>
</tr>
<tr>
<td>L2.</td>
<td>-.0759918</td>
<td>.0854356</td>
<td>-0.89</td>
<td>0.374</td>
</tr>
<tr>
<td>w</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pr.</td>
<td>-.6475372</td>
<td>.1687931</td>
<td>-3.84</td>
<td>0.000</td>
</tr>
<tr>
<td>L1.</td>
<td>.6906238</td>
<td>.1789698</td>
<td>3.86</td>
<td>0.000</td>
</tr>
<tr>
<td>k</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pr.</td>
<td>.3788106</td>
<td>.1848137</td>
<td>2.05</td>
<td>0.040</td>
</tr>
<tr>
<td>L1.</td>
<td>-.2158533</td>
<td>.1446198</td>
<td>-1.49</td>
<td>0.136</td>
</tr>
<tr>
<td>L2.</td>
<td>-.0914584</td>
<td>.0852267</td>
<td>-1.07</td>
<td>0.283</td>
</tr>
<tr>
<td>ys</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pr.</td>
<td>.7324964</td>
<td>.176748</td>
<td>4.14</td>
<td>0.000</td>
</tr>
<tr>
<td>L1.</td>
<td>-.9428141</td>
<td>.2735472</td>
<td>-3.45</td>
<td>0.001</td>
</tr>
<tr>
<td>yr1980</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pr.</td>
<td>-.0102389</td>
<td>.0172473</td>
<td>-0.59</td>
<td>0.553</td>
</tr>
<tr>
<td>yr1981</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pr.</td>
<td>-.0763495</td>
<td>.0296992</td>
<td>-2.57</td>
<td>0.010</td>
</tr>
<tr>
<td>yr1982</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pr.</td>
<td>-.1373829</td>
<td>.0441833</td>
<td>-3.11</td>
<td>0.002</td>
</tr>
<tr>
<td>yr1983</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pr.</td>
<td>-.1826149</td>
<td>.0613674</td>
<td>-2.97</td>
<td>0.003</td>
</tr>
<tr>
<td>yr1984</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pr.</td>
<td>-.2314023</td>
<td>.0753669</td>
<td>-3.07</td>
<td>0.002</td>
</tr>
<tr>
<td>year</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pr.</td>
<td>.0310012</td>
<td>.0119167</td>
<td>2.60</td>
<td>0.009</td>
</tr>
</tbody>
</table>

Instruments for differenced equation
GMM-type: L(2/). n L(1/3). L.w L(1/3). L2.k
Example 9: Missing observations in the middle of panels

*xtabond* handles data in which there are missing observations in the middle of the panels. In this example, we deliberately set the dependent variable to missing in the year 1980:

```
. replace n=. if year==1980
   (140 real changes made, 140 to missing)
. xtabond n l(0/1).w l(0/2).(k ys) yr1980-yr1984 year, lags(2) noconstant
   vce(robust)
```

Arellano-Bond dynamic panel-data estimation

Number of obs = 115
Group variable: id
Number of groups = 101
Time variable: year

Obs per group: min = 1
    avg = 1.138614
    max = 2

Number of instruments = 18
Wald chi2(12) = 44.48
Prob > chi2 = 0.0000

One-step results

(Std. Err. adjusted for clustering on id)

|        | Coef.  | Std. Err. | z       | P>|z|  | [95% Conf. Interval] |
|--------|--------|-----------|---------|------|---------------------|
| n      |        |           |         |      |                     |
| L1.    | .1790577 | .2204682  | 0.81    | 0.417| -.253052  .6111674  |
| L2.    | .0214253 | .0488476  | 0.44    | 0.661| -.0743143 .1171649 |
| w      | -.2513405 | .1402114  | -1.79   | 0.073| -.5261498 .0234689 |
| L1.    | .1983952 | .1445875  | 1.37    | 0.170| -.0849912 .4817815 |
| k      | -.3983149 | .0883352  | 4.51    | 0.000| .2251811  .5714488 |
| L1.    | -.025125 | .0909236  | -0.28   | 0.782| -.203332  .1530821 |
| L2.    | -.0395338 | .0623382  | -0.58   | 0.564| -.1581144 .0862468 |
| ys     | -.3663201 | .3824893  | 0.96    | 0.338| -.3833451 1.115985  |
| L1.    | -.6319976 | .4823958  | -1.31   | 0.190| -.577476  .3134807 |
| L2.    | .5318404 | .4105269  | 1.30    | 0.195| -.2727775 1.336458 |
| yr1983 | -.0047543 | .024855   | -0.19   | 0.848| -.0534692 .0439606 |
| yr1984 | 0 (omitted) |          |         |      |                     |
| year   | .0014465 | .010355   | 0.14    | 0.889| -.0188489 .0217419 |

Instruments for differenced equation

GMM-type: L(2/).n

There are two important aspects to this example. First, *xtabond* reports that variables have been dropped from the model and from the *div()* instrument list. For *xtabond*, the *div()* instrument list is the list of instruments created from the strictly exogenous variables; see [XT] *xtdpd* for more about the *div()* instrument list. Second, because *xtabond* uses time-series operators in its computations, *if* statements and missing values are not equivalent. An *if* statement causes the false observations to
be excluded from the sample, but it computes the time-series operators wherever possible. In contrast, missing data prevent evaluation of the time-series operators that involve missing observations. Thus the example above is not equivalent to the following one:

```
. use http://www.stata-press.com/data/r13/abdata, clear
. xtabond n l(0/1).w l(0/2).(k ys) yr1980-yr1984 year if year!=1980,
   > lags(2) noconstant vce(robust)
```

```
note: yr1980 dropped from div() because of collinearity
```

```
ote: yr1980 dropped because of collinearity
```

Arellano-Bond dynamic panel-data estimation
Number of obs = 473
Group variable: id Number of groups = 140
Time variable: year
Obs per group:
  min = 3
  avg = 3.378571
  max = 5

Number of instruments = 37

Wald chi2(15) = 1041.61
Prob > chi2 = 0.0000

One-step results
(Std. Err. adjusted for clustering on id)

<table>
<thead>
<tr>
<th></th>
<th>Robust</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coef.</td>
<td>Std. Err.</td>
<td>z</td>
<td>P&gt;</td>
<td>z</td>
</tr>
<tr>
<td>n</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L1.</td>
<td>.7210062</td>
<td>.1321214</td>
<td>5.46</td>
<td>0.000</td>
<td>.4620531</td>
</tr>
<tr>
<td>L2.</td>
<td>-.0960646</td>
<td>.0570547</td>
<td>-1.68</td>
<td>0.092</td>
<td>-.2078898</td>
</tr>
<tr>
<td>w</td>
<td>-.6684175</td>
<td>.1739484</td>
<td>-3.84</td>
<td>0.000</td>
<td>-1.00935</td>
</tr>
<tr>
<td>L1.</td>
<td>.482322</td>
<td>.1647185</td>
<td>2.93</td>
<td>0.003</td>
<td>.1594797</td>
</tr>
<tr>
<td>k</td>
<td>.3802777</td>
<td>.0728546</td>
<td>5.22</td>
<td>0.000</td>
<td>.2374853</td>
</tr>
<tr>
<td>L1.</td>
<td>-.104598</td>
<td>.088597</td>
<td>-1.18</td>
<td>0.238</td>
<td>-.278245</td>
</tr>
<tr>
<td>L2.</td>
<td>-.0272055</td>
<td>.0379994</td>
<td>-0.72</td>
<td>0.474</td>
<td>-.101683</td>
</tr>
<tr>
<td>ys</td>
<td>.4655989</td>
<td>.1864368</td>
<td>2.50</td>
<td>0.013</td>
<td>.1001895</td>
</tr>
<tr>
<td>L1.</td>
<td>-.8562492</td>
<td>.2187886</td>
<td>-3.91</td>
<td>0.000</td>
<td>-1.285067</td>
</tr>
<tr>
<td>L2.</td>
<td>.0896556</td>
<td>.1440035</td>
<td>0.62</td>
<td>0.534</td>
<td>-.192586</td>
</tr>
<tr>
<td>yr1981</td>
<td>-.0711626</td>
<td>.0205299</td>
<td>-3.47</td>
<td>0.001</td>
<td>-.1114005</td>
</tr>
<tr>
<td>yr1982</td>
<td>-.1212749</td>
<td>.0334659</td>
<td>-3.62</td>
<td>0.000</td>
<td>-.1868669</td>
</tr>
<tr>
<td>yr1983</td>
<td>-.1470248</td>
<td>.0461714</td>
<td>-3.18</td>
<td>0.001</td>
<td>-.2375191</td>
</tr>
<tr>
<td>yr1984</td>
<td>-.1519021</td>
<td>.0543904</td>
<td>-2.79</td>
<td>0.005</td>
<td>-.2585054</td>
</tr>
<tr>
<td>year</td>
<td>.0203277</td>
<td>.0108732</td>
<td>1.87</td>
<td>0.062</td>
<td>-.0009833</td>
</tr>
</tbody>
</table>

Instruments for differenced equation
GMM-type: L(2/.)\_n
Standard: D.w LD.w D.k LD.k D.ys LD ys L2.ys D.yr1981
D.yr1982 D.yr1983 D.yr1984 D.year

The year 1980 is dropped from the sample, but when the value of a variable from 1980 is required because a lag or difference is required, the 1980 value is used.
xtabond stores the following in e():

Scalars
- e(N): number of observations
- e(N_g): number of groups
- e(df_m): model degrees of freedom
- e(g_min): smallest group size
- e(g_avg): average group size
- e(g_max): largest group size
- e(t_min): minimum time in sample
- e(t_max): maximum time in sample
- e(chi2): \( \chi^2 \)
- e(arm#): test for autocorrelation of order \#\n- e(artests): number of AR tests computed
- e(sig2): estimate of \( \sigma^2 \)
- e(rss): sum of squared differenced residuals
- e(sargan): Sargan test statistic
- e(rank): rank of e(V)
- e(zrank): rank of instrument matrix

Macros
- e(cmd): xtabond
- e(cmdline): command as typed
- e(depvar): name of dependent variable
- e(twostep): twostep, if specified
- e(ivar): variable denoting groups
- e(tvar): variable denoting time within groups
- e(vce): vcetype specified in vce()
- e(vcetype): title used to label Std. Err.
- e(system): system, if system estimator
- e(hascons): hascons, if specified
- e(transform): specified transform
- e(diffvars): already differenced variables
- e(datasignature): checksum from datasignature
- e(properties): b V
- e(estat_cmd): program used to implement estat
- e(predict): program used to implement predict
- e(marginsok): predictions allowed by margins

Matrices
- e(b): coefficient vector
- e(V): variance–covariance matrix of the estimators

Functions
- e(sample): marks estimation sample

### Methods and formulas

A dynamic panel-data model has the form

\[
y_{it} = \sum_{j=1}^{p} \alpha_{j} y_{i,t-j} + x_{it} \beta_1 + w_{it} \beta_2 + \nu_i + \epsilon_{it} \quad i = 1, \ldots, N \quad t = 1, \ldots, T_i
\]  

(1)
where

\[
\begin{align*}
\alpha_j & \text{ are } p \text{ parameters to be estimated,} \\
x_{it} & \text{ is a } 1 \times k_1 \text{ vector of strictly exogenous covariates,} \\
\beta_1 & \text{ is a } k_1 \times 1 \text{ vector of parameters to be estimated,} \\
w_{it} & \text{ is a } 1 \times k_2 \text{ vector of predetermined and endogenous covariates,} \\
\beta_2 & \text{ is a } k_2 \times 1 \text{ vector of parameters to be estimated,} \\
\nu_i & \text{ are the panel-level effects (which may be correlated with the covariates), and} \\
\epsilon_{it} & \text{ are i.i.d. over the whole sample with variance } \sigma^2_{\epsilon}.
\end{align*}
\]

The \( \nu_i \) and the \( \epsilon_{it} \) are assumed to be independent for each \( i \) over all \( t \).

By construction, the lagged dependent variables are correlated with the unobserved panel-level effects, making standard estimators inconsistent. With many panels and few periods, estimators are constructed by first-differencing to remove the panel-level effects and using instruments to form moment conditions.

\texttt{xtabond} uses a GMM estimator to estimate \( \alpha_1, \ldots, \alpha_p, \beta_1, \) and \( \beta_2 \). The moment conditions are formed from the first-differenced errors from (1) and instruments. Lagged levels of the dependent variable, the predetermined variables, and the endogenous variables are used to form GMM-type instruments. See Arellano and Bond (1991) and Holtz-Eakin, Newey, and Rosen (1988) for discussions of GMM-type instruments. First differences of the strictly exogenous variables are used as standard instruments.

\texttt{xtabond} uses \texttt{xtdpd} to perform its computations, so the formulas are given in \textit{Methods and formulas} of \texttt{[XT] xtdpd}.

References


Also see

[XT] xtabond postestimation — Postestimation tools for xtabond
[XT] xtset — Declare data to be panel data
[XT] xtdpd — Linear dynamic panel-data estimation
[XT] xtdpdsys — Arellano–Bover/Blundell–Bond linear dynamic panel-data estimation
[XT] xtivreg — Instrumental variables and two-stage least squares for panel-data models
[XT] xtreg — Fixed-, between-, and random-effects and population-averaged linear models
[XT] xtregar — Fixed- and random-effects linear models with an AR(1) disturbance

[U] 20 Estimation and postestimation commands
The following postestimation commands are of special interest after `xtabond`:

<table>
<thead>
<tr>
<th>Command</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>estat abond</code></td>
<td>test for autocorrelation</td>
</tr>
<tr>
<td><code>estat sargan</code></td>
<td>Sargan test of overidentifying restrictions</td>
</tr>
</tbody>
</table>

The following standard postestimation commands are also available:

<table>
<thead>
<tr>
<th>Command</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>estat summarize</code></td>
<td>summary statistics for the estimation sample</td>
</tr>
<tr>
<td><code>estat vce</code></td>
<td>variance–covariance matrix of the estimators (VCE)</td>
</tr>
<tr>
<td><code>estimates</code></td>
<td>cataloging estimation results</td>
</tr>
<tr>
<td><code>forecast</code></td>
<td>dynamic forecasts and simulations</td>
</tr>
<tr>
<td><code>lincom</code></td>
<td>point estimates, standard errors, testing, and inference for linear combinations of coefficients</td>
</tr>
<tr>
<td><code>margins</code></td>
<td>marginal means, predictive margins, marginal effects, and average marginal effects</td>
</tr>
<tr>
<td><code>marginsplot</code></td>
<td>graph the results from margins (profile plots, interaction plots, etc.)</td>
</tr>
<tr>
<td><code>nlcom</code></td>
<td>point estimates, standard errors, testing, and inference for nonlinear combinations of coefficients</td>
</tr>
<tr>
<td><code>predict</code></td>
<td>predictions, residuals, influence statistics, and other diagnostic measures</td>
</tr>
<tr>
<td><code>predictnl</code></td>
<td>point estimates, standard errors, testing, and inference for generalized predictions</td>
</tr>
<tr>
<td><code>test</code></td>
<td>Wald tests of simple and composite linear hypotheses</td>
</tr>
<tr>
<td><code>testnl</code></td>
<td>Wald tests of nonlinear hypotheses</td>
</tr>
</tbody>
</table>

**Special-interest postestimation commands**

- `estat abond` reports the Arellano–Bond tests for serial correlation in the first-differenced errors.
- `estat sargan` reports the Sargan test of the overidentifying restrictions.
Syntax for predict

predict [type] newvar [if] [in] [, xb e stdp difference]

Menu for predict

Statistics > Postestimation > Predictions, residuals, etc.

Options for predict

- **xb**: the default, calculates the linear prediction.
- **e**: calculates the residual error.
- **stdp**: calculates the standard error of the prediction, which can be thought of as the standard error of the predicted expected value or mean for the observation’s covariate pattern. The standard error of the prediction is also referred to as the standard error of the fitted value. **stdp** may not be combined with **difference**.
- **difference**: specifies that the statistic be calculated for the first differences instead of the levels, the default.

Syntax for estat

**Test for autocorrelation**

estat abond [, artests(#)]

**Sargan test of overidentifying restrictions**

estat sargan

Menu for estat

Statistics > Postestimation > Reports and statistics

Option for estat abond

- **artests(#)** specifies the highest order of serial correlation to be tested. By default, the tests computed during estimation are reported. The model will be refit when **artests(#)** specifies a higher order than that computed during the original estimation. The model can be refit only if the data have not changed.

Remarks and examples

Remarks are presented under the following headings:

- estat abond
- estat sargan
estat abond

`estat abond` reports the Arellano–Bond test for serial correlation in the first-differenced errors at order $m$. Rejecting the null hypothesis of no serial correlation in the first-differenced errors at order zero does not imply model misspecification because the first-differenced errors are serially correlated if the idiosyncratic errors are independent and identically distributed. Rejecting the null hypothesis of no serial correlation in the first-differenced errors at an order greater than one implies model misspecification; see example 5 in `[XT] xtdpd` for an alternative estimator that allows for idiosyncratic errors that follow a first-order moving average process.

After the one-step system estimator, the test can be computed only when `vce(robust)` has been specified. (The system estimator is used to estimate the constant in `xtabond`.)

See `Remarks and examples` in `[XT] xtabond` for more remarks about `estat abond` that are made in the context of the examples analyzed therein.

estat sargan

The distribution of the Sargan test is known only when the errors are independently and identically distributed. For this reason, `estat sargan` does not produce a test statistic when `vce(robust)` was specified in the call to `xtabond`.

See `Remarks and examples` in `[XT] xtabond` for more remarks about `estat sargan` that are made in the context of the examples analyzed therein.

Methods and formulas

See `[XT] xtdpd postestimation` for the formulas.

Also see

[[XT] xtabond] — Arellano–Bond linear dynamic panel-data estimation

[[U] 20 Estimation and postestimation commands]
**xtcloglog — Random-effects and population-averaged cloglog models**

### Syntax

**Random-effects (RE) model**

\[
\text{xtcloglog depvar \{indepvars\} \{if\} \{in\} \{weight\} \{, re \ RE\_options\}}
\]

**Population-averaged (PA) model**

\[
\text{xtcloglog depvar \{indepvars\} \{if\} \{in\} \{weight\}, pa \{PA\_options\}}
\]

<table>
<thead>
<tr>
<th><strong>RE_options</strong></th>
<th><strong>Description</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Model</strong></td>
<td></td>
</tr>
<tr>
<td><code>noconstant</code></td>
<td>suppress constant term</td>
</tr>
<tr>
<td><code>re</code></td>
<td>use random-effects estimator; the default</td>
</tr>
<tr>
<td><code>offset(varname)</code></td>
<td>include <code>varname</code> in model with coefficient constrained to 1</td>
</tr>
<tr>
<td><code>constraints(constraints)</code></td>
<td>apply specified linear constraints</td>
</tr>
<tr>
<td><code>collinear</code></td>
<td>keep collinear variables</td>
</tr>
<tr>
<td><code>asis</code></td>
<td>retain perfect predictor variables</td>
</tr>
</tbody>
</table>

| **SE/Robust**    |                  |
| `vce(vcetype)`  | `vcetype` may be oim, robust, \texttt{cluster clustvar}, bootstrap, or jackknife |

| **Reporting**    |                  |
| `level(#)`      | set confidence level; default is `level(95)` |
| `noskip`        | perform overall model test as a likelihood-ratio test |
| `eform`         | report exponentiated coefficients |
| `noconsreport`  | do not display constraints |
| `display_options` | control column formats, row spacing, line width, display of omitted variables and base and empty cells, and factor-variable labeling |

| **Integration**  |                  |
| `intmethod(intmethod)` | integration method; `intmethod` may be \texttt{mvaghermite} (the default) or \texttt{ghermite} |
| `intpoints(#)`   | use \# quadrature points; default is `intpoints(12)` |

| **Maximization** |                  |
| `maximize\_options` | control the maximization process; seldom used |
| `coeflegend`      | display legend instead of statistics |
**xtcloglog** — Random-effects and population-averaged cloglog models

### PA options

<table>
<thead>
<tr>
<th>Model</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>noconstant</code></td>
<td>suppress constant term</td>
</tr>
<tr>
<td><code>pa</code></td>
<td>use population-averaged estimator</td>
</tr>
<tr>
<td><code>offset(varname)</code></td>
<td>include <code>varname</code> in model with coefficient constrained to 1</td>
</tr>
<tr>
<td><code>asis</code></td>
<td>retain perfect predictor variables</td>
</tr>
</tbody>
</table>

### Correlation

<table>
<thead>
<tr>
<th>Correlation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>corr(correlation)</code></td>
<td>within-panel correlation structure</td>
</tr>
<tr>
<td><code>force</code></td>
<td>estimate even if observations unequally spaced in time</td>
</tr>
</tbody>
</table>

### SE/Robust

<table>
<thead>
<tr>
<th>SE/Robust</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>vce(vcetype)</code></td>
<td><code>vcetype</code> may be <code>conventional</code>, <code>robust</code>, <code>bootstrap</code>, or <code>jackknife</code></td>
</tr>
<tr>
<td><code>nmp</code></td>
<td>use divisor (N - P) instead of the default (N)</td>
</tr>
<tr>
<td><code>scale(parm)</code></td>
<td>overrides the default scale parameter; <code>parm</code> may be <code>x2</code>, <code>dev</code>, <code>phi</code>, or <code>#</code></td>
</tr>
</tbody>
</table>

### Reporting

<table>
<thead>
<tr>
<th>Reporting</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>level(#)</code></td>
<td>set confidence level; default is <code>level(95)</code></td>
</tr>
<tr>
<td><code>eform</code></td>
<td>report exponentiated coefficients</td>
</tr>
<tr>
<td><code>display_options</code></td>
<td>control column formats, row spacing, line width, display of omitted variables and base and empty cells, and factor-variable labeling</td>
</tr>
</tbody>
</table>

### Optimization

<table>
<thead>
<tr>
<th>Optimization</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>optimize_options</code></td>
<td>control the optimization process; seldom used</td>
</tr>
<tr>
<td><code>coeflegend</code></td>
<td>display legend instead of statistics</td>
</tr>
</tbody>
</table>

### correlation

<table>
<thead>
<tr>
<th>correlation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>exchangeable</code></td>
<td>exchangeable; the default</td>
</tr>
<tr>
<td><code>independent</code></td>
<td>independent</td>
</tr>
<tr>
<td><code>unstructured</code></td>
<td>unstructured</td>
</tr>
<tr>
<td><code>fixed matname</code></td>
<td>user-specified</td>
</tr>
<tr>
<td><code>ar #</code></td>
<td>autoregressive of order <code>#</code></td>
</tr>
<tr>
<td><code>stationary #</code></td>
<td>stationary of order <code>#</code></td>
</tr>
<tr>
<td><code>nonstationary #</code></td>
<td>nonstationary of order <code>#</code></td>
</tr>
</tbody>
</table>

A panel variable must be specified. For `xtcloglog`, `pa`, correlation structures other than `exchangeable` and `independent` require that a time variable also be specified. Use `xtset`; see [XT] `xtset`.

`indepv`ars may contain factor variables; see [U] 11.4.3 Factor variables.

by, `mi estimate`, and `statsby` are allowed; see [U] 11.1.10 Prefix commands. `fp` is allowed for the random-effects model.

`vce(bootstrap)` and `vce(jackknife)` are not allowed with the `mi estimate` prefix; see [MI] `mi estimate`.

`iweights`, `fweights`, and `pweights` are allowed for the population-averaged model, and `iweights` are allowed for the random-effects model; see [U] 11.1.6 weight. Weights must be constant within panel.

`coeflegend` does not appear in the dialog box.

See [U] 20 Estimation and postestimation commands for more capabilities of estimation commands.
xtcloglog fits population-averaged and random-effects complementary log-log (cloglog) models. There is no command for a conditional fixed-effects model, as there does not exist a sufficient statistic allowing the fixed effects to be conditioned out of the likelihood. Unconditional fixed-effects cloglog models may be fit with cloglog with indicator variables for the panels. However, unconditional fixed-effects estimates are biased.

By default, the population-averaged model is an equal-correlation model; that is, xtcloglog, pa assumes corr(exchangeable). See [XT] xtgee for information on fitting other population-averaged models.

See [R] logistic for a list of related estimation commands.

Options for RE model

- **noconstant;** see [R] estimation options.
- **re** requests the random-effects estimator, which is the default.
- **offset(varname), constraints(constraints), collinear;** see [R] estimation options.
- **asis** forces retention of perfect predictor variables and their associated, perfectly predicted observations and may produce instabilities in maximization; see [R] probit.

**SE/Robust**
- **vce(vcetype)** specifies the type of standard error reported, which includes types that are derived from asymptotic theory (oim), that are robust to some kinds of misspecification (robust), that allow for intragroup correlation (cluster clustvar), and that use bootstrap or jackknife methods (bootstrap, jackknife); see [XT] vce_options.

Specifying vce(robust) is equivalent to specifying vce(cluster panelvar); see xtcloglog, re and the robust VCE estimator in Methods and formulas.

**Reporting**
- **level(#), noskip;** see [R] estimation options.
- **eform** displays the exponentiated coefficients and corresponding standard errors and confidence intervals.
- **nocnsreport;** see [R] estimation options.
- **display_options: noomitted, vsquish, noemptycells, baselevels, allbaselevels, nofvlabel, fvwrap(#), fvwrapon(style), cformat(%fmt), pformat(%fmt), sformat(%fmt), and nolstretch;** see [R] estimation options.

**Integration**
- **intmethod(intmethod), intpoints(#);** see [R] estimation options.
Maximization

`maximize_options: difficult, technique(algorithm_spec), iterate(#), [no]log, trace, gradient, showstep, hessian, showtolerance, tolerance(#), ltolerance(#), nrtolerance(#), nonrtolerance, and from(init_specs);` see [R] `maximize`. These options are seldom used.

The following option is available with xtcloglog but is not shown in the dialog box:
`coeflegend;` see [R] `estimation options`.

**Options for PA model**

**Model**

`nocodeconstant;` see [R] `estimation options`.

`pa` requests the population-averaged estimator.

`offset(varname);` see [R] `estimation options`

`asis` forces retention of perfect predictor variables and their associated, perfectly predicted observations and may produce instabilities in maximization; see [R] `probit`.

**Correlation**

`corr(correlation)` specifies the within-panel correlation structure; the default corresponds to the equal-correlation model, `corr(exchangeable)`.

When you specify a correlation structure that requires a lag, you indicate the lag after the structure’s name with or without a blank; for example, `corr(ar 1)` or `corr(ar1)`.

If you specify the fixed correlation structure, you specify the name of the matrix containing the assumed correlations following the word `fixed`, for example, `corr(fixed myr)`.

`force` specifies that estimation be forced even though the time variable is not equally spaced.

This is relevant only for correlation structures that require knowledge of the time variable. These correlation structures require that observations be equally spaced so that calculations based on lags correspond to a constant time change. If you specify a time variable indicating that observations are not equally spaced, the (time dependent) model will not be fit. If you also specify `force`, the model will be fit, and it will be assumed that the lags based on the data ordered by the time variable are appropriate.

**SE/Robust**

`vce(vcetype)` specifies the type of standard error reported, which includes types that are derived from asymptotic theory (conventional), that are robust to some kinds of misspecification (robust), and that use bootstrap or jackknife methods (bootstrap, jackknife); see [XT] `vce_options`.

`vce(conventional),` the default, uses the conventionally derived variance estimator for generalized least-squares regression.

`nomp, scale(x2|dev|phi|#);` see [XT] `vce_options`.

**Reporting**

`level(#)`; see [R] `estimation options`.

`eform` displays the exponentiated coefficients and corresponding standard errors and confidence intervals.
display_options: noomitted, vsquish, noemptycells, baselevels, allbaselevels, nofvlabel, fvwrap(#), fvwrapon(style), cformat(%,fmt), pformat(%,fmt), sformat(%,fmt), and nolstretch; see [R] estimation options.

Optimization

Optimization options control the iterative optimization process. These options are seldom used.

iterate(#) specifies the maximum number of iterations. When the number of iterations equals #, the optimization stops and presents the current results, even if convergence has not been reached. The default is iterate(100).

tolerance(#) specifies the tolerance for the coefficient vector. When the relative change in the coefficient vector from one iteration to the next is less than or equal to #, the optimization process is stopped. tolerance(1e-6) is the default.

nolog suppresses display of the iteration log.

trace specifies that the current estimates be printed at each iteration.

The following option is available with xtcloglog but is not shown in the dialog box: coeflegend; see [R] estimation options.

Remarks and examples

xtcloglog, pa is a shortcut command for fitting the population-averaged model. Typing

    . xtcloglog ..., pa ... 

is equivalent to typing

    . xtgee ..., ... family(binomial) link(cloglog) corr(exchangeable) 

Also see [XT] xtgee for information about xtcloglog.

By default or when re is specified, xtcloglog fits, via maximum likelihood, the random-effects model

$$\Pr(y_{it} \neq 0| x_{it}) = P(x_{it}\beta + \nu_i)$$

for \(i = 1, \ldots, n\) panels, where \(t = 1, \ldots, n_i\), \(\nu_i\) are i.i.d., \(N(0, \sigma^2_{\nu})\), and \(P(z) = 1 - \exp\{-\exp(z)\}\).

Underlying this model is the variance-components model

$$y_{it} \neq 0 \iff x_{it}\beta + \nu_i + \epsilon_{it} > 0$$

where \(\epsilon_{it}\) are i.i.d. extreme-value (Gumbel) distributed with the mean equal to Euler’s constant and variance \(\sigma^2_{\epsilon} = \pi^2/6\), independently of \(\nu_i\). The nonsymmetric error distribution is an alternative to logit and probit analysis and is typically used when the positive (or negative) outcome is rare.
Example 1

Suppose that we are studying unionization of women in the United States and are using the union dataset; see [XT] xt. We wish to fit a random-effects model of union membership:

```
. use http://www.stata-press.com/data/r13/union
(NLS Women 14–24 in 1968)
. xtcloglog union age grade not_smsa south##c.year
(output omitted)
```

<table>
<thead>
<tr>
<th>Random-effects complementary log-log model</th>
<th>Number of obs</th>
<th>26200</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group variable: idcode</td>
<td>Number of groups</td>
<td>4434</td>
</tr>
<tr>
<td>Random effects u_i ~ Gaussian</td>
<td>Obs per group: min</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>avg</td>
<td>5.9</td>
</tr>
<tr>
<td></td>
<td>max</td>
<td>12</td>
</tr>
<tr>
<td>Integration method: mvaghermite</td>
<td>Integration points</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>Wald chi2(6)</td>
<td>248.58</td>
</tr>
<tr>
<td>Log likelihood</td>
<td>Prob &gt; chi2</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

| union | Coef. | Std. Err. | z      | P>|z|    | [95% Conf. Interval] |
|-------|-------|-----------|--------|--------|----------------------|
| age   | .0128659 | .0119004  | 1.08   | 0.280  | -.0104586-.0361903   |
| grade | .06985 | .0138135  | 5.06   | 0.000  | .042776-.096924      |
| not_smsa | -.198416 | .0647943 | -3.06 | 0.002  | -.3254104-.0714215   |
| 1.south | -2.047645 | .488965  | -4.19  | 0.000  | -3.005999-1.089291   |
| year   | -.0006432 | .0123569  | -0.05  | 0.958  | -.0248623-.0235759   |
| south#c.year 1 | .0164259 | .006065  | 2.71   | 0.007  | .0045387-.0283132    |
| _cons  | -3.269158 | .659029  | -4.96  | 0.000  | -4.560831-1.977485   |

| /lnsig2u | 1.24128 | .0461705 | 1.150787 | 1.331772 |
| sigma_u  | 1.860118 | .0429413 | 1.77783  | 1.946214 |
| rho      | .677778 | .0100834 | .6577057 | .6972152 |

Likelihood-ratio test of rho=0: chibar2(01) = 6009.36 Prob >= chibar2 = 0.000

The output includes the additional panel-level variance component, which is parameterized as the log of the standard deviation, \( \ln \sigma_\nu \) (labeled \( \text{lnsig2u} \) in the output). The standard deviation \( \sigma_\nu \) is also included in the output, labeled \( \text{sigma}_u \), together with \( \rho \) (labeled \( \text{rho} \)),

\[
\rho = \frac{\sigma_\nu^2}{\sigma_\nu^2 + \sigma_\epsilon^2}
\]

which is the proportion of the total variance contributed by the panel-level variance component.

When \( \rho \) is zero, the panel-level variance component is not important, and the panel estimator is no different from the pooled estimator (\( \text{cloglog} \)). A likelihood-ratio test of this is included at the bottom of the output, which formally compares the pooled estimator with the panel estimator.
As an alternative to the random-effects specification, you might want to fit an equal-correlation population-averaged cloglog model by typing

```
.xtcloglog union age grade not_smsa south##c.year, pa
```

Iteration 1: tolerance = .11878399
Iteration 2: tolerance = .01424628
Iteration 3: tolerance = .00075278
Iteration 4: tolerance = .00003195
Iteration 5: tolerance = 1.661e-06
Iteration 6: tolerance = 8.308e-08

GEE population-averaged model

|                | Coef. | Std. Err. | z   | P>|z| | [95% Conf. Interval] |
|----------------|-------|-----------|-----|-----|---------------------|
| age            | .0153737 | .0081156 | 1.89| 0.058| -.0005326 -.03128  |
| grade          | .0549518 | .0095093 | 5.78| 0.000| .0363139 .0735897 |
| not_smsa       | -.1045232 | .0431082 | -2.42| 0.015| -.1890138 -.0200326 |
| 1.south        | -1.714868 | .3384558 | -5.07| 0.000| -2.378229 -1.051507 |
| year           | -.0115881 | .0084125 | -1.38| 0.168| -.0280763 .0049001 |
| south#c.year 1 | .0149796 | .0041687 | 3.59| 0.000| .0068091 .0231501 |
| _cons          | -1.488278 | .4468005 | -3.33| 0.001| -2.363991 -0.6125652 |

Example 2

In [R] cloglog, we showed these results and compared them with cloglog, vce(cluster id). xtcloglog with the pa option allows a vce(robust) option (the random-effects estimator does not allow the vce(robust) specification), so we can obtain the population-averaged cloglog estimator with the robust variance calculation by typing
. xtcloglog union age grade not_smsa south##c.year, pa vce(robust)

(output omitted)

GEE population-averaged model

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of obs</td>
<td>2620</td>
</tr>
<tr>
<td>Number of groups</td>
<td>4434</td>
</tr>
</tbody>
</table>

Link: cloglog

Family: binomial

Correlation: exchangeable

Wald chi2(6) = 157.24

Scale parameter: 1

Prob > chi2 = 0.0000

(Std. Err. adjusted for clustering on idcode)

Semirobust

|         | Coef. | Std. Err. | z     | P>|z|  | [95% Conf. Interval] |
|---------|-------|-----------|-------|------|---------------------|
| union   |       |           |       |      |                     |
| age     | .0153737 | .0079446 | 1.94  | 0.053 | -.0001974 to .0309448 |
| grade   | .0549518 | .0117258 | 4.69  | 0.000 | .0319697 to .077934  |
| not_smsa| -.1045232 | .0548598 | -1.91 | 0.057 | -.2120465 to .0030001 |
| 1.south | -1.714868 | .4864999 | -3.52 | 0.000 | -2.66839 to -.7613455 |
| year    | -.0115881 | .0085742 | -1.35 | 0.177 | -.0283932 to .005217  |
| south#c.year | 1  | .0149796 | .0060548 | 2.47  | 0.013 | .0031124 to .0268468  |
| _cons   | -1.488278 | .4924738 | -3.02 | 0.003 | -2.453509 to -.5230472 |

These standard errors are similar to those shown for cloglog, vce(cluster id) in [R] cloglog.

Technical note

The random-effects model is calculated using quadrature, which is an approximation whose accuracy depends partially on the number of integration points used. We can use the quadchk command to see if changing the number of integration points affects the results. If the results change, the quadrature approximation is not accurate given the number of integration points. Try increasing the number of integration points using the intpoints() option and run quadchk again. Do not attempt to interpret the results of estimates when the coefficients reported by quadchk differ substantially. See [XT] quadchk for details and [XT] xtprob for an example.

Because the xtcloglog likelihood function is calculated by Gauss–Hermite quadrature, on large problems the computations can be slow. Computation time is roughly proportional to the number of points used for the quadrature.
Stored results

xtcloglog, re stores the following in e():

Scalars
- e(N): number of observations
- e(N_g): number of groups
- e(N_cd): number of completely determined observations
- e(k): number of parameters
- e(k_aux): number of auxiliary parameters
- e(k_eq): number of equations in e(b)
- e(k_eq_model): number of equations in overall model test
- e(k_dv): number of dependent variables
- e(df_m): model degrees of freedom
- e(ll): log likelihood
- e(ll_0): log likelihood, constant-only model
- e(ll_c): log likelihood, comparison model
- e(chi2): $\chi^2$
- e(chi2_c): $\chi^2$ for comparison test
- e(N_clust): number of clusters
- e(rho): $\rho$
- e(sigma_u): panel-level standard deviation
- e(n_quad): number of quadrature points
- e(g_min): smallest group size
- e(g_avg): average group size
- e(g_max): largest group size
- e(p): significance
- e(rank): rank of e(V)
- e(rank0): rank of e(V) for constant-only model
- e(ic): number of iterations
- e(rc): return code
- e(converged): 1 if converged, 0 otherwise
Macros

\[\text{e(cmd)}\] xtcloglog
\[\text{e(cmdline)}\] command as typed
\[\text{e(depvar)}\] name of dependent variable
\[\text{e(ivar)}\] variable denoting groups
\[\text{e(model)}\] re
\[\text{e(wtype)}\] weight type
\[\text{e(wexp)}\] weight expression
\[\text{e(title)}\] title in estimation output
\[\text{e(clustvar)}\] name of cluster variable
\[\text{e(offset)}\] linear offset variable
\[\text{e(chi2type)}\] Wald or LR; type of model \(\chi^2\) test
\[\text{e(chi2_ct)}\] Wald or LR; type of model \(\chi^2\) test corresponding to \[\text{e(chi2_c)}\]
\[\text{e(vce)}\] vcetype specified in vce()
\[\text{e(vcetype)}\] title used to label Std. Err.
\[\text{e(intmethod)}\] integration method
\[\text{e(distrib)}\] Gaussian; the distribution of the random effect
\[\text{e(which)}\] \text{max or min}; whether optimizer is to perform maximization or minimization
\[\text{e(ml_method)}\] type of ml method
\[\text{e(user)}\] name of likelihood-evaluator program
\[\text{e(technique)}\] maximization technique
\[\text{e(properties)}\] b V
\[\text{e(predict)}\] program used to implement predict
\[\text{e(asbalanced)}\] factor variables fvset as asbalanced
\[\text{e(asobserved)}\] factor variables fvset as asobserved

Matrices

\[\text{e(b)}\] coefficient vector
\[\text{e(Cns)}\] constraints matrix
\[\text{e(ilog)}\] iteration log
\[\text{e(gradient)}\] gradient vector
\[\text{e(V)}\] variance–covariance matrix of the estimators
\[\text{e(V_modelbased)}\] model-based variance

Functions

\[\text{e(sample)}\] marks estimation sample
**xtcloglog, pa** stores the following in `e()`:

Scalars

- `e(N)` number of observations
- `e(N_g)` number of groups
- `e(df_m)` model degrees of freedom
- `e(chi2)` \( \chi^2 \)
- `e(p)` significance
- `e(df_pear)` degrees of freedom for Pearson \( \chi^2 \)
- `e(chi2_dev)` \( \chi^2 \) test of deviance
- `e(chi2_dis)` \( \chi^2 \) test of deviance dispersion
- `e(deviance)` deviance
- `e(dispers)` deviance dispersion
- `e(phi)` scale parameter
- `e(g_min)` smallest group size
- `e(g_avg)` average group size
- `e(g_max)` largest group size
- `e(rank)` rank of `e(V)`
- `e(tol)` target tolerance
- `e(dif)` achieved tolerance
- `e(rc)` return code

Macros

- `e(cmd)` `xtgee`
- `e(cmd2)` `xtcloglog`
- `e(cmdline)` command as typed
- `e(depvar)` name of dependent variable
- `e(ivar)` variable denoting groups
- `e(tvar)` variable denoting time within groups
- `e(model)` `pa`
- `e(family)` `binomial`
- `e(link)` `cloglog`; link function
- `e(corr)` correlation structure
- `e(scale)` `x2`, `dev`, `phi`, or `#`; scale parameter
- `e(type)` weight type
- `e(wexp)` weight expression
- `e(offset)` linear offset variable
- `e(chi2type)` Wald; type of model \( \chi^2 \) test
- `e(vce)` `vctype` specified in `vce()`
- `e(vcetype)` title used to label Std. Err.
- `e(nmp)` `nmp`, if specified
- `e(properties)` `b V`
- `e(predict)` program used to implement `predict`
- `e(marginsnotok)` predictions disallowed by `margins`
- `e(asbalanced)` factor variables `fvset` as `asbalanced`
- `e(asobserved)` factor variables `fvset` as `asobserved`

Matrices

- `e(b)` coefficient vector
- `e(Cns)` constraints matrix
- `e(R)` estimated working correlation matrix
- `e(V)` variance–covariance matrix of the estimators
- `e(V_modelbased)` model-based variance

Functions

- `e(sample)` marks estimation sample

**Methods and formulas**

`xtcloglog, pa` reports the population-averaged results obtained using `xtgee`, `family(binomial) link(cloglog)` to obtain estimates.
For the random-effects model, assume a normal distribution, $N(0, \sigma^2)$, for the random effects $\nu_i$,

$$\Pr(y_{i1}, \ldots, y_{in_i} | x_{i1}, \ldots, x_{in_i}) = \int_{-\infty}^{\infty} \frac{e^{-\nu_i^2/2\sigma^2}}{\sqrt{2\pi}\sigma} \left\{ \prod_{t=1}^{n_i} F(y_{it}, x_{it}\beta + \nu_i) \right\} d\nu_i$$

where

$$F(y, z) = \begin{cases} 1 - \exp \{- \exp(z)\} & \text{if } y \neq 0 \\ \exp\{- \exp(z)\} & \text{otherwise} \end{cases}$$

The panel-level likelihood $l_i$ is given by

$$l_i = \int_{-\infty}^{\infty} \frac{e^{-\nu_i^2/2\sigma^2}}{\sqrt{2\pi}\sigma} \left\{ \prod_{t=1}^{n_i} F(y_{it}, x_{it}\beta + \nu_i) \right\} d\nu_i$$

$$\equiv \int_{-\infty}^{\infty} g(y_{it}, x_{it}, \nu_i) d\nu_i$$

This integral can be approximated with $M$-point Gauss–Hermite quadrature

$$\int_{-\infty}^{\infty} e^{-x^2} h(x) dx \approx \sum_{m=1}^{M} w^*_m h(a^*_m)$$

This is equivalent to

$$\int_{-\infty}^{\infty} f(x) dx \approx \sum_{m=1}^{M} w^*_m \exp\{(a^*_m)^2\} f(a^*_m)$$

where the $w^*_m$ denote the quadrature weights and the $a^*_m$ denote the quadrature abscissas. The log likelihood, $L$, is the sum of the logs of the panel-level likelihoods $l_i$.

The default approximation of the log likelihood is by adaptive Gauss–Hermite quadrature, which approximates the panel-level likelihood with

$$l_i \approx \sqrt{2\hat{\sigma}_i} \sum_{m=1}^{M} w^*_m \exp\{(a^*_m)^2\} g(y_{it}, x_{it}, \sqrt{2\hat{\sigma}_i}a^*_m + \hat{\mu}_i)$$

where $\hat{\sigma}_i$ and $\hat{\mu}_i$ are the adaptive parameters for panel $i$. Therefore, with the definition of $g(y_{it}, x_{it}, \nu_i)$, the total log likelihood is approximated by
\[ L \approx \sum_{i=1}^{n} w_i \log \left[ \sqrt{2\sigma_i} \sum_{m=1}^{M} w_m^* \exp \left\{ (a_m^*)^2 \right\} \exp \left\{ -\frac{(\sqrt{2\sigma_i} a_m^* + \hat{\mu}_i)^2}{2\sigma_v^2} \right\} \prod_{t=1}^{n_i} F(y_{it}, x_{it}\beta + \sqrt{2\sigma_i} a_m^* + \hat{\mu}_i) \right] \]

where \( w_i \) is the user-specified weight for panel \( i \); if no weights are specified, \( w_i = 1 \).

The default method of adaptive Gauss–Hermite quadrature is to calculate the posterior mean and variance and use those parameters for \( \hat{\mu}_i \) and \( \hat{\sigma}_i \) by following the method of Naylor and Smith (1982), further discussed in Skrondal and Rabe-Hesketh (2004). We start with \( \hat{\sigma}_{i,0} = 1 \) and \( \hat{\mu}_{i,0} = 0 \), and the posterior means and variances are updated in the \( k \)th iteration. That is, at the \( k \)th iteration of the optimization for \( l_i \), we use

\[ l_{i,k} \approx \sum_{m=1}^{M} \sqrt{2\hat{\sigma}_{i,k-1}} w_m^* \exp \left\{ (a_m^*)^2 \right\} g(y_{it}, x_{it}, \sqrt{2\hat{\sigma}_{i,k-1}} a_m^* + \hat{\mu}_{i,k-1}) \]

Letting

\[ \tau_{i,m,k-1} = \sqrt{2\hat{\sigma}_{i,k-1}} a_m^* + \hat{\mu}_{i,k-1} \]

\[ \hat{\mu}_{i,k} = \frac{\sum_{m=1}^{M} \left( \tau_{i,m,k-1} \right) \sqrt{2\hat{\sigma}_{i,k-1}} w_m^* \exp \left\{ (a_m^*)^2 \right\} g(y_{it}, x_{it}, \tau_{i,m,k-1})}{l_{i,k}} \]

and

\[ \hat{\sigma}_{i,k} = \frac{\sum_{m=1}^{M} \left( \tau_{i,m,k-1} \right)^2 \sqrt{2\hat{\sigma}_{i,k-1}} w_m^* \exp \left\{ (a_m^*)^2 \right\} g(y_{it}, x_{it}, \tau_{i,m,k-1})}{l_{i,k}} - (\hat{\mu}_{i,k})^2 \]

and this is repeated until \( \hat{\mu}_{i,k} \) and \( \hat{\sigma}_{i,k} \) have converged for this iteration of the maximization algorithm. This adaptation is applied on every iteration until the log-likelihood change from the preceding iteration is less than a relative difference of 1e–6; after this, the quadrature parameters are fixed.

The log likelihood can also be calculated by nonadaptive Gauss–Hermite quadrature, the int-method(ghermite) option, where \( \rho = \sigma_v^2/(\sigma_v^2 + 1) \):

\[ L = \sum_{i=1}^{n} w_i \log \left\{ \Pr(y_{i1}, \ldots, y_{in_i} | x_{i1}, \ldots, x_{in_i}) \right\} \]

\[ \approx \sum_{i=1}^{n} w_i \log \left[ \frac{1}{\sqrt{\pi}} \sum_{m=1}^{M} w_m^* \prod_{t=1}^{n_i} F \left( y_{it}, x_{it}\beta + a_m^* \left( \frac{2\rho}{1-\rho} \right)^{1/2} \right) \right] \]

Both quadrature formulas require that the integrated function be well approximated by a polynomial of degree equal to the number of quadrature points. The number of periods (panel size) can affect whether

\[ \prod_{t=1}^{n_i} F(y_{it}, x_{it}\beta + \nu_i) \]
is well approximated by a polynomial. As panel size and $\rho$ increase, the quadrature approximation can become less accurate. For large $\rho$, the random-effects model can also become unidentified. Adaptive quadrature gives better results for correlated data and large panels than nonadaptive quadrature; however, we recommend that you use the \texttt{quadchk} command (see \texttt{[XT] quadchk}) to verify the quadrature approximation used in this command, whichever approximation you choose.

\textbf{xtcloglog, re and the robust VCE estimator}

Specifying \texttt{vce(robust)} or \texttt{vce(cluster clustvar)} causes the Huber/White/sandwich VCE estimator to be calculated for the coefficients estimated in this regression. See \texttt{[P] _robust}, particularly \textit{Introduction} and \textit{Methods and formulas}. Wooldridge (2013) and Arellano (2003) discuss this application of the Huber/White/sandwich VCE estimator. As discussed by Wooldridge (2013), Stock and Watson (2008), and Arellano (2003), specifying \texttt{vce(robust)} is equivalent to specifying \texttt{vce(cluster panelvar)}, where \texttt{panelvar} is the variable that identifies the panels.

Clustering on the panel variable produces a consistent VCE estimator when the disturbances are not identically distributed over the panels or there is serial correlation in $\epsilon_{it}$.

The cluster–robust VCE estimator requires that there are many clusters and the disturbances are uncorrelated across the clusters. The panel variable must be nested within the cluster variable because of the within-panel correlation that is generally induced by the random-effects transform when there is heteroskedasticity or within-panel serial correlation in the idiosyncratic errors.

\textbf{References}


Also see

[XT] **xtcloglog postestimation** — Postestimation tools for xtcloglog

[XT] **quadchk** — Check sensitivity of quadrature approximation

[XT] **xtgee** — Fit population-averaged panel-data models by using GEE

[XT] **xtlogit** — Fixed-effects, random-effects, and population-averaged logit models

[XT] **xtprobit** — Random-effects and population-averaged probit models

[XT] **xtset** — Declare data to be panel data

[ME] **mecloglog** — Multilevel mixed-effects complementary log-log regression

[MI] **estimation** — Estimation commands for use with mi estimate

[R] **cloglog** — Complementary log-log regression

[U] 20 Estimation and postestimation commands
Description

The following postestimation commands are available after xtcloglog:

<table>
<thead>
<tr>
<th>Command</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>contrast</td>
<td>contrasts and ANOVA-style joint tests of estimates</td>
</tr>
<tr>
<td>estat ic</td>
<td>Akaike’s and Schwarz’s Bayesian information criteria (AIC and BIC)</td>
</tr>
<tr>
<td>estat summarize</td>
<td>summary statistics for the estimation sample</td>
</tr>
<tr>
<td>estat vce</td>
<td>variance–covariance matrix of the estimators (VCE)</td>
</tr>
<tr>
<td>estimates</td>
<td>cataloging estimation results</td>
</tr>
<tr>
<td>forecast</td>
<td>dynamic forecasts and simulations</td>
</tr>
<tr>
<td>lincom</td>
<td>point estimates, standard errors, testing, and inference for linear combinations of coefficients</td>
</tr>
<tr>
<td>lrtest</td>
<td>likelihood-ratio test</td>
</tr>
<tr>
<td>margins</td>
<td>marginal means, predictive margins, marginal effects, and average marginal effects</td>
</tr>
<tr>
<td>marginsplot</td>
<td>graph the results from margins (profile plots, interaction plots, etc.)</td>
</tr>
<tr>
<td>nlcom</td>
<td>point estimates, standard errors, testing, and inference for nonlinear combinations of coefficients</td>
</tr>
<tr>
<td>predict</td>
<td>predictions, residuals, influence statistics, and other diagnostic measures</td>
</tr>
<tr>
<td>predictnl</td>
<td>point estimates, standard errors, testing, and inference for generalized predictions</td>
</tr>
<tr>
<td>pwcompare</td>
<td>pairwise comparisons of estimates</td>
</tr>
<tr>
<td>test</td>
<td>Wald tests of simple and composite linear hypotheses</td>
</tr>
<tr>
<td>testnl</td>
<td>Wald tests of nonlinear hypotheses</td>
</tr>
</tbody>
</table>

1 estat ic is not appropriate after xtcloglog, pa.
2 forecast is not appropriate with mi estimation results.

Syntax for predict

Random-effects (RE) model

```
predict [type] newvar [if] [in] [, RE_statistic nooffset]
```

Population-averaged (PA) model

```
predict [type] newvar [if] [in] [, PA_statistic nooffset]
```
**xtcloglog postestimation** — Postestimation tools for xtcloglog

### RE statistic

<table>
<thead>
<tr>
<th>Main</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>xb</code></td>
<td>linear prediction; the default</td>
</tr>
<tr>
<td><code>pu0</code></td>
<td>probability of a positive outcome</td>
</tr>
<tr>
<td><code>stdp</code></td>
<td>standard error of the linear prediction</td>
</tr>
</tbody>
</table>

### PA statistic

<table>
<thead>
<tr>
<th>Main</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>mu</code></td>
<td>predicted probability of <code>depvar</code>; considers the <code>offset()</code>; the default</td>
</tr>
<tr>
<td><code>rate</code></td>
<td>predicted probability of <code>depvar</code></td>
</tr>
<tr>
<td><code>xb</code></td>
<td>linear prediction</td>
</tr>
<tr>
<td><code>stdp</code></td>
<td>standard error of the linear prediction</td>
</tr>
<tr>
<td><code>score</code></td>
<td>first derivative of the log likelihood with respect to $x_j\beta$</td>
</tr>
</tbody>
</table>

These statistics are available both in and out of sample; type `predict ... if e(sample)` ... if wanted only for the estimation sample.

### Menu for predict

Statistics > Postestimation > Predictions, residuals, etc.

### Options for predict

- **Main**
  - `xb` calculates the linear prediction. This is the default for the random-effects model.
  - `pu0` calculates the probability of a positive outcome, assuming that the random effect for that observation’s panel is zero ($\nu = 0$). This may not be similar to the proportion of observed outcomes in the group.
  - `stdp` calculates the standard error of the linear prediction.
  - `mu` and `rate` both calculate the predicted probability of `depvar`. `mu` takes into account the `offset()`. `rate` ignores those adjustments. `mu` and `rate` are equivalent if you did not specify `offset()`. `mu` is the default for the population-averaged model.
  - `score` calculates the equation-level score, $u_j = \partial \ln L_j(x_j\beta)/\partial(x_j\beta)$.
  - `nooffset` is relevant only if you specified `offset(varname)` for xtcloglog. It modifies the calculations made by `predict` so that they ignore the offset variable; the linear prediction is treated as $x_{it}\beta$ rather than $x_{it}\beta + \text{offset}_{it}$.
Remarks and examples

Example 1

In example 1 of [XT] xtcloglog, we fit the model

```
use http://www.stata-press.com/data/r13/union
(NLS Women 14-24 in 1968)
.xtcloglog union age grade not_smsa south##c.year, pa
(output omitted)
```

Here we use `margins` to determine the average effect each regressor has on the probability of a positive response in the sample.

```
.margins, dydx(*)
Average marginal effects Number of obs = 26200
Model VCE : Conventional
Expression : Pr(union != 0), predict()
dy/dx w.r.t. : age grade not_smsa 1.south year

<table>
<thead>
<tr>
<th></th>
<th>Delta-method</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>dy/dx</td>
<td>Std. Err.</td>
<td>z</td>
<td>P&gt;</td>
</tr>
<tr>
<td>age</td>
<td>.0028297</td>
<td>.0014952</td>
<td>1.89</td>
<td>0.058</td>
</tr>
<tr>
<td>grade</td>
<td>.0101144</td>
<td>.0017498</td>
<td>5.78</td>
<td>0.000</td>
</tr>
<tr>
<td>not_smsa</td>
<td>-.0192384</td>
<td>.0079304</td>
<td>-2.43</td>
<td>0.015</td>
</tr>
<tr>
<td>1.south</td>
<td>-.0913197</td>
<td>.0073101</td>
<td>-12.49</td>
<td>0.000</td>
</tr>
<tr>
<td>year</td>
<td>-.0012694</td>
<td>.001534</td>
<td>-0.83</td>
<td>0.408</td>
</tr>
</tbody>
</table>
```

Note: dy/dx for factor levels is the discrete change from the base level.

We see that an additional year of schooling (covariate `grade`) increases the probability that a woman belongs to a union by an average of about one percentage point.

Also see

[XT] xtcloglog — Random-effects and population-averaged cloglog models

[U] 20 Estimation and postestimation commands
### Syntax

`xtdata [varlist] [if] [in] [, options]`

### Options

<table>
<thead>
<tr>
<th>Main</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>`re</td>
<td>convert data to a form suitable for random-effects estimation</td>
</tr>
<tr>
<td>`ratio(#)</td>
<td>ratio of random effect to pure residual (standard deviations)</td>
</tr>
<tr>
<td>`be</td>
<td>convert data to a form suitable for between estimation</td>
</tr>
<tr>
<td>`fe</td>
<td>convert data to a form suitable for fixed-effects (within) estimation</td>
</tr>
<tr>
<td>`nodouble</td>
<td>keep original variable type; default is to recast type as double</td>
</tr>
<tr>
<td>`clear</td>
<td>overwrite current data in memory</td>
</tr>
</tbody>
</table>

A panel variable must be specified; use `xtset`; see [XT] `xtset`.

### Description

`xtdata` produces a transformed dataset of the variables specified in `varlist` or of all the variables in the data. Once the data are transformed, Stata’s `regress` command may be used to perform specification searches more quickly than `xtreg`; see [R] `regress` and [XT] `xtreg`. Using `xtdata`, `re` also creates a variable named `constant`. When using `regress` after `xtdata`, `re`, specify `noconstant` and include `constant` in the regression. After `xtdata`, `be` and `xtdata`, `fe`, you need not include `constant` or specify `regress`’s `noconstant` option.

### Options

- `re` specifies that the data are to be converted into a form suitable for random-effects estimation. `re` is the default if `be`, `fe`, or `re` is not specified. `ratio(#)` must also be specified.
- `ratio(#)` (use with `xtdata, re` only) specifies the ratio $\frac{\sigma_v}{\sigma_e}$, which is the ratio of the random effect to the pure residual. This is the ratio of the standard deviations, not the variances.
- `be` specifies that the data are to be converted into a form suitable for between estimation.
- `fe` specifies that the data are to be converted into a form suitable for fixed-effects (within) estimation.
nodouble specifies that transformed variables keep their original types, if possible. The default is to recast variables to double.

Remember that xtdata transforms variables to be differences from group means, pseudodifferences from group means, or group means. Specifying nodouble will decrease the size of the resulting dataset but may introduce roundoff errors in these calculations.

clear specifies that the data may be converted even though the dataset has changed since it was last saved on disk.

Remarks and examples

If you have not read [XT] xt and [XT] xtreg, please do so.

The formal estimation commands of xtreg—see [XT] xtreg—do not produce results instantaneously, especially with large datasets. Equations (2), (3), and (4) of [XT] xtreg describe the data necessary to fit each of the models with OLS. The idea here is to transform the data once to the appropriate form and then use regress to fit such models more quickly.

> Example 1

We will use the example in [XT] xtreg demonstrating between-effects regression. Another way to estimate the between equation is to convert the data in memory to the between data:

```
. use http://www.stata-press.com/data/r13/nlswork
    (National Longitudinal Survey. Young Women 14-26 years of age in 1968)
. generate age2=age^2
    (24 missing values generated)
. generate ttl_exp2 = ttl_exp^2
. generate tenure2=tenure^2
    (433 missing values generated)
. generate byte black = race==2
. xtdata ln_w grade age* ttl_exp* tenure* black not_smsa south, be clear
. regress ln_w grade age* ttl_exp* tenure* black not_smsa south
```

The output is the same as that produced by xtreg, be; the reported $R^2$ is the $R^2$ between. Using xtdata followed by just one regress does not save time. Using xtdata is justified when you intend to explore the specification of the model by running many alternative regressions.
Technical note

When using `xtdata`, you must eliminate any variables that you do not intend to use and that have missing values. `xtdata` follows a casewise-deletion rule, which means that an observation is excluded from the conversion if it is missing on any of the variables. In the example above, we specified that the variables be converted on the command line. We could also drop the variables first, and it might even be useful to preserve our estimation sample:

```
. use http://www.stata-press.com/data/r13/nlswork, clear
    (National Longitudinal Survey. Young Women 14-26 years of age in 1968)
. generate age2=age^2
    (24 missing values generated)
. generate ttl_exp2 = ttl_exp^2
. generate tenure2=tenure^2
    (433 missing values generated)
. generate byte black = race==2
. keep id year ln_w grade age* ttl_exp* tenure* black not_smsa south
. save xtdatasmpl
    file xtdatasmpl.dta saved
```

Example 2

`xtdata` with the `fe` option converts the data so that results are equivalent to those from estimating by using `xtreg` with the `fe` option.

```
. xtdata, fe
. regress ln_w grade age* ttl_exp* tenure* black not_smsa south
    note: grade omitted because of collinearity
    note: black omitted because of collinearity

    Source | SS      df    MS
    ------|---------|--------|--------
    Model | 412.443881  8  51.554852
    Residual | 1976.1223 28082  .070369714
    Total | 2388.5662 28090  .085032617

    Number of obs = 28091
    F(  8, 28082) = 732.64
    Prob > F      =  0.0000
    R-squared    =  0.1727
    Adj R-squared =  0.1724
    Root MSE     =  0.26527

    ln_wage | Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
    --------|---------|---------|--------|---------|------------------------
    grade   | 0 (omitted)
    age     | .0359987  .0030903  11.65  0.000    .0299415    .0420558
    age2    | -.000723  .0000486  -14.88  0.000   -.0008183   -.0006277
    ttl_exp | .0334668  .0027061  12.37  0.000     .0281627    .0387708
    ttl_exp2 | .0002163  .0001166   1.86  0.064    -.0000122    .0004447
    tenure  | .0357539  .0016871  21.19  0.000     .0324472    .0390606
    tenure2 | -.0019701  .0001141  -17.27  0.000   -.0021937   -.0017465
    black   | 0 (omitted)
    not_smsa | -.0890108  .0088982  -10.23  0.000   -.1060597   -.0719619
    south   | -.0606309  .0099761  -6.08  0.000   -.0801845   -.0410772
    _cons   | 1.03732  .0443093  23.41  0.000     .9504716    1.124168
```

The coefficients reported by `regress` after `xtdata, fe` are the same as those reported by `xtreg, fe`, but the standard errors are slightly smaller. This is because no adjustment has been made to the estimated covariance matrix for the estimation of the person means. The difference is small, however, and results are adequate for a specification search.
Example 3

To use `xtdata, re`, you must specify the ratio \( \sigma_v/\sigma_\epsilon \), which is the ratio of the standard deviations of the random effect and pure residual. Merely to show the relationship of `regress` after `xtdata, re` to `xtreg, re`, we will specify this ratio as \( 0.25790526/0.29068923 = 0.88721987 \), which is the number `xtreg` reports when the model is fit from the outset; see the random-effects example in [XT] `xtreg`. For specification searches, however, it is adequate to specify this number more crudely, and, when performing the specification search for this manual entry, we used `ratio(1)`.

```
. use http://www.stata-press.com/data/r13/xtdatasmpl, clear
    (National Longitudinal Survey. Young Women 14-26 years of age in 1968)
. xtdata, clear re ratio(.88721987)
```

<table>
<thead>
<tr>
<th>theta</th>
<th>5%</th>
<th>median</th>
<th>95%</th>
<th>max</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2520</td>
<td>0.2520</td>
<td>0.5499</td>
<td>0.7016</td>
<td>0.7206</td>
</tr>
</tbody>
</table>

`xtdata` reports the distribution of \( \theta \) based on the specified ratio. If these were balanced data, \( \theta \) would have been constant.

When running regressions with these data, you must specify the `noconstant` option and include the variable constant:

```
. regress ln_w grade age* ttl_exp* tenure* black not_smsa south constant, > noconstant
```

```
Source | SS    | df  | MS       | Number of obs = 28091
        |       |     |          | F( 11, 28080) =14302.56
Model  | 13271.7208 | 11 | 1206.52007 | Prob > F = 0.0000
Residual | 2368.74223 | 28080 | .084356917 |
Total | 15640.463 | 28091 | .556778435 |

`ln_wage` | Coef. | Std. Err. | t     | P>|t| | [95% Conf. Interval]
---|------|----------|------|------|-----------------
grade | .0646499 | .0017812 | 36.30 | 0.000 | .0611587 -.0681411
age | .0368059 | .0031195 | 11.80 | 0.000 | .0306915 -.0429203
age2 | -.0007133 | .00005 | -14.27 | 0.000 | -.0008113 -.0006153
south | -.0868922 | .0073032 | -11.90 | 0.000 | -.1012068 -.0725775
constant | .2387206 | .049469 | 4.83 | 0.000 | .141759 -.3356822

Results are the same coefficients and standard errors that `xtreg, re` estimated in example 4 of [XT] `xtreg`. The summaries at the top, however, should be ignored, as they are expressed in terms of (4) of [XT] `xtreg`, and, moreover, for a model without a constant.

⚠️ Technical note

Using `xtdata` requires some caution. The following guidelines may help:

1. `xtdata` is intended for use only during the specification search phase of analysis. Results should be estimated with `xtreg` on unconverted data.
2. After converting the data, you may use `regress` to obtain estimates of the coefficients and their standard errors. For `regress` after `xtdata, fe`, the standard errors are too small, but only slightly.
3. You may loosely interpret the coefficient’s significance tests and confidence intervals. However, for results after `xtdata, fe` and `re`, an incorrect (but close to correct) distribution is assumed.
4. You should ignore the summary statistics reported at the top of \texttt{regress}'s output.

5. After converting the data, you may form linear, but not nonlinear, combinations of regressors; that is, if your data contained age, it would not be correct to convert the data and then form age squared. All nonlinear transformations should be done before conversion. (For \texttt{xtdata, be}, you can get away with forming nonlinear combinations ex post, but the results will not be exact.)

\begin{quote}
\textbf{Technical note}

The \texttt{xtdata} command can be used to help you examine data, especially with \texttt{scatter}.

\begin{verbatim}
. use http://www.stata-press.com/data/r13/xtdatasmpl, clear
(National Longitudinal Survey. Young Women 14-26 years of age in 1968)
. xtdata, be
. scatter ln_wage age, title(Between data) msymbol(o) msize(tiny)
\end{verbatim}
\end{quote}
. use http://www.stata-press.com/data/r13/xtdatasmpl, clear
(National Longitudinal Survey. Young Women 14-26 years of age in 1968)
. xtdata, fe
. scatter ln_wage age, title(Within data) msymbol(o) msize(tiny)

Within data

. use http://www.stata-press.com/data/r13/xtdatasmpl, clear
(National Longitudinal Survey. Young Women 14-26 years of age in 1968)
. scatter ln_wage age, title(Overall data) msymbol(o) msize(tiny)

Overall data
Methods and formulas

(This section is a continuation of the Methods and formulas of [XT] xtreg.)

xtdata, be, fe, and re transform the data according to (2), (3), and (4), respectively, of [XT] xtreg, except that xtdata, fe adds back in the overall mean, thus forming the transformation

\[ x_{it} = \bar{x}_i + \bar{x} \]

xtdata, re requires the user to specify \( r \) as an estimate of \( \sigma_\nu/\sigma_\epsilon \). \( \theta_i \) is calculated from

\[ \theta_i = 1 - \frac{1}{\sqrt{T_i r^2 + 1}} \]

Also see

[XT] xtsum — Summarize xt data
xtdescribe — Describe pattern of xt data

Syntax

```
xtdescribe [ if ] [ in ] [ , options ]
```

options Description

<table>
<thead>
<tr>
<th>Main</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>patterns(#)</td>
<td>maximum participation patterns; default is patterns(9)</td>
</tr>
<tr>
<td>width(#)</td>
<td>display # width of participation patterns; default is width(100)</td>
</tr>
</tbody>
</table>

A panel variable and a time variable must be specified; use xtset; see [XT] xtset. by is allowed; see [D] by.

Menu

Statistics > Longitudinal/panel data > Setup and utilities > Describe pattern of xt data

Description

`xtdescribe` describes the participation pattern of cross-sectional time-series (xt) data.

Options

- `patterns(#)`: specifies the maximum number of participation patterns to be reported; `patterns(9)` is the default. Specifying `patterns(50)` would list up to 50 patterns. Specifying `patterns(1000)` is taken to mean `patterns(∞)`; all the patterns will be listed.
- `width(#)`: specifies the desired width of the participation patterns to be displayed; `width(100)` is the default. If the number of times is greater than `width()`, then each column in the participation pattern represents multiple periods as indicated in a footnote at the bottom of the table. The actual width may differ slightly from the requested width depending on the span of the time variable and the number of periods.

Remarks and examples

If you have not read [XT] xt, please do so.

`xtdescribe` describes the cross-sectional and time-series aspects of the data in memory.
Example 1

In [XT] xt, we introduced data based on a subsample of the NLSY data on young women aged 14–26 years in 1968. Here is a description of the data used in many of the [XT] xt examples:

```
use http://www.stata-press.com/data/r13/nlswork
(National Longitudinal Survey. Young Women 14-26 years of age in 1968)
.xtdescribe
idcode: 1, 2, ..., 5159  n =  4711
year: 68, 69, ..., 88  T =  15
Delta(year) = 1 unit
Span(year) = 21 periods
(idcode*year uniquely identifies each observation)

Distribution of T_i:  min 5% 25% 50% 75% 95% max
1 1 3 5 9 13 15
Freq. Percent Cum. Pattern
136  2.89  2.89 1....................
114  2.42  5.31 1....................1
 89  1.89  7.20 1....................11
 87  1.85  9.04 1....................11
 86  1.83 10.87 111111.1.11.1.11.11
 61  1.29 12.16 111111.1.11.1.11.11
 56  1.19 13.35 111111.1.11.1.11.11
 54  1.15 14.50 111111.1.11.1.11.11
 54  1.15 15.64 111111.1.11.1.11.11
 3974 84.36 100.00 (other patterns)
```

xtdescribe tells us that we have 4,711 women in our data and that the idcode that identifies each ranges from 1 to 5,159. We are also told that the maximum number of individual years over which we observe any woman is 15, though the year variable spans 21 years. The delta or periodicity of year is one unit, meaning that in principle we could observe each woman yearly. We are reassured that idcode and year, taken together, uniquely identify each observation in our data. We are also shown the distribution of $T_i$; 50% of our women are observed 5 years or less. Only 5% of our women are observed for 13 years or more.

Finally, we are shown the participation pattern. A 1 in the pattern means one observation that year; a dot means no observation. The largest fraction of our women (still only 2.89%) was observed in the single year 1968 and not thereafter; the next largest fraction was observed in 1988 but not before; and the next largest fraction was observed in 1985, 1987, and 1988.

At the bottom is the sum of the participation patterns, including the patterns that were not shown. We can see that none of the women were observed in six of the years (there are six dots). (The survey was not administered in those six years.)

We could see more of the patterns by specifying the patterns() option, or we could see all the patterns by specifying patterns(1000).

Example 2

The strange participation patterns shown above have to do with our subsampling of the data, not with the administrators of the survey. Here are the data from which we drew the sample used in the [XT] xt examples:
We have multiple observations per year. In the pattern, 2 indicates that a woman appears twice in the year, 3 indicates 3 times, and so on—X indicates 10 or more, should that be necessary.

In fact, this is a dataset that was itself extracted from the NLSY, in which \( t \) is not time but job number. To simplify exposition, we made a simpler dataset by selecting the last job in each year.

#### Example 3

When the number of periods is greater than the width of the participation pattern, each column will represent more than one period.
have an observation taken at either 4:00 PM or 5:00 PM on March 9, but we do not have observations for both times. There are three patients for whom we are missing both the 10:00 PM and 11:00 PM observations on March 10, and there are two patients for whom we are missing the 4:00 PM and 5:00 PM observations for March 9.

Reference

Also see

[XT] *xtsum* — Summarize xt data
[XT] *xttab* — Tabulate xt data
**xtdpd — Linear dynamic panel-data estimation**

### Syntax

```plaintext
xtdpd  depvar [ indepvars ] [ if ] [ in ], dgmmiv(varlist [...]) [ options ]
```

**options**  
**Description**

<table>
<thead>
<tr>
<th>Model</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>*dgmmiv(varlist[...])</code></td>
<td>GMM-type instruments for the difference equation; can be specified more than once</td>
</tr>
<tr>
<td><code>lgmmiv(varlist[...])</code></td>
<td>GMM-type instruments for the level equation; can be specified more than once</td>
</tr>
<tr>
<td><code>iv(varlist[...])</code></td>
<td>standard instruments for the difference and level equations; can be specified more than once</td>
</tr>
<tr>
<td><code>div(varlist[...])</code></td>
<td>standard instruments for the difference equation only; can be specified more than once</td>
</tr>
<tr>
<td><code>liv(varlist)</code></td>
<td>standard instruments for the level equation only; can be specified more than once</td>
</tr>
<tr>
<td><code>noconstant</code></td>
<td>suppress constant term</td>
</tr>
<tr>
<td><code>twostep</code></td>
<td>compute the two-step estimator instead of the one-step estimator</td>
</tr>
<tr>
<td><code>hascons</code></td>
<td>check for collinearity only among levels of independent variables; by default checks occur among levels and differences</td>
</tr>
<tr>
<td><code>fodeviation</code></td>
<td>use forward-orthogonal deviations instead of first differences</td>
</tr>
</tbody>
</table>

**SE/Robust**  
`vce(vcetype)`  
`vcetype` may be `gmm` or `robust`

**Reporting**  
`level(#)`  
set confidence level; default is `level(95)`
| `artests(#)` | use `#` as maximum order for AR tests; default is `artests(2)` |
| `display_options` | control spacing and line width |
| `coeflegend` | display legend instead of statistics |

*`dgmmiv()` is required.*

A panel variable and a time variable must be specified; use `xtset`; see [XT] `xtset`.  
`depvar`, `indepvars`, and all `varlists` may contain time-series operators; see [U] 11.4.4 Time-series `varlists`.  
`by`, `statsby`, and `xi` are allowed; see [U] 11.1.10 Prefix commands.  
`coeflegend` does not appear in the dialog box.  
See [U] 20 Estimation and postestimation commands for more capabilities of estimation commands.
**Description**

Linear dynamic panel-data models include \( p \) lags of the dependent variable as covariates and contain unobserved panel-level effects, fixed or random. By construction, the unobserved panel-level effects are correlated with the lagged dependent variables, making standard estimators inconsistent. \texttt{xtdpd} fits a dynamic panel-data model by using the Arellano–Bond (1991) or the Arellano–Bover/Blundell–Bond (1995, 1998) estimator.

At the cost of a more complicated syntax, \texttt{xtdpd} can fit models with low-order moving-average correlation in the idiosyncratic errors or predetermined variables with a more complicated structure than allowed for \texttt{xtabond} or \texttt{xtdpdsys}; see \cite{xtabond} and \cite{xtdpdsys}.

**Options**

\texttt{dgmmiv(}varlist \texttt{[, lagrange(flag [llag ])]\texttt{)} specifies GMM-type instruments for the differenced equation. Levels of the variables are used to form GMM-type instruments for the difference equation. All possible lags are used, unless \texttt{lagrange(flag llag)} restricts the lags to begin with \texttt{flag} and end with \texttt{llag}. You may specify as many sets of GMM-type instruments for the differenced equation as you need within the standard Stata limits on matrix size. Each set may have its own \texttt{flag} and \texttt{llag}. \texttt{dgmmiv()} is required.

\texttt{lgmmiv(}varlist \texttt{[, lag(#)]\texttt{)} specifies GMM-type instruments for the level equation. Differences of the variables are used to form GMM-type instruments for the level equation. The first lag of the differences is used unless \texttt{lag(#)} is specified, indicating that \texttt{#}th lag of the differences be used. You may specify as many sets of GMM-type instruments for the level equation as you need within the standard Stata limits on matrix size. Each set may have its own \texttt{lag}.

\texttt{iv(}varlist \texttt{[, nodifference]\texttt{)} specifies standard instruments for both the differenced and level equations. Differences of the variables are used as instruments for the differenced equations, unless \texttt{nodifference} is specified, which requests that levels be used. Levels of the variables are used as instruments for the level equations. You may specify as many sets of standard instruments for both the differenced and level equations as you need within the standard Stata limits on matrix size.

\texttt{div(}varlist \texttt{[, nodifference]\texttt{)} specifies additional standard instruments for the differenced equation. Specified variables may not be included in \texttt{iv()} or in \texttt{liv()}. Differences of the variables are used, unless \texttt{nodifference} is specified, which requests that levels of the variables be used as instruments for the differenced equation. You may specify as many additional sets of standard instruments for the differenced equation as you need within the standard Stata limits on matrix size.

\texttt{liv(}varlist \texttt{)} specifies additional standard instruments for the level equation. Specified variables may not be included in \texttt{iv()} or in \texttt{div()}. Levels of the variables are used as instruments for the level equation. You may specify as many additional sets of standard instruments for the level equation as you need within the standard Stata limits on matrix size.

\texttt{noconstant}; see \cite{estimation_options}.

\texttt{twostep} specifies that the two-step estimator be calculated.
hascons specifies that xtdpd check for collinearity only among levels of independent variables; by default checks occur among levels and differences.

fodeviation specifies that forward-orthogonal deviations are to be used instead of first differences. 

fodeviation is not allowed when there are gaps in the data or when lgmmiv() is specified.

vce(vcetype) specifies the type of standard error reported, which includes types that are derived from asymptotic theory and that are robust to some kinds of misspecification; see Methods and formulas.

vce(gmm), the default, uses the conventionally derived variance estimator for generalized method of moments estimation.

vce(robust) uses the robust estimator. For the one-step estimator, this is the Arellano–Bond robust VCE estimator. For the two-step estimator, this is the Windmeijer (2005) WC-robust estimator.

level(#); see [R] estimation options.

artests(#) specifies the maximum order of the autocorrelation test to be calculated. The tests are reported by estat abond; see [XT] xtdpd postestimation. Specifying the order of the highest test at estimation time is more efficient than specifying it to estat abond, because estat abond must refit the model to obtain the test statistics. The maximum order must be less than or equal to the number of periods in the longest panel. The default is artests(2).

display_options: vsquish and nolstretch; see [R] estimation options.

The following option is available with xtdpd but is not shown in the dialog box:

coeflegend; see [R] estimation options.

Remarks and examples

If you have not read [XT] xtabond and [XT] xtdpdsys, you should do so before continuing.

Consider the dynamic panel-data model

\[ y_{it} = \sum_{j=1}^{p} \alpha_j y_{i,t-j} + x_{it} \beta_1 + w_{it} \beta_2 + \nu_i + \epsilon_{it} \quad i = \{1, \ldots, N\}; \quad t = \{1, \ldots, T_i\} \quad (1) \]

where

the \( \alpha_1, \ldots, \alpha_p \) are \( p \) parameters to be estimated,

\( x_{it} \) is a \( 1 \times k_1 \) vector of strictly exogenous covariates,

\( \beta_1 \) is a \( k_1 \times 1 \) vector of parameters to be estimated,

\( w_{it} \) is a \( 1 \times k_2 \) vector of predetermined covariates,

\( \beta_2 \) is a \( k_2 \times 1 \) vector of parameters to be estimated,

\( \nu_i \) are the panel-level effects (which may be correlated with \( x_{it} \) or \( w_{it} \)), and

and \( \epsilon_{it} \) are i.i.d. or come from a low-order moving-average process, with variance \( \sigma^2_{\epsilon} \).
Building on the work of Anderson and Hsiao (1981, 1982) and Holtz-Eakin, Newey, and Rosen (1988), Arellano and Bond (1991) derived one-step and two-step GMM estimators using moment conditions in which lagged levels of the dependent and predetermined variables were instruments for the differenced equation. Blundell and Bond (1998) show that the lagged-level instruments in the Arellano–Bond estimator become weak as the autoregressive process becomes too persistent or the ratio of the variance of the panel-level effect \( \nu_i \) to the variance of the idiosyncratic error \( \epsilon_{it} \) becomes too large. Building on the work of Arellano and Bover (1995), Blundell and Bond (1998) proposed a system estimator that uses moment conditions in which lagged differences are used as instruments for the level equation in addition to the moment conditions of lagged levels as instruments for the differenced equation. The additional moment conditions are valid only if the initial condition \( E[\nu_i \Delta y_{i2}] = 0 \) holds for all \( i \); see Blundell and Bond (1998) and Blundell, Bond, and Windmeijer (2000).

\texttt{xtdpd} fits dynamic panel-data models by using the Arellano–Bond or the Arellano–Bover/Blundell–Bond system estimator. The parameters of many standard models can be more easily estimated using the Arellano–Bond estimator implemented in \texttt{xtabond} or using the Arellano–Bover/Blundell–Bond system estimator implemented in \texttt{xtdpdsys}; see [\texttt{XT}] \texttt{xtabond} and [\texttt{XT}] \texttt{xtdpdsys}. \texttt{xtdpd} can fit more complex models at the cost of a more complicated syntax. That the idiosyncratic errors follow a low-order MA process and that the predetermined variables have a more complicated structure than accommodated by \texttt{xtabond} and \texttt{xtdpdsys} are two common reasons for using \texttt{xtdpd} instead of \texttt{xtabond} or \texttt{xtdpdsys}.

The standard GMM robust two-step estimator of the VCE is known to be seriously biased. Windmeijer (2005) derived a bias-corrected robust estimator for two-step VCEs from GMM estimators known as the WC-robust estimator, which is implemented in \texttt{xtdpd}.

The Arellano–Bond test of autocorrelation of order \( m \) and the Sargan test of overidentifying restrictions derived by Arellano and Bond (1991) are computed by \texttt{xtdpd} but reported by \texttt{estat abond} and \texttt{estat sargan}, respectively; see [\texttt{XT}] \texttt{xtdpd postestimation}.

Because \texttt{xtdpd} extends \texttt{xtabond} and \texttt{xtdpdsys}, [\texttt{XT}] \texttt{xtabond} and [\texttt{XT}] \texttt{xtdpdsys} provide useful background.

\textbf{Example 1: An Arellano–Bond estimator}

Arellano and Bond (1991) apply their new estimators and test statistics to a model of dynamic labor demand that had previously been considered by Layard and Nickell (1986), using data from an unbalanced panel of firms from the United Kingdom. All variables are indexed over the firm \( i \) and time \( t \). In this dataset, \( n_{it} \) is the log of employment in firm \( i \) inside the United Kingdom at time \( t \), \( w_{it} \) is the natural log of the real product wage, \( k_{it} \) is the natural log of the gross capital stock, and \( y_{s, it} \) is the natural log of industry output. The model also includes time dummies \( yr1980, yr1981, yr1982, yr1983, \) and \( yr1984 \). To gain some insight into the syntax for \texttt{xtdpd}, we reproduce the first example from [\texttt{XT}] \texttt{xtabond} using \texttt{xtdpd}:
. use http://www.stata-press.com/data/r13/abdata
. xtdpd L(0/2).n L(0/1).w L(0/2).(k ys) yr1980-yr1984 year, noconstant
> div(L(0/1).w L(0/2).(k ys) yr1980-yr1984 year) dgmmiv(n)

Dynamic panel-data estimation
Group variable: id
Time variable: year
Number of obs = 611
Group variable: id
Number of groups = 140
Time variable: year
Obs per group: min = 4
avg = 4.364286
max = 6
Number of instruments = 41
Wald chi2(16) = 1757.07
Prob > chi2 = 0.0000

One-step results

|    | Coef. | Std. Err. | z    | P>|z| | [95% Conf. Interval] |
|----|-------|-----------|------|------|----------------------|
| n  | .6862261 | .1486163 | 4.62 | 0.000 | .3949435 .9775088   |
| L1 | -.0853582 | .0444365 | -1.92 | 0.055 | -.1724523 .0017358 |
| L2 | -.6078208 | .0657694 | -9.24 | 0.000 | -.7367265 -.4789151 |
| w  | .3926237 | .1092374 | 3.59 | 0.000 | .1785222 .6067251   |
| L1 | .3568456 | .0370314 | 9.64 | 0.000 | .2842653 .4294259   |
| L2 | -.0580012 | .0583051 | -0.99 | 0.320 | -.172777 .0562747   |
| k  | .1057969 | .0416274 | 2.50 | 0.012 | .0177873 .2938062   |
| L1 | .0029062 | .0212705 | 0.14 | 0.891 | -.0387832 .0445975 |
| L2 | -.0404378 | .0354707 | -1.14 | 0.254 | -.1099591 .0290836 |
| yrs | -.0652767 | .0482092 | -1.35 | 0.176 | -.1597646 .0292111 |
| yrs | -.0690928 | .0627354 | -1.10 | 0.271 | -.1920521 .0538664 |
| yrs | -.0650302 | .0781322 | -0.83 | 0.405 | -.2181665 .0881061 |
| year | .0095545 | .0142073 | 0.67 | 0.501 | -.0182912 .0374002 |

Instruments for differenced equation
GMM-type: L(2/).n
Standard: D.w LD.w D.k LD.k L2D.k D.ys LD.ys L2D.ys D.yr1980

Unlike most instrumental-variables estimation commands, the independent variables in the varlist
are not automatically used as instruments. In this example, all the independent variables are strictly
exogenous, so we include them in div(), a list of variables whose first differences will be instruments
for the differenced equation. We include the dependent variable in dgmmiv(), a list of variables whose
lagged levels will be used to create GMM-type instruments for the differenced equation. (GMM-type
instruments are discussed in a technical note below.)

The footer in the output reports the instruments used. The first line indicates that xtdpd used
lags from 2 on back to create the GMM-type instruments described in Arellano and Bond (1991)
and Holtz-Eakin, Newey, and Rosen (1988). The second line says that the first difference of all the
variables included in the div() varlist were used as standard instruments for the differenced equation.

Technical Note: GMM-type instruments are discussed in Arellano and Bond (1991) and Holtz-Eakin, Newey, and Rosen (1988).
GMM-type instruments are built from lags of one variable. Ignoring the strictly exogenous variables for simplicity, our model is

\[ n_{it} = \alpha_1 n_{it-1} + \alpha_2 n_{it-2} + \nu_i + \epsilon_{it} \]  

(2)

After differencing we have

\[ \Delta n_{it} = \Delta \alpha_1 n_{it-1} + \Delta \alpha_2 n_{it-2} + \Delta \epsilon_{it} \]  

(3)

Equation (3) implies that we need instruments that are not correlated with either \( \epsilon_{it} \) or \( \epsilon_{it-1} \). Equation (2) shows that \( L2.n \) is the first lag of \( n \) that is not correlated with \( \epsilon_{it} \) or \( \epsilon_{it-1} \), so it is the first lag of \( n \) that can be used to instrument the differenced equation.

Consider the following data from one of the complete panels in the previous example:

```
. list id year n L2.n dl2.n if id==140
```

<table>
<thead>
<tr>
<th>id</th>
<th>year</th>
<th>n</th>
<th>L2.</th>
<th>L2D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1023</td>
<td>1976</td>
<td>.4324315</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>1024</td>
<td>1977</td>
<td>.3694925</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>1026</td>
<td>1979</td>
<td>.3632532</td>
<td>.3694925</td>
<td>-.0629391</td>
</tr>
<tr>
<td>1027</td>
<td>1980</td>
<td>.3371863</td>
<td>.3541718</td>
<td>-.0153207</td>
</tr>
<tr>
<td>1028</td>
<td>1981</td>
<td>.285179</td>
<td>.3632532</td>
<td>.0090815</td>
</tr>
<tr>
<td>1029</td>
<td>1982</td>
<td>.1756326</td>
<td>.3371863</td>
<td>-.026067</td>
</tr>
<tr>
<td>1030</td>
<td>1983</td>
<td>.1275133</td>
<td>.285179</td>
<td>-.0520073</td>
</tr>
<tr>
<td>1031</td>
<td>1984</td>
<td>.0889263</td>
<td>.1756326</td>
<td>-.1095464</td>
</tr>
</tbody>
</table>
```

The missing values in L2D.n show that we lose 3 observations because of lags and the difference that removes the panel-level effects. The first nonmissing observation occurs in 1979 and observations on \( n \) from 1976 and 1977 are available to instrument the 1979 differenced equation. The table below gives the observations available to instrument the differenced equation for the data above.

<table>
<thead>
<tr>
<th>Year of difference errors</th>
<th>Years of instruments</th>
<th>Number of instruments</th>
</tr>
</thead>
</table>

The table shows that there are a total of 27 GMM-type instruments.

The output in the example above informs us that there were a total of 41 instruments applied to the differenced equation. Because there are 14 standard instruments, there must have been 27 GMM-type instruments, which matches our above calculation.
Example 2: An Arellano–Bond estimator with predetermined variables

Sometimes we cannot assume strict exogeneity. Recall that a variable $x_{it}$ is said to be strictly exogenous if $E[x_{it} \epsilon_{is}] = 0$ for all $t$ and $s$. If $E[x_{it} \epsilon_{is}] \neq 0$ for $s < t$ but $E[x_{it} \epsilon_{is}] = 0$ for all $s \geq t$, the variable is said to be predetermined. Intuitively, if the error term at time $t$ has some feedback on the subsequent realizations of $x_{it}$, $x_{it}$ is a predetermined variable. In the output below, we use xtdpd to reproduce example 6 in [XT] xtabond.

```
. xtdpd L(0/2).n L(0/1).(w ys) L(0/2).k yr1980-yr1984 year,
  > div(L(0/1).(ys) yr1980-yr1984 year) dgmmiv(n) dgmmiv(L.w L2.k, lag(1 .))
  > twostep noconstant vce(robust)
Dynamic panel-data estimation
Number of obs = 611
Group variable: id Number of groups = 140
Time variable: year
Obs per group: min = 4
  avg = 4.364286
  max = 6
Number of instruments = 83
Wald chi2(15) = 958.30
Prob > chi2 = 0.0000
```

Two-step results

| n    | WC-Robust Coef. (Std. Err.) | z     | P>|z| | [95% Conf. Interval] |
|------|-----------------------------|-------|------|------------------------|
| n    |                             |       |      |                        |
| L1.  | .8580958 (.1265515)         | 6.78  | .000 | .6100594 1.106132     |
| L2.  | -.081207 (.0760703)         | -1.07 | .286 | -.2303022 .0678881    |
| w    | -.6910855 (.1387684)        | -4.98 | .000 | -.9630666 -.4191044   |
| L1.  | .5961712 (.1497338)         | 3.98  | .000 | .3026982 .8896441     |
| ys   | .6936392 (.1728623)         | 4.01  | .000 | .3548354 1.032443     |
| L1.  | -.8773678 (.2183085)        | -4.02 | .000 | -1.305245 -.449491    |
| k    | -.4140654 (.1382788)        | 2.99  | .003 | .1430439 .6850868     |
| L1.  | -.1537048 (.1220244)        | -1.26 | .208 | -.3928681 .0854586    |
| L2.  | -.1025833 (.0710886)        | -1.44 | .149 | -.2419143 .0367477    |
| yr1980| -.0072451 (.017163)         | -0.42 | .673 | -.0408839 .0263938    |
| yr1981| -.0069608 (.030207)         | -2.02 | .044 | -.1201655 -.0017561   |
| yr1982| -.1130369 (.0454826)        | -2.49 | .013 | -.2021812 -.0238926   |
| yr1983| -.1335249 (.0600213)        | -2.22 | .026 | -.2511645 -.0158853   |
| yr1984| -.1623177 (.0725434)        | -2.24 | .025 | -.3045001 -.0201352   |
| year | .0264501 (.0119329)         | 2.22  | .027 | .003062 .0498381      |

Instruments for differenced equation

`GMM-type: L(2/.).n L(1/.).L.w L(1/.).L2.k`


The footer informs us that we are now including GMM-type instruments from the first lag of $L.w$ on back and from the first lag of $L2.k$ on back.
**Example 3: A weaker definition of predetermined variables**

As discussed in [XT] `xtabond` and [XT] `xtdpdsys`, `xtabond` and `xtdpdsys` both use a strict definition of predetermined variables with lags. In the strict definition, the most recent lag of the variable in `pre()` is considered predetermined. (Here specifying `pre(w, lag(1, .))` to `xtabond` means that $L.w$ is a predetermined variable and `pre(k, lag(2, .))` means that $L2.k$ is a predetermined variable.) In a weaker definition, the current observation is considered predetermined, but subsequent lags are included in the model. Here $w$ and $k$ would be predetermined instead of $L.w$ and $L2.w$. The output below implements this weaker definition for the previous example.

```
  . xtdpd L(0/2).n L(0/1).(w ys) L(0/2).k yr1980-yr1984 year,
        > div(L(0/1).(ys) yr1980-yr1984 year) dgmmiv(n) dgmmiv(w k, lag(1 .))
        > twostep noconstant vce(robust)
```

```
Dynamic panel-data estimation
Number of obs = 611
Group variable: id
Number of groups = 140
Time variable: year
Obs per group: min = 4
              avg = 4.364286
              max = 6
Number of instruments = 101
Wald chi2(15) = 879.53
Prob > chi2 = 0.0000
Two-step results
(Std. Err. adjusted for clustering on id)
```

|   | WC-Robust | Std. Err. | z    | P>|z| | [95% Conf. Interval] |
|---|-----------|-----------|------|------|----------------------|
| n |           |           |      |      |                      |
| L1.n | .6343155 | .1221058 | 5.19 | 0.000 | .3949925 .8736384 |
| L2.n | -.0871247| .0704816 | -1.24| 0.216 | -.2252661 .0510168 |
| w  | --.       | -.720063 | .1133359 | -6.35 | 0.000 | -.9421973 -.4979287 |
| L1.w | .238069 | .1223186 | 1.95 | 0.052 | -.0016712 .4778091 |
| ys | --.       | .5999718 | .1653036 | 3.63 | 0.000 | .2759827 .923961 |
| L1.ys | -.5674808 | .1656411 | -3.43 | 0.001 | -.8921314 -.2428303 |
| k  | --.       | .3931997 | .0986673 | 3.99 | 0.000 | .1998153 .5865842 |
| L1.k | -.0019641 | .0772814 | -0.03 | 0.980 | -.1534329 .1495047 |
| L2.k | -.0231165 | .0487317 | -0.47 | 0.635 | -.1186288 .0723958 |
| yr1980 | -.006209 | .0162138 | -0.38 | 0.702 | -.0379875 .0255694 |
| yr1981 | -.0398491 | .0313794 | -1.27 | 0.204 | -.1013516 .0216535 |
| yr1982 | -.0525715 | .0397346 | -1.32 | 0.186 | -.1304498 .0253068 |
| yr1983 | -.0451175 | .051418 | -0.88 | 0.380 | -.145895 .05566 |
| yr1984 | -.0437772 | .0614391 | -0.71 | 0.476 | -.1641955 .0766412 |
| year | .0173374 | .0108665 | 1.60 | 0.111 | -.0039605 .0386352 |

Instruments for differenced equation
GMM-type: L(2/.).n L(1/.).w L(1/.).k

As expected, the output shows that the additional 18 instruments available under the weaker definition can affect the magnitudes of the estimates. Applying the stricter definition when the true model was generated by the weaker definition yielded consistent but inefficient results; there were some additional...
Example 4: A system estimator of a dynamic panel-data model

Here we use *xtdpd* to reproduce example 2 from [XT] *xtdpdsys* in which we used the system estimator to fit a model with predetermined variables.

```stata
. xtdpd L(0/1).n L(0/2). (w k) yr1980-yr1984 year,
   > div(yr1980-yr1984 year) dgmmiv(n) dgmmiv(L2.(w k), lag(1 .))
   > lgmmiv(n L1.(w k)) vce(robust) hascons
```

```
Dynamic panel-data estimation
Number of obs = 751
Group variable: id
Number of groups = 140
Time variable: year
Obs per group: min = 5
   avg = 5.364286
   max = 7
Number of instruments = 95
Wald chi2(13) = 7562.80
Prob > chi2 = 0.0000
One-step results
(Std. Err. adjusted for clustering on id)
```

|      | Robust Coef. | Std. Err. | z     | P>|z| | [95% Conf. Interval] |
|------|--------------|------------|-------|------|---------------------|
| n    | .913278      | .0460602   | 19.83 | 0.000 | .8230017             |
| L1.  | .913278      | .0460602   | 19.83 | 0.000 | 1.003554             |
| w    | -.728159     | .1019044   | -7.15 | 0.000 | -.927888             |
| L1.  | .5602737     | .1939617   | 2.89  | 0.004 | .1801156             |
| L2.  | -.0523028    | .1487653   | -0.35 | 0.725 | -.3438775            |
| k    | .4820097     | .0760787   | 6.34  | 0.000 | .3328983             |
| L1.  | -.2846944    | .0831902   | -3.42 | 0.001 | -.4477442            |
| L2.  | -.1394181    | .0405709   | -3.44 | 0.001 | -.2189356            |
| yr1980 | -.0325146   | .0216371   | -1.50 | 0.133 | -.0749226            |
| yr1981 | -.0726116   | .0346482   | -2.10 | 0.036 | -.1405207            |
| yr1982 | -.0477038   | .0451914   | -1.06 | 0.291 | -.1362772            |
| yr1983 | -.0396264   | .0585734   | -0.71 | 0.478 | -.1491362            |
| yr1984 | -.0810383   | .0736648   | -1.10 | 0.271 | -.2254186            |
| year  | .0192741     | .0145326   | 1.33  | 0.185 | -.0092092            |
| _cons| -.3734972    | 28.77747   | -1.30 | 0.194 | 93.75253             |

The first lags of the variables included in *lgmmiv()* are used to create GMM-type instruments for the level equation. Only the first lags of the variables in *lgmmiv()* are used because the moment conditions using higher lags are redundant; see Blundell and Bond (1998) and Blundell, Bond, and Windmeijer (2000).
Example 5: Allowing for MA(1) errors

All the previous examples have used moment conditions that are valid only if the idiosyncratic errors are i.i.d. This example shows how to use xtdpd to estimate the parameters of a model with first-order moving-average [MA(1)] errors using the Arellano–Bond estimator, the Arellano–Bover/Blundell–Bond system estimator, or any other consistent GMM estimator you want to specify. For simplicity, we assume that the independent variables are strictly exogenous. Also, to highlight the fact that we can specify the instrument list flexibly, we only include the levels and first lags of the exogenous variables in the instrument list. An Arellano–Bond estimator, for instance, would have included levels and first and second lags of the exogenous variables.

We begin by noting that the Sargan test rejects the null hypothesis that the overidentifying restrictions are valid in the model with i.i.d. errors.

```
xtdpd L(0/1).n L(0/2).(w k) yr1980-yr1984 year, > div(L(0/1).(w k) yr1980-yr1984 year) dgmmiv(n) hascons 
(output omitted)
.esstat sargan
Sargan test of overidentifying restrictions
H0: overidentifying restrictions are valid
  chi2(24) = 49.70094
  Prob > chi2 = 0.0015
```

Assuming that the idiosyncratic errors are MA(1) implies that only lags three or higher are valid instruments for the differenced equation. (See the technical note below.)

```
xtdpd L(0/1).n L(0/2).(w k) yr1980-yr1984 year, > div(L(0/1).(w k) yr1980-yr1984 year) dgmmiv(n, lag(3 .)) hascons
Dynamic panel-data estimation Number of obs = 751
Group variable: id Number of groups = 140
Time variable: year
Obs per group: min = 5
avg = 5.364286
max = 7
Number of instruments = 32 Wald chi2(13) = 1195.04
Prob > chi2 = 0.0000
One-step results

| n  | Coef.  | Std. Err. | z     | P>|z|  | [95% Conf. Interval] |
|----|--------|-----------|-------|-------|---------------------|
| L1. | .8696303 | .2014473 | 4.32  | 0.000 | .4748008          | 1.26446 |
| w  | -.5802971 | .0762659 | -7.61 | 0.000 | -.7297756         | -.4308187 |
   | L1. | .2918658 | .1543883 | 1.89  | 0.059 | -.0107295         | .5944613 |
   | L2. | -.5903459 | .2995123 | -1.97 | 0.049 | -1.177379         | -.0033126 |
| k  | -.3428139 | .0447916 | 7.65  | 0.000 | .2550239          | .4306039 |
   | L1. | -.1383918 | .0825823 | -1.68 | 0.094 | -.3002502         | .0234665 |
   | L2. | -.0260956 | .1535855 | -0.17 | .865 | -.3271177         | .2749265 |
| yr1980 | -.0036873 | .0301587 | -0.12 | 0.903 | -.0627973         | .0554226 |
| yr1981 | .00218 | .0592014 | 0.04  | 0.971 | -.1138526         | .1182125 |
| yr1982 | .0782939 | .0897622 | 0.87  | 0.383 | -.0976367         | .2542246 |
| yr1983 | .1734231 | .1309194 | 1.32  | 0.185 | -.0831193         | .4299655 |
| yr1984 | .2400685 | .1734456 | 1.38  | 0.166 | -.0998787         | .5800157 |
| year | -.0354681 | .0309963 | -1.14 | 0.253 | -.0962198         | .0252836 |
| _cons | 73.13706 | 62.61443 | 1.17  | 0.243 | -.49.58496         | 195.8591 |
```

---

Technical Note:
The technical note explains why lags three or higher are only valid instruments for the differenced equation when the idiosyncratic errors are MA(1).
Instruments for differenced equation
GMM-type: L(3/.) n
D.yr1984 D.year

Instruments for level equation
Standard: _cons

The results from estat sargan no longer reject the null hypothesis that the overidentifying restrictions are valid.

. estat sargan
Sargan test of overidentifying restrictions
H0: overidentifying restrictions are valid
   chi2(18) = 20.80081
   Prob > chi2 = 0.2896

Moving on to the system estimator, we note that the Sargan test rejects the null hypothesis after fitting the model with i.i.d. errors.

. xtdpd L(0/1).n L(0/2).(w k) yr1980-yr1984 year,
   > div(L(0/1).(w k) yr1980-yr1984 year) dgmmiv(n) lgmmiv(n) hascons
(output omitted)
. estat sargan
Sargan test of overidentifying restrictions
H0: overidentifying restrictions are valid
   chi2(31) = 59.22907
   Prob > chi2 = 0.0017

Now we fit the model using the additional moment conditions constructed from the second lag of n as an instrument for the level equation.
. xtdpd L(0/1).n L(0/2).(w k) yr1980-yr1984 year,
> div(L(0/1).(w k) yr1980-yr1984 year) dgmmiv(n, lag(3 .)) lgmmiv(n, lag(2))
> hascons

Dynamic panel-data estimation
Number of obs = 751
Group variable: id Number of groups = 140
Time variable: year
Obs per group: min = 5
avg = 5.364286
max = 7

Number of instruments = 38
Wald chi2(13) = 3680.01
Prob > chi2 = 0.0000

One-step results

|    | Coef. | Std. Err. | z     | P>|z|  | [95% Conf. Interval] |
|----|-------|-----------|-------|-------|----------------------|
| n  | .9603675 | .095608 | 10.04 | 0.000 | .7729794 1.147756   |
| L1 |       |          |       |       |                      |
| w  | -.5433987 | .068835 | -7.89 | 0.000 | -.6783128 -.4084845 |
| L1 | .4356183 | .0881727 | 4.94 | 0.000 | .262803 .6084336   |
| L2 | -.2785721 | .1115061 | -2.50 | 0.012 | -.4971201 -.0600241 |
| k  |       |          |       |       |                      |
| L1 | -.160103 | .0546915 | -2.93 | 0.003 | -.2672963 -.0529096 |
| L2 | -.1295766 | .0507752 | -2.55 | 0.011 | -.2290943 -.030059  |
| yr1980 | -.0200704 | .0248954 | -0.81 | 0.420 | -.0688644 .0287236 |
| yr1981 | -.0426838 | .0422155 | -1.01 | 0.313 | -.1253246 .040157  |
| yr1982 | .0048723 | .0600938 | 0.08 | 0.935 | -.1129093 .122654  |
| yr1983 | .0458978 | .0785687 | 0.58 | 0.569 | -.1080941 .1988897 |
| yr1984 | .0633219 | .1026188 | 0.62 | 0.537 | -.1378074 .2644511 |
| _cons | -.0075599 | .019059 | -0.40 | 0.692 | -.0449148 .029795  |
|     | 16.20856 | 38.00619 | 0.43 | 0.670 | -58.28221 90.69932 |

Instruments for differenced equation
GMM-type: L(3/).n
D.yr1984 D.year

Instruments for level equation
GMM-type: L2D.n
Standard: _cons

The estimate of the coefficient on L.n is now .96. Blundell, Bond, and Windmeijer (2000, 63–65) show that the moment conditions in the system estimator remain informative as the true coefficient on L.n approaches unity. Holtz-Eakin, Newey, and Rosen (1988) show that because the large-sample distribution of the estimator is derived for fixed number of periods and a growing number of individuals there is no “unit-root” problem.

The results from estat sargan no longer reject the null hypothesis that the overidentifying restrictions are valid.

. estat sargan
Sargan test of overidentifying restrictions
HO: overidentifying restrictions are valid
chi2(24) = 27.22585
Prob > chi2 = 0.2940
Technical note

To find the valid moment conditions for the model with MA(1) errors, we begin by writing the model

\[ n_{it} = \alpha n_{it-1} + \beta x_{it} + \nu_i + \epsilon_{it} + \gamma \epsilon_{it-1} \]

where the \( \epsilon_{it} \) are assumed to be i.i.d.

Because the composite error, \( \epsilon_{it} + \gamma \epsilon_{it-1} \), is MA(1), only lags two or higher are valid instruments for the level equation, assuming the initial condition that \( E[\nu_i \Delta n_{it2}] = 0 \). The key to this point is that lagging the above equation two periods shows that \( \epsilon_{it-2} \) and \( \epsilon_{it-3} \) appear in the equation for \( n_{it-2} \). Because the \( \epsilon_{it} \) are i.i.d., \( n_{it-2} \) is a valid instrument for the level equation with errors \( \nu_i + \epsilon_{it} + \gamma \epsilon_{it-1} \). (\( n_{it-2} \) will be correlated with \( n_{it-1} \) but uncorrelated with the errors \( \nu_i + \epsilon_{it} + \gamma \epsilon_{it-1} \).) An analogous argument works for higher lags.

First-differencing the above equation yields

\[ \Delta n_{it} = \alpha \Delta n_{it-1} + \beta \Delta x_{it} + \Delta \epsilon_{it} + \gamma \Delta \epsilon_{it-1} \]

Because \( \epsilon_{it-2} \) is the farthest lag of \( \epsilon_{it} \) that appears in the differenced equation, lags three or higher are valid instruments for the differenced composite errors. (Lagging the level equation three periods shows that only \( \epsilon_{it-3} \) and \( \epsilon_{it-4} \) appear in the equation for \( n_{it-3} \), which implies that \( n_{it-3} \) is a valid instrument for the current differenced equation. An analogous argument works for higher lags.)

Stored results

`xtdpd` stores the following in \( e() \):

Scalars
- `e(N)` number of observations
- `e(N_g)` number of groups
- `e(df_m)` model degrees of freedom
- `e(g_min)` smallest group size
- `e(g_avg)` average group size
- `e(g_max)` largest group size
- `e(t_min)` minimum time in sample
- `e(t_max)` maximum time in sample
- `e(ch12)` \( \chi^2 \)
- `e(arm#)` test for autocorrelation of order \#`e(artests)` number of AR tests computed
- `e(sig2)` estimate of \( \sigma^2 \)
- `e(rss)` sum of squared differenced residuals
- `e(sargan)` Sargan test statistic
- `e(rank)` rank of \( e(V) \)
- `e(zrank)` rank of instrument matrix
Consider dynamic panel-data models of the form

$$y_{it} = \sum_{j=1}^{p} \alpha_j y_{i,t-j} + x_{it} \beta_1 + w_{it} \beta_2 + \nu_i + \epsilon_{it}$$

where the variables are as defined as in (1).

$x$ and $w$ may contain lagged independent variables and time dummies.

Let $X_{it}^L = (y_{i,t-1}, y_{i,t-2}, \ldots, y_{i,t-p}, x_{it}, w_{it})$ be the $1 \times K$ vector of covariates for $i$ at time $t$, where $K = p + k_1 + k_2$, $p$ is the number of included lags, $k_1$ is the number of strictly exogenous variables in $x_{it}$, and $k_2$ is the number of predetermined variables in $w_{it}$. (The superscript $L$ stands for levels.)

Now rewrite this relationship as a set of $T_i$ equations for each individual,

$$y_{i}^L = X_{i}^L \delta + \nu_i \iota_i + \epsilon_i$$

where $T_i$ is the number of observations available for individual $i$; $y_{i}$, $\iota_{i}$, and $\epsilon_{i}$ are $T_i \times 1$, whereas $X_{i}$ is $T_i \times K$.

The estimators use both the levels and a transform of the variables in the above equation. Denote the transformed variables by an $*$, so that $y_{i}^*$ is the transformed $y_{i}^L$ and $X_{i}^*$ is the transformed $X_{i}^L$. The transform may be either the first difference or the forward-orthogonal deviations (FOD) transform. The $(i, t)$th observation of the FOD transform of a variable $x$ is given by

$$x_{it}^* = c_t \left\{ x_{it} - \frac{1}{T-t} (x_{it+1} + x_{it+2} + \cdots + x_{iT}) \right\}$$

where $c_t^2 = (T - t)/(T - t + 1)$ and $T$ is the number of observations on $x$; see Arellano and Bover (1995) and Arellano (2003).
Here we present the formulas for the Arellano–Bover/Blundell–Bond system estimator. The formulas for the Arellano–Bond estimator are obtained by setting the additional level matrices in the system estimator to null matrices.

Stacking the transformed and untransformed vectors of the dependent variable for a given \( i \) yields

\[
y_i = \left( \begin{array}{c} y_i^* \\ y_i^L \end{array} \right)
\]

Similarly, stacking the transformed and untransformed matrices of the covariates for a given \( i \) yields

\[
X_i = \left( \begin{array}{c} X_i^* \\ X_i^L \end{array} \right)
\]

\( Z_i \) is a matrix of instruments,

\[
Z_i = \left( \begin{array}{cccc} Z_{di} & 0 & D_i & 0 \\ 0 & Z_{Li} & 0 & L_i \end{array} \right)
\]

where \( Z_{di} \) is the matrix of GMM-type instruments created from the \texttt{dgmmiv()} options, \( Z_{Li} \) is the matrix of GMM-type instruments created from the \texttt{lgmmiv()} options, \( D_i \) is the matrix of standard instruments created from the \texttt{div()} options, \( L_i \) is the matrix of standard instruments created from the \texttt{liv()} options, \( I_i^d \) is the matrix of standard instruments created from the \texttt{iv()} options for the differenced errors, and \( I_i^L \) is the matrix of standard instruments created from the \texttt{iv()} options for the level errors.

\( \texttt{div()} \), \( \texttt{liv()} \), and \( \texttt{iv()} \) simply add columns to instrument matrix. The GMM-type instruments are more involved. Begin by considering a simple balanced-panel example in which our model is

\[
y_{it} = \alpha_1 y_{i,t-1} + \alpha_2 y_{i,t-2} + \nu_i + \epsilon_{it}
\]

We do not need to consider covariates because strictly exogenous variables are handled using \texttt{div()} , \texttt{iv()} , or \texttt{liv()} , and predetermined or endogenous variables are handled analogous to the dependent variable.

Assume that the data come from a balanced panel in which there are no missing values. After first-differencing the equation, we have

\[
\Delta y_{it} = \alpha_1 \Delta y_{i,t-1} + \alpha_2 \Delta y_{i,t-2} + \Delta \epsilon_{it}
\]

The first 3 observations are lost to lags and differencing. If we assume that the \( \epsilon_{it} \) are not autocorrelated, for each \( i \) at \( t = 4 \), \( y_{i1} \) and \( y_{i2} \) are valid instruments for the differenced equation. Similarly, at \( t = 5 \), \( y_{i1} \), \( y_{i2} \), and \( y_{i3} \) are valid instruments. We specify \texttt{dgmmiv(y)} to obtain an instrument matrix with one row for each period that we are instrumenting:

\[
Z_{di} = \begin{pmatrix}
y_{i1} & y_{i2} & 0 & 0 & 0 & \ldots & 0 & 0 & 0 \\
0 & 0 & y_{i1} & y_{i2} & y_{i3} & \ldots & 0 & 0 & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots \\
0 & 0 & 0 & \ldots & 0 & y_{i1} & \ldots & y_{i,T-2}
\end{pmatrix}
\]

Because \( p = 2 \), \( Z_{di} \) has \( T - p - 1 \) rows and \( \sum_{m=p}^{T-2} m \) columns.
Specifying \( \text{lgmmiv}(y) \) creates the instrument matrix

\[
Z_{Li} = \begin{pmatrix}
\Delta.y_{i2} & 0 & 0 & \ldots & 0 \\
0 & \Delta.y_{i3} & 0 & \ldots & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & 0 & \ldots & \Delta.y_{i(T_i-1)}
\end{pmatrix}
\]

This extends to other lag structures with complete data. Unbalanced data and missing observations are handled by dropping the rows for which there are no data and filling in zeros in columns where missing data are required. Suppose that, for some \( i \), the \( t = 1 \) observation was missing but was not missing for some other panels. \( \text{dgmmiv}(y) \) would then create the instrument matrix

\[
Z_{di} = \begin{pmatrix}
0 & 0 & 0 & y_{i2} & y_{i3} & 0 & 0 & 0 & \ldots & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & y_{i2} & y_{i3} & 0 & \ldots & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \ldots & 0 & y_{i2} & \ldots & y_{iT-2}
\end{pmatrix}
\]

\( Z_{di} \) has \( T_i - p - 1 \) rows and \( \sum_{m=p}^{\tau-2} \) columns, where \( \tau = \max_i \tau_i \) and \( \tau_i \) is the number of nonmissing observations in panel \( i \).

After defining

\[
Q_{xz} = \sum_i X_i'Z_i
\]

\[
Q_{zy} = \sum_i Z_i'y_i
\]

\[
W_1 = Q_{xz}A_1Q_{xz}'
\]

\[
A_1 = \left( \sum_i Z_i'H_{1i}Z_i \right)^{-1}
\]

and

\[
H_{1i} = \begin{pmatrix} H_{di} & 0 \\
0 & H_{Li} \end{pmatrix}
\]

the one-step estimates are given by

\[
\hat{\beta}_1 = W_1^{-1}Q_{xz}A_1Q_{zy}
\]
When using the first-difference transform $H_{di}$, is given by

$$H_{di} = \begin{pmatrix}
1 & -0.5 & 0 & \ldots & 0 & 0 \\
-0.5 & 1 & -0.5 & \ldots & 0 & 0 \\
& & & \ddots & & \\
0 & 0 & 0 & \ldots & 1 & -0.5 \\
0 & 0 & 0 & \ldots & -0.5 & 1
\end{pmatrix}$$

and $H_{Li}$ is given by 0.5 times the identity matrix. When using the FOD transform, both $H_{di}$ and $H_{Li}$ are equal to the identity matrix.

The transformed one-step residuals are given by

$$\hat{\epsilon}_{1i}^* = y_i^* - \hat{\beta}_1 X_i^*$$

which are used to compute

$$\hat{\sigma}_1^2 = (1/(N - K)) \sum_{i} \hat{\epsilon}_{1i}^* \hat{\epsilon}_{1i}^*$$

The GMM one-step VCE is then given by

$$\hat{V}_{GMM}[\hat{\beta}_1] = \hat{\sigma}_1^2 W_1^{-1}$$

The one-step level residuals are given by

$$\hat{\epsilon}_{1i}^L = y_i^L - \hat{\beta}_1 X_i^L$$

Stacking the residual vectors yields

$$\hat{\epsilon}_{1i} = \begin{pmatrix} \hat{\epsilon}_{1i}^* \\ \hat{\epsilon}_{1i}^L \end{pmatrix}$$

which is used to compute $H_{2i} = \hat{\epsilon}_{1i}^T \hat{\epsilon}_{1i}$, which is used in

$$A_2 = \left( \sum_i Z_i'H_{2i}Z_i \right)^{-1}$$

and the robust one-step VCE is given by

$$\hat{V}_{\text{robust}}[\hat{\beta}_1] = W_1^{-1} Q_{xz} A_1 A_2^{-1} A_1 Q'_{xz} W_1^{-1}$$

$\hat{V}_{\text{robust}}[\hat{\beta}_1]$ is robust to heteroskedasticity in the errors.
After defining

\[ W_2 = Q_{xz}A_2Q'_{xz} \]

the two-step estimates are given by

\[ \hat{\beta}_2 = W_2^{-1}Q_{xz}A_2Q_{zy} \]

The GMM two-step VCE is then given by

\[ \hat{V}_{\text{GMM}}[\hat{\beta}_2] = W_2^{-1} \]

The GMM two-step VCE is known to be severely biased. Windmeijer (2005) derived the Windmeijer bias-corrected (WC) estimator for the robust VCE of two-step GMM estimators. \texttt{xtdpd} implements this WC-robust estimator of the VCE. The formulas for this method are involved; see Windmeijer (2005). The WC-robust estimator of the VCE is robust to heteroskedasticity in the errors.

Acknowledgment

We thank David Roodman of the Center for Global Development, who wrote \texttt{xtabond2}.

References


Also see

[XT] **xtdpd postestimation** — Postestimation tools for xtdpd

[XT] **xtset** — Declare data to be panel data

[XT] **xtabond** — Arellano–Bond linear dynamic panel-data estimation

[XT] **xtdpdsys** — Arellano–Bover/Blundell–Bond linear dynamic panel-data estimation

[XT] **xtivreg** — Instrumental variables and two-stage least squares for panel-data models

[XT] **xtreg** — Fixed-, between-, and random-effects and population-averaged linear models

[XT] **xtregar** — Fixed- and random-effects linear models with an AR(1) disturbance

[R] **gmm** — Generalized method of moments estimation

[U] **20 Estimation and postestimation commands**
Description

The following postestimation commands are of special interest after *xtdpd*:

<table>
<thead>
<tr>
<th>Command</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>estat abond</code></td>
<td>test for autocorrelation</td>
</tr>
<tr>
<td><code>estat sargan</code></td>
<td>Sargan test of overidentifying restrictions</td>
</tr>
</tbody>
</table>

The following standard postestimation commands are also available:

<table>
<thead>
<tr>
<th>Command</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>estat summarize</code></td>
<td>summary statistics for the estimation sample</td>
</tr>
<tr>
<td><code>estat vce</code></td>
<td>variance–covariance matrix of the estimators (VCE)</td>
</tr>
<tr>
<td><code>estimates</code></td>
<td>cataloging estimation results</td>
</tr>
<tr>
<td><code>forecast</code></td>
<td>dynamic forecasts and simulations</td>
</tr>
<tr>
<td><code>lincom</code></td>
<td>point estimates, standard errors, testing, and inference for linear combinations of coefficients</td>
</tr>
<tr>
<td><code>margins</code></td>
<td>marginal means, predictive margins, marginal effects, and average marginal effects</td>
</tr>
<tr>
<td><code>marginsplot</code></td>
<td>graph the results from margins (profile plots, interaction plots, etc.)</td>
</tr>
<tr>
<td><code>nlcom</code></td>
<td>point estimates, standard errors, testing, and inference for nonlinear combinations of coefficients</td>
</tr>
<tr>
<td><code>predict</code></td>
<td>predictions, residuals, influence statistics, and other diagnostic measures</td>
</tr>
<tr>
<td><code>predictnl</code></td>
<td>point estimates, standard errors, testing, and inference for generalized predictions</td>
</tr>
<tr>
<td><code>test</code></td>
<td>Wald tests of simple and composite linear hypotheses</td>
</tr>
<tr>
<td><code>testnl</code></td>
<td>Wald tests of nonlinear hypotheses</td>
</tr>
</tbody>
</table>

**Special-interest postestimation commands**

`estat abond` reports the Arellano–Bond test for serial correlation in the first-differenced residuals.

`estat sargan` reports the Sargan test of the overidentifying restrictions.
Syntax for predict

```
predict [type] newvar [if] [in] [, xb e stdp difference]
```

Menu for predict

Statistics > Postestimation > Predictions, residuals, etc.

Options for predict

- **xb**, the default, calculates the linear prediction.
- **e** calculates the residual error.
- **stdp** calculates the standard error of the prediction, which can be thought of as the standard error of the predicted expected value or mean for the observation’s covariate pattern. The standard error of the prediction is also referred to as the standard error of the fitted value. **stdp** may not be combined with **difference**.
- **difference** specifies that the statistic be calculated for the first differences instead of the levels, the default.

Syntax for estat

- **Test for autocorrelation**
  ```
estat abond [, artests(#)]
```
- **Sargan test of overidentifying restrictions**
  ```
estat sargan
```

Menu for estat

Statistics > Postestimation > Reports and statistics

Option for estat abond

- **artests(#)** specifies highest order of serial correlation to be tested. By default, the tests computed during estimation are reported. The model will be refit when **artests(#)** specifies a higher order than that computed during the original estimation. The model can only be refit if the data have not changed.

Remarks and examples

Remarks are presented under the following headings:

- **estat abond**
- **estat sargan**
estat abond

The moment conditions used by `xtdpd` are valid only if there is no serial correlation in the idiosyncratic errors. Testing for serial correlation in dynamic panel-data models is tricky because one needs to apply a transform to remove the panel-level effects, but the transformed errors have a more complicated error structure than the idiosyncratic errors. The Arellano–Bond test for serial correlation reported by `estat abond` tests for serial correlation in the first-differenced errors.

Because the first difference of independently and identically distributed idiosyncratic errors will be autocorrelated, rejecting the null hypothesis of no serial correlation at order one in the first-differenced errors does not imply that the model is misspecified. Rejecting the null hypothesis at higher orders implies that the moment conditions are not valid. See example 5 in [XT] `xtdpd` for an alternative estimator that allows for idiosyncratic errors that follow a first-order moving average process.

After the one-step system estimator, the test can be computed only when `vce(robust)` has been specified.

estat sargan

Like all GMM estimators, the estimator in `xtdpd` can produce consistent estimates only if the moment conditions used are valid. Although there is no method to test if the moment conditions from an exactly identified model are valid, one can test whether the overidentifying moment conditions are valid. `estat sargan` implements the Sargan test of overidentifying conditions discussed in Arellano and Bond (1991).

Only for a homoskedastic error term does the Sargan test have an asymptotic chi-squared distribution. In fact, Arellano and Bond (1991) show that the one-step Sargan test overrejects in the presence of heteroskedasticity. Because its asymptotic distribution is not known under the assumptions of the `vce(robust)` model, `xtdpd` does not compute it when `vce(robust)` is specified.

Methods and formulas

The notation for $\hat{\epsilon}_1, \hat{\epsilon}_i, H_1, H_i, X_i, Z_i, W_1, W_2, \hat{V}_s[\hat{\beta}_s], A_1, A_2, Q_{xz}$, and $\hat{\sigma}_1^2$ has been defined in Methods and formulas of [XT] `xtdpd`.

The Arellano–Bond test for zero $m$th-order autocorrelation in the first-differenced errors is given by

$$A(m) = \frac{s_0}{\sqrt{s_1 + s_2 + s_3}}$$

where the definitions of $s_0, s_1, s_2,$ and $s_3$ vary over the estimators and transforms.

We begin by defining $\hat{u}_{1i}^* = Lm.\hat{\epsilon}_1^*$, with the missing values filled in with zeros. Letting $j = 1$ for the one-step estimator, $j = 2$ for the two-step estimator, $c = \text{GMM}$ for the GMM VCE estimator, and $c = \text{robust}$ for the robust VCE estimator, we can now define $s_0, s_1, s_2,$ and $s_3$:

$$s_0 = \sum_i \hat{u}_{ji}^*\hat{\epsilon}_{ji}^*$$

$$s_1 = \sum_i \hat{u}_{ji}^*H_{ji}\hat{u}_{ji}^*$$

$$s_2 = -2q_{ji} W_j^{-1} Q_{xz} A_j Q_{zu}$$
\[ s_3 = q_{jx} \hat{V}_c [\hat{\beta}_j] q'_{jx} \]

where

\[ q_{jx} = \left( \sum_i \hat{u}^{*'}_{ji} X_i \right) \]

and \( Q_{zu} \) varies over estimator and transform.

For the Arellano–Bond estimator with the first-differenced transform,

\[ Q_{zu} = \left( \sum_i Z_i' H_{ji} \hat{u}^*_{ji} \right) \]

For the Arellano–Bond estimator with the FOD transform,

\[ Q_{zu} = \left( \sum_i Z_i' Q_{fod} \right) \]

where

\[ Q_{fod} = \begin{pmatrix}
-\sqrt{\frac{T_i+1}{T_i}} & 0 & \cdots & 0 \\
\sqrt{\frac{T_i-1}{T_i}} & \sqrt{\frac{T_i}{T_i-1}} & \cdots & 0 \\
0 & \ddots & \ddots & \ddots \\
0 & \cdots & \sqrt{\frac{1}{2}} & -\sqrt{\frac{2}{1}}
\end{pmatrix} \hat{u}^*_{ji} \]

and * implies the first-differenced transform instead of the FOD transform.

For the Arellano–Bover/Blundell–Bond system estimator with the first-differenced transform,

\[ Q_{zu} = \left( \sum_i Z_i' \hat{\epsilon}_{ji} \hat{\epsilon}^{*'}_{ji} \hat{u}^*_{ji} \right) \]
After a one-step estimator, the Sargan test is

\[ S_1 = \frac{1}{\sigma_1^2} \left( \sum_i \hat{\epsilon}_{1i}' Z_i \right) A_1 \left( \sum_i Z_i' \hat{\epsilon}_{1i} \right) \]

The transformed two-step residuals are given by

\[ \hat{\epsilon}^*_{2i} = y^*_i - \hat{\beta}_2 X^*_i \]

and the level two-step residuals are given by

\[ \hat{\epsilon}^L_{2i} = y^L_i - \hat{\beta}_2 X^L_i \]

Stacking the residual vectors yields

\[ \hat{\epsilon}_{2i} = \begin{pmatrix} \hat{\epsilon}^*_{2i} \\ \hat{\epsilon}^L_{2i} \end{pmatrix} \]

After a two-step estimator, the Sargan test is

\[ S_2 = \left( \sum_i \hat{\epsilon}_{2i}' Z_i \right) A_2 \left( \sum_i Z_i' \hat{\epsilon}_{2i} \right) \]

Reference


Also see

[ XT] xtdpd — Linear dynamic panel-data estimation

[U] 20 Estimation and postestimation commands
Syntax

```
xtdpdsys depvar [ indepvars ] [ if ] [ in ] [ , options ]
```

**options**

<table>
<thead>
<tr>
<th>Model</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>noconstant</code></td>
<td>suppress constant term</td>
</tr>
<tr>
<td><code>lags(#)</code></td>
<td>use # lags of dependent variable as covariates; default is <code>lags(1)</code></td>
</tr>
<tr>
<td><code>maxldep(#)</code></td>
<td>maximum lags of dependent variable for use as instruments</td>
</tr>
<tr>
<td><code>maxlags(#)</code></td>
<td>maximum lags of predetermined and endogenous variables for use as instruments</td>
</tr>
<tr>
<td><code>twostep</code></td>
<td>compute the two-step estimator instead of the one-step estimator</td>
</tr>
</tbody>
</table>

| Predetermined              | predetermained variables; can be specified more than once |
| `pre(varlist[...])`       |                                                        |

| Endogenous                 | endogenous variables; can be specified more than once |
| `endogenous(varlist[...])`|                                                        |

| SE/Robust                  | `vcetype` may be `gmm` or `robust`                     |
| `vce(vcetype)`             |                                                        |

| Reporting                  |                                                        |
| `level(#)`                 | set confidence level; default is `level(95)`           |
| `artests(#)`               | use # as maximum order for AR tests; default is `artests(2)` |
| `display_options`          | control spacing and line width                         |

| `coeflegend`               | display legend instead of statistics                   |

A panel variable and a time variable must be specified; use [XT] `xtset`. `indepvars` and all `varlists`, except `pre(varlist[...])` and `endogenous(varlist[...])`, may contain time-series operators; see [U] 11.4.4 Time-series varlists. The specification of `depvar` may not contain time-series operators. `by`, `statsby`, and `xi` are allowed; see [U] 11.1.10 Prefix commands. `coeflegend` does not appear in the dialog box. See [U] 20 Estimation and postestimation commands for more capabilities of estimation commands.
Description

Linear dynamic panel-data models include $p$ lags of the dependent variable as covariates and contain unobserved panel-level effects, fixed or random. By construction, the unobserved panel-level effects are correlated with the lagged dependent variables, making standard estimators inconsistent. Arellano and Bond (1991) derived a consistent generalized method of moments (GMM) estimator for this model. The Arellano and Bond estimator can perform poorly if the autoregressive parameters are too large or the ratio of the variance of the panel-level effect to the variance of idiosyncratic error is too large. Building on the work of Arellano and Bover (1995), Blundell and Bond (1998) developed a system estimator that uses additional moment conditions; xtdpdsys implements this estimator.

This estimator is designed for datasets with many panels and few periods. This method assumes that there is no autocorrelation in the idiosyncratic errors and requires the initial condition that the panel-level effects be uncorrelated with the first difference of the first observation of the dependent variable.

Options

### Model

- **noconstant**: see [R] estimation options.
- **lags(#)**: sets $p$, the number of lags of the dependent variable to be included in the model. The default is $p = 1$.
- **maxldep(#)**: sets the maximum number of lags of the dependent variable that can be used as instruments. The default is to use all $T_i - p - 2$ lags.
- **maxlags(#)**: sets the maximum number of lags of the predetermined and endogenous variables that can be used as instruments. For predetermined variables, the default is to use all $T_i - p - 1$ lags. For endogenous variables, the default is to use all $T_i - p - 2$ lags.
- **twostep**: specifies that the two-step estimator be calculated.

### Predetermined

- **pre(varlist [ , lagstruct(prelags, premaxlags)] )**: specifies that a set of predetermined variables be included in the model. Optionally, you may specify that prelags lags of the specified variables also be included. The default for prelags is 0. Specifying premaxlags sets the maximum number of further lags of the predetermined variables that can be used as instruments. The default is to include $T_i - p - 1$ lagged levels as instruments for predetermined variables. You may specify as many sets of predetermined variables as you need within the standard Stata limits on matrix size. Each set of predetermined variables may have its own number of prelags and premaxlags.

### Endogenous

- **endogenous(varlist [ , lagstruct(endlags, endmaxlags)] )**: specifies that a set of endogenous variables be included in the model. Optionally, you may specify that endlags lags of the specified variables also be included. The default for endlags is 0. Specifying endmaxlags sets the maximum number of further lags of the endogenous variables that can be used as instruments. The default is to include $T_i - p - 2$ lagged levels as instruments for endogenous variables. You may specify as many sets of endogenous variables as you need within the standard Stata limits on matrix size. Each set of endogenous variables may have its own number of endlags and endmaxlags.
SE/Robust

tdpdsys — Arellano–Bover/Blundell–Bond linear dynamic panel-data estimation

vce(vcetype) specifies the type of standard error reported, which includes types that are derived from asymptotic theory and that are robust to some kinds of misspecification; see Methods and formulas in [XT] xtdpd.

vce(gmm), the default, uses the conventionally derived variance estimator for generalized method of moments estimation.

vce(robust) uses the robust estimator. For the one-step estimator, this is the Arellano–Bond robust VCE estimator. For the two-step estimator, this is the Windmeijer (2005) WC-robust estimator.

Reporting

level(#) specifies the maximum order of the autocorrelation test to be calculated. The tests are reported by estat abond; see [XT] xtdpdsys postestimation. Specifying the order of the highest test at estimation time is more efficient than specifying it to estat abond, because estat abond must refit the model to obtain the test statistics. The maximum order must be less than or equal the number of periods in the longest panel. The default is artests(2).

display_options: vsquish and nolstretch; see [R] estimation options.

The following option is available with xtdpdsys but is not shown in the dialog box:

coeflegend; see [R] estimation options.

Remarks and examples

If you have not read [XT] xtabond, you may want to do so before continuing.

Consider the dynamic panel-data model

\[ y_{it} = \sum_{j=1}^{p} \alpha_j y_{i,t-j} + x_{it} \beta_1 + w_{it} \beta_2 + \nu_i + \epsilon_{it} \quad i = 1, \ldots, N \quad t = 1, \ldots, T_i \]  

(1)

where

the \( \alpha_j \) are \( p \) parameters to be estimated,
\( x_{it} \) is a \( 1 \times k_1 \) vector of strictly exogenous covariates,
\( \beta_1 \) is a \( k_1 \times 1 \) vector of parameters to be estimated,
\( w_{it} \) is a \( 1 \times k_2 \) vector of predetermined or endogenous covariates,
\( \beta_2 \) is a \( k_2 \times 1 \) vector of parameters to be estimated,
\( \nu_i \) are the panel-level effects (which may be correlated with the covariates), and
\( \epsilon_{it} \) are i.i.d. over the whole sample with variance \( \sigma^2_{\epsilon} \).

The \( \nu_i \) and the \( \epsilon_{it} \) are assumed to be independent for each \( i \) over all \( t \).

By construction, the lagged dependent variables are correlated with the unobserved panel-level effects, making standard estimators inconsistent. With many panels and few periods, the Arellano–Bond estimator is constructed by first-differencing to remove the panel-level effects and using instruments to form moment conditions.
Blundell and Bond (1998) show that the lagged-level instruments in the Arellano–Bond estimator become weak as the autoregressive process becomes too persistent or the ratio of the variance of the panel-level effects $\nu_i$ to the variance of the idiosyncratic error $\epsilon_{it}$ becomes too large. Building on the work of Arellano and Bover (1995), Blundell and Bond (1998) proposed a system estimator that uses moment conditions in which lagged differences are used as instruments for the level equation in addition to the moment conditions of lagged levels as instruments for the differenced equation. The additional moment conditions are valid only if the initial condition $E[\nu_i \Delta y_{i2}] = 0$ holds for all $i$; see Blundell and Bond (1998) and Blundell, Bond, and Windmeijer (2000).

_xtdpdsys_ fits dynamic panel-data estimators with the Arellano–Bover/Blundell–Bond system estimator. Because _xtdpdsys_ extends _xtabond_, [XT] _xtabond_ provides useful background.

➢ Example 1: A dynamic panel model

In their article, Arellano and Bond (1991) apply their estimators and test statistics to a model of dynamic labor demand that had previously been considered by Layard and Nickell (1986), using data from an unbalanced panel of firms from the United Kingdom. All variables are indexed over the firm $i$ and time $t$. In this dataset, $n_{it}$ is the log of employment in firm $i$ at time $t$, $w_{it}$ is the natural log of the real product wage, $k_{it}$ is the natural log of the gross capital stock, and $y_{s_{it}}$ is the natural log of industry output. The model also includes time dummies $yr_{1980}$, $yr_{1981}$, $yr_{1982}$, $yr_{1983}$, and $yr_{1984}$.

For comparison, we begin by using _xtabond_ to fit a model to these data.
. use http://www.stata-press.com/data/r13/abdata
. xtabond n L(0/2).(w k) yr1980-yr1984 year, vce(robust)

Arellano–Bond dynamic panel-data estimation
Number of obs = 611
Group variable: id
Number of groups = 140
Time variable: year
Obs per group: min = 4
avg = 4.364286
max = 6
Number of instruments = 40
Wald chi2(13) = 1318.68
Prob > chi2 = 0.0000

One-step results
(Std. Err. adjusted for clustering on id)

<table>
<thead>
<tr>
<th></th>
<th>Robust</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coef. Std. Err.</td>
<td>z</td>
<td>P&gt;</td>
<td>z</td>
</tr>
<tr>
<td>n</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L1.</td>
<td>.6286618</td>
<td>.1161942</td>
<td>5.41</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>.4009254</td>
<td>.8563983</td>
<td></td>
<td></td>
</tr>
<tr>
<td>w</td>
<td>-.5104249</td>
<td>.1904292</td>
<td>-2.68</td>
<td>0.007</td>
</tr>
<tr>
<td></td>
<td>-.8836592</td>
<td>-.1371906</td>
<td></td>
<td></td>
</tr>
<tr>
<td>L1.</td>
<td>.2891446</td>
<td>.140946</td>
<td>2.05</td>
<td>0.040</td>
</tr>
<tr>
<td></td>
<td>.0128954</td>
<td>.5653937</td>
<td></td>
<td></td>
</tr>
<tr>
<td>L2.</td>
<td>-.0443653</td>
<td>.0768135</td>
<td>-0.58</td>
<td>0.564</td>
</tr>
<tr>
<td></td>
<td>-.194917</td>
<td>.1061865</td>
<td></td>
<td></td>
</tr>
<tr>
<td>k</td>
<td>.3556923</td>
<td>.0603274</td>
<td>5.90</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>.2374528</td>
<td>.4739318</td>
<td></td>
<td></td>
</tr>
<tr>
<td>L1.</td>
<td>-.0457102</td>
<td>.0699732</td>
<td>-0.65</td>
<td>0.514</td>
</tr>
<tr>
<td></td>
<td>-.1828552</td>
<td>.0914348</td>
<td></td>
<td></td>
</tr>
<tr>
<td>L2.</td>
<td>-.0619721</td>
<td>.0328589</td>
<td>-1.89</td>
<td>0.059</td>
</tr>
<tr>
<td></td>
<td>-.1263743</td>
<td>.0024301</td>
<td></td>
<td></td>
</tr>
<tr>
<td>yr1980</td>
<td>-.0282422</td>
<td>.0166363</td>
<td>-1.70</td>
<td>0.090</td>
</tr>
<tr>
<td></td>
<td>-.0608488</td>
<td>.0043643</td>
<td></td>
<td></td>
</tr>
<tr>
<td>yr1981</td>
<td>-.0694052</td>
<td>.028961</td>
<td>-2.40</td>
<td>0.017</td>
</tr>
<tr>
<td></td>
<td>-.1261677</td>
<td>-.0126426</td>
<td></td>
<td></td>
</tr>
<tr>
<td>yr1982</td>
<td>-.0523678</td>
<td>.0423433</td>
<td>-1.24</td>
<td>0.216</td>
</tr>
<tr>
<td></td>
<td>-.1353591</td>
<td>.0306235</td>
<td></td>
<td></td>
</tr>
<tr>
<td>yr1983</td>
<td>-.0256599</td>
<td>.0533747</td>
<td>-0.48</td>
<td>0.631</td>
</tr>
<tr>
<td></td>
<td>-.1302723</td>
<td>.0789525</td>
<td></td>
<td></td>
</tr>
<tr>
<td>yr1984</td>
<td>-.0093229</td>
<td>.0696241</td>
<td>-0.13</td>
<td>0.893</td>
</tr>
<tr>
<td></td>
<td>-.1457837</td>
<td>.1271379</td>
<td></td>
<td></td>
</tr>
<tr>
<td>year</td>
<td>.0019575</td>
<td>.0119481</td>
<td>0.16</td>
<td>0.870</td>
</tr>
<tr>
<td></td>
<td>-.0214604</td>
<td>.0253754</td>
<td></td>
<td></td>
</tr>
<tr>
<td>_cons</td>
<td>-2.543221</td>
<td>23.97919</td>
<td>-0.11</td>
<td>0.916</td>
</tr>
<tr>
<td></td>
<td>-49.54158</td>
<td>44.45514</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Instruments for differenced equation
GMM-type: L(2/.).n
D.yr1983 D.yr1984 D.year

Instruments for level equation
Standard: _cons
Now we fit the same model by using \texttt{xtdpdsys}:

\begin{verbatim}
. xtdpdsys n L(0/2).w k yr1980-yr1984 year, vce(robust)
\end{verbatim}

System dynamic panel-data estimation
Number of obs = 751
Group variable: id
Number of groups = 140
Time variable: year
Obs per group:
min = 5
avg = 5.364286
max = 7
Number of instruments = 47
Wald chi2(13) = 2579.96
Prob > chi2 = 0.0000

One-step results

\begin{verbatim}

| Coef. | Std. Err. | z    | P>|z| | 95% Conf. Interval |
|-------|-----------|------|------|-------------------|
| n     | .8221535  | .093387 | 8.80 | 0.000 | .6391184 - 1.005189 |
| L1.   | -1.5427935| .1881721| -2.88| 0.004 | -.911604 - .1793831 |
| w     | -3.703602 | .1656364| 2.24 | 0.025 | .0457189 - .6950015 |
| L1.   | .0726314  | .0907148| -0.80| 0.423 | -.2504292 - .1051664 |
| L2.   | -3.0726314| .1656364| 2.24 | 0.025 | .0457189 - .6950015 |
| k     | -3.638069 | .0657524| 5.53 | 0.000 | .2349346 - .4926792 |
| L1.   | -1.1222996| .0701521| -1.74| 0.081 | -.2597951 - .015196 |
| L2.   | -3.001355 | .0344142| 2.62 | 0.009 | -.1575862 - .0226849 |
| yr1980| -.0308622 | .016946 | -1.82| 0.069 | -.0640757 - .0023512 |
| yr1981| -.0718417 | .0293223| -2.45| 0.014 | -.1293123 - .014371 |
| yr1982| -0.0384806| .0373631| -1.03| 0.303 | -.1117111 - .0347998 |
| yr1983| -.0121768 | .0498519| -0.24| 0.807 | -.1098847 - .0855311 |
| yr1984| -.0050903 | .0555011| -0.08| 0.938 | -.1334701 - .1232895 |
| year  | .0058631  | .0119887| 0.49 | 0.625 | -.0176304 - .0293566 |
| _cons | -10.59198 | 23.92087| -0.44| 0.658 | -.57.47602 - 36.29207 |

Instruments for differenced equation
GMM-type: L(2/2)\(,n\)

Instruments for level equation
GMM-type: LD.n
Standard: _cons

\end{verbatim}

If you are unfamiliar with the \texttt{L(\()\(,\)\) notation, see [U] \ref{time-series-operators}. That the system estimator produces a much higher estimate of the coefficient on lagged employment agrees with the results in Blundell and Bond (1998), who show that the system estimator does not have the downward bias that the Arellano–Bond estimator has when the true value is high.

Comparing the footers illustrates the difference between the two estimators; \texttt{xtdpdsys} includes lagged differences of \(n\) as instruments for the level equation, whereas \texttt{xtabond} does not. Comparing the headers shows that \texttt{xtdpdsys} has seven more instruments than \texttt{xtabond}. (As it should; there are 7 observations on LD.\(n\) available in the complete panels that run from 1976–1984, after accounting for the first two years that are lost because the model has two lags.) Only the first lags of the variables are used because the moment conditions using higher lags are redundant; see Blundell and Bond (1998) and Blundell, Bond, and Windmeijer (2000).

\texttt{estat abond} reports the Arellano–Bond test for serial correlation in the first-differenced errors. The moment conditions are valid only if there is no serial correlation in the idiosyncratic errors.
Because the first difference of independently and identically distributed idiosyncratic errors will be autoregressed, rejecting the null hypothesis of no serial correlation at order one in the first-differenced errors does not imply that the model is misspecified. Rejecting the null hypothesis at higher orders implies that the moment conditions are not valid. See [XT] xtdpd for an alternative estimator in this case.

. estat abond

Arellano-Bond test for zero autocorrelation in first-differenced errors

<table>
<thead>
<tr>
<th>Order</th>
<th>z</th>
<th>Prob &gt; z</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-4.6414</td>
<td>0.0000</td>
</tr>
<tr>
<td>2</td>
<td>-1.0572</td>
<td>0.2904</td>
</tr>
</tbody>
</table>

H0: no autocorrelation

The above output does not present evidence that the model is misspecified.

Example 2: Including predetermined covariates

Sometimes we cannot assume strict exogeneity. Recall that a variable \( x_{it} \) is said to be strictly exogenous if \( E[x_{it}\epsilon_{is}] = 0 \) for all \( t \) and \( s \). If \( E[x_{it}\epsilon_{is}] \neq 0 \) for \( s < t \) but \( E[x_{it}\epsilon_{is}] = 0 \) for all \( s \geq t \), the variable is said to be predetermined. Intuitively, if the error term at time \( t \) has some feedback on the subsequent realizations of \( x_{it} \), \( x_{it} \) is a predetermined variable. Because unforecastable errors today might affect future changes in the real wage and in the capital stock, we might suspect that the log of the real product wage and the log of the gross capital stock are predetermined instead of strictly exogenous.
. xtdpdsys n yr1980-yr1984 year, pre(w k, lag(2, .)) vce(robust)

System dynamic panel-data estimation

<table>
<thead>
<tr>
<th>Number of obs</th>
<th>751</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group variable: id</td>
<td>140</td>
</tr>
<tr>
<td>Time variable: year</td>
<td></td>
</tr>
<tr>
<td>Obs per group: min</td>
<td>5</td>
</tr>
<tr>
<td>avg</td>
<td>5.364286</td>
</tr>
<tr>
<td>max</td>
<td>7</td>
</tr>
<tr>
<td>Number of instruments</td>
<td>95</td>
</tr>
<tr>
<td>Wald chi2(13)</td>
<td>7562.80</td>
</tr>
<tr>
<td>Prob &gt; chi2</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

One-step results

| n | Coef. | Std. Err. | z | P>|z| | [95% Conf. Interval] |
|---|-------|-----------|---|-----|------------------|
| L1. | .913278 | .0460602 | 19.83 | 0.000 | .8230017 1.003554 |
| w | --. | -.728159 | .1019044 | -7.15 | 0.000 | -.927888 -.5284301 |
| L1. | .5602737 | .1939617 | 2.89 | 0.004 | .1801156 .9404317 |
| L2. | -.0523028 | .1487653 | -0.35 | 0.725 | -.3438775 .2392718 |
| k | --. | .4820097 | .0760787 | 6.34 | 0.000 | .3328983 .6311212 |
| L1. | -.2846944 | .0831902 | -3.42 | 0.001 | -.4477442 -.1216446 |
| L2. | -.1394181 | .0405709 | -3.44 | 0.001 | -.2189356 -.0599006 |
| yr1980 | -.0325146 | .0216371 | -1.50 | 0.133 | -.0749226 .0098935 |
| yr1981 | -.0726116 | .0346482 | -2.10 | 0.036 | -.1405207 -.0047024 |
| yr1982 | -.0477038 | .0451914 | -1.06 | 0.291 | -.1362772 .0408696 |
| yr1983 | -.0396264 | .0558734 | -0.71 | 0.478 | -.1491362 .0698835 |
| yr1984 | -.0810383 | .0736648 | -1.10 | 0.271 | -.2254186 .063342 |
| year | .0192741 | .0145326 | 1.33 | 0.185 | -.0092092 .0477574 |
| _cons | -37.34972 | 28.77747 | -1.30 | 0.194 | -93.75253 19.05308 |

Instruments for differenced equation

GMM-type: L(2/.)n L(1/.)L2.w L(1/.)L2.k

Instruments for level equation

GMM-type: L. n L2.D.w L2.D.k
Standard: _cons

The footer informs us that we are now including GMM-type instruments from the first lag of L.w on back and from the first lag of L2.k on back for the differenced errors and the second lags of the differences of w and k as instruments for the level errors.

Technical note

The above example illustrates that xtdpdsys understands pre(w k, lag(2, .)) to mean that L2.w and L2.k are predetermined variables. This is a stricter definition than the alternative that pre(w k, lag(2, .)) means only that w k are predetermined but to include two lags of w and two lags of k in the model. If you prefer the weaker definition, xtdpdsys still gives you consistent estimates, but it is not using all possible instruments; see [XT] xtdpd for an example of how to include all possible instruments.
Stored results

xtdpdsys stores the following in e():

Scalars
- e(N) number of observations
- e(N_g) number of groups
- e(df_m) model degrees of freedom
- e(g_min) smallest group size
- e(g_avg) average group size
- e(g_max) largest group size
- e(t_min) minimum time in sample
- e(t_max) maximum time in sample
- e(chi2) \( \chi^2 \)
- e(arm#) test for autocorrelation of order #
- e(artests) number of AR tests computed
- e(sig2) estimate of \( \sigma^2 \)
- e(rss) sum of squared differenced residuals
- e(sargan) Sargan test statistic
- e(rank) rank of e(V)
- e(zrank) rank of instrument matrix

Macros
- e(cmd) xtdpdsys
- e(cmdline) command as typed
- e(depvar) name of dependent variable
- e(twostep) twostep, if specified
- e(ivar) variable denoting groups
- e(tvar) variable denoting time within groups
- e(vce) vcetype specified in vce()
- e(vcetype) title used to label Std. Err.
- e(system) system, if system estimator
- e(hascons) hascons, if specified
- e(transform) specified transform
- e(datasignature) checksum from datasignature
- e(properties) b V
- e(estat_cmd) program used to implement estat
- e(predict) program used to implement predict
- e(marginsok) predictions allowed by margins

Matrices
- e(b) coefficient vector
- e(V) variance–covariance matrix of the estimators

Functions
- e(sample) marks estimation sample

Methods and formulas

xtdpdsys uses xtdpd to perform its computations, so the formulas are given in Methods and formulas of [XT] xtdpd.

Acknowledgment

We thank David Roodman of the Center for Global Development, who wrote xtabond2.
References


Also see

[XT] xtdpdsys postestimation — Postestimation tools for xtdpdsys

[XT] xtset — Declare data to be panel data

[XT] xtabond — Arellano–Bond linear dynamic panel-data estimation

[XT] xtdpd — Linear dynamic panel-data estimation

[XT] xtitreg — Instrumental variables and two-stage least squares for panel-data models

[XT] xtreg — Fixed-, between-, and random-effects and population-averaged linear models

[XT] xtregar — Fixed- and random-effects linear models with an AR(1) disturbance

[U] 20 Estimation and postestimation commands
## Description

The following postestimation commands are of special interest after `xtdpdsys`:

<table>
<thead>
<tr>
<th>Command</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>estat abond</code></td>
<td>test for autocorrelation</td>
</tr>
<tr>
<td><code>estat sargan</code></td>
<td>Sargan test of overidentifying restrictions</td>
</tr>
</tbody>
</table>

The following standard postestimation commands are also available:

<table>
<thead>
<tr>
<th>Command</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>estat summarize</code></td>
<td>summary statistics for the estimation sample</td>
</tr>
<tr>
<td><code>estat vce</code></td>
<td>variance–covariance matrix of the estimators (VCE)</td>
</tr>
<tr>
<td><code>estimates</code></td>
<td>cataloging estimation results</td>
</tr>
<tr>
<td><code>forecast</code></td>
<td>dynamic forecasts and simulations</td>
</tr>
<tr>
<td><code>lincom</code></td>
<td>point estimates, standard errors, testing, and inference for linear combinations of coefficients</td>
</tr>
<tr>
<td><code>margins</code></td>
<td>marginal means, predictive margins, marginal effects, and average marginal effects</td>
</tr>
<tr>
<td><code>marginsplot</code></td>
<td>graph the results from margins (profile plots, interaction plots, etc.)</td>
</tr>
<tr>
<td><code>nlcom</code></td>
<td>point estimates, standard errors, testing, and inference for nonlinear combinations of coefficients</td>
</tr>
<tr>
<td><code>predict</code></td>
<td>predictions, residuals, influence statistics, and other diagnostic measures</td>
</tr>
<tr>
<td><code>predictnl</code></td>
<td>point estimates, standard errors, testing, and inference for generalized predictions</td>
</tr>
<tr>
<td><code>test</code></td>
<td>Wald tests of simple and composite linear hypotheses</td>
</tr>
<tr>
<td><code>testnl</code></td>
<td>Wald tests of nonlinear hypotheses</td>
</tr>
</tbody>
</table>

### Special-interest postestimation commands

- `estat abond` reports the Arellano–Bond test for serial correlation in the first-differenced residuals.
- `estat sargan` reports the Sargan test of the overidentifying restrictions.
Syntax for predict

\[ \text{predict [type] newvar [if] [in]} [, \text{xb e stdp difference}] \]

Menu for predict

Statistics > Postestimation > Predictions, residuals, etc.

Options for predict

- **Main**
  - `xb`, the default, calculates the linear prediction.
  - `e` calculates the residual error.
  - `stdp` calculates the standard error of the prediction, which can be thought of as the standard error of the predicted expected value or mean for the observation’s covariate pattern. The standard error of the prediction is also referred to as the standard error of the fitted value. `stdp` may not be combined with `difference`.
  - `difference` specifies that the statistic be calculated for the first differences instead of the levels, the default.

Syntax for estat

**Test for autocorrelation**

\[ \text{estat abond} [, \text{artests(#)}] \]

**Sargan test of overidentifying restrictions**

\[ \text{estat sargan} \]

Menu for estat

Statistics > Postestimation > Reports and statistics

Option for estat abond

`artests(#)` specifies highest order of serial correlation to be tested. By default, the tests computed during estimation are reported. The model will be refit when `artests(#)` specifies a higher order than that computed during the original estimation. The model can only be refit if the data have not changed.

Remarks and examples

Remarks are presented under the following headings:

- `estat abond`
- `estat sargan`
The moment conditions used by `xtdpdsys` are valid only if there is no serial correlation in the idiosyncratic errors. Testing for serial correlation in dynamic panel-data models is tricky because a transform is required to remove the panel-level effects, but the transformed errors have a more complicated error structure than that of the idiosyncratic errors. The Arellano–Bond test for serial correlation reported by `estat abond` tests for serial correlation in the first-differenced errors.

Because the first difference of independently and identically distributed idiosyncratic errors will be serially correlated, rejecting the null hypothesis of no serial correlation in the first-differenced errors at order one does not imply that the model is misspecified. Rejecting the null hypothesis at higher orders implies that the moment conditions are not valid. See example 5 in [XT] `xtdpd` for an alternative estimator that allows for idiosyncratic errors that follow a first-order moving average process.

After the one-step system estimator, the test can be computed only when `vce(robust)` has been specified.

Like all GMM estimators, the estimator in `xtdpdsys` can produce consistent estimates only if the moment conditions used are valid. Although there is no method to test if the moment conditions from an exactly identified model are valid, one can test whether the overidentifying moment conditions are valid. `estat sargan` implements the Sargan test of overidentifying conditions discussed in Arellano and Bond (1991).

Only for a homoskedastic error term does the Sargan test have an asymptotic chi-squared distribution. In fact, Arellano and Bond (1991) show that the one-step Sargan test overrejects in the presence of heteroskedasticity. Because its asymptotic distribution is not known under the assumptions of the `vce(robust)` model, `xtdpdsys` does not compute it when `vce(robust)` is specified. See [XT] `xtdpd` for an example in which the null hypothesis of the Sargan test is not rejected.

```
. use http://www.stata-press.com/data/r13/abdata
. xtdpdsys n L(0/2).(w k) yr1980-yr1984 year
(output omitted)
. estat sargan
Sargan test of overidentifying restrictions
H0: overidentifying restrictions are valid
    chi2(33) =  63.63911
    Prob > chi2 =  0.0011
```

The output above presents strong evidence against the null hypothesis that the overidentifying restrictions are valid. Rejecting this null hypothesis implies that we need to reconsider our model or our instruments, unless we attribute the rejection to heteroskedasticity in the data-generating process. Although performing the Sargan test after the two-step estimator is an alternative, Arellano and Bond (1991) found a tendency for this test to underreject in the presence of heteroskedasticity.

**Methods and formulas**

The formulas are given in *Methods and formulas* of [XT] `xtdpd postestimation`. 
Reference


Also see

[XT] xtdpdsys — Arellano–Bover/Blundell–Bond linear dynamic panel-data estimation

[U] 20 Estimation and postestimation commands
Title

**xtfrontier — Stochastic frontier models for panel data**

Syntax

---

**Time-invariant model**

```
xfrontier  depvar  [indepvars]  [if]  [in]  [weight],  ti  [ti_options]
```

**Time-varying decay model**

```
xfrontier  depvar  [indepvars]  [if]  [in]  [weight],  tvd  [tvd_options]
```

**ti_options**

---

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>noconstant</code></td>
<td>suppress constant term</td>
</tr>
<tr>
<td><code>ti</code></td>
<td>use time-invariant model</td>
</tr>
<tr>
<td><code>cost</code></td>
<td>fit cost frontier model</td>
</tr>
<tr>
<td><code>constraints(constraints)</code></td>
<td>apply specified linear constraints</td>
</tr>
<tr>
<td><code>collinear</code></td>
<td>keep collinear variables</td>
</tr>
</tbody>
</table>

**SE**

```
vce(vcetype)
```

`vcetype` may be `oim`, `bootstrap`, or `jackknife`

**Reporting**

```
level(#)
```

set confidence level; default is `level(95)`

```
nocnsreport
display_options
```

control column formats, row spacing, line width, display of omitted variables and base and empty cells, and factor-variable labeling

**Maximization**

```
maximize_options
```

control the maximization process; seldom used

```
coeflegend
```

display legend instead of statistics
xtfrontier — Stochastic frontier models for panel data

### tvc_options

<table>
<thead>
<tr>
<th>Model</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>noconstant</td>
<td>suppress constant term</td>
</tr>
<tr>
<td>tvd</td>
<td>use time-varying decay model</td>
</tr>
<tr>
<td>cost</td>
<td>fit cost frontier model</td>
</tr>
<tr>
<td>constraints(Constraints)</td>
<td>apply specified linear constraints</td>
</tr>
<tr>
<td>collinear</td>
<td>keep collinear variables</td>
</tr>
</tbody>
</table>

SE

vce(vcetype)    vcetype may be oim, bootstrap, or jackknife

### Reporting

<table>
<thead>
<tr>
<th>vce(vcetype)</th>
<th>level(#)</th>
<th>nocnsreport</th>
<th>display_options</th>
</tr>
</thead>
<tbody>
<tr>
<td>vcetype may be oim, bootstrap, or jackknife</td>
<td>set confidence level; default is level(95)</td>
<td>do not display constraints</td>
<td>control column formats, row spacing, line width, display of omitted variables and base and empty cells, and factor-variable labeling</td>
</tr>
</tbody>
</table>

Maximization

| maximize_options                | control the maximization process; seldom used |
| coeflegend                       | display legend instead of statistics |

A panel variable must be specified. For xtfrontier, tvd, a time variable must also be specified. Use xtset; see [XT] xtset.

indepvars may contain factor variables; see [U] 11.4.3 Factor variables.

depvar and indepvars may contain time-series operators; see [U] 11.4.4 Time-series varlists.

by, fp, and statsby are allowed; see [U] 11.1.10 Prefix commands.

fweights and iweights are allowed; see [U] 11.1.6 weight. Weights must be constant within panel.

coeftrend does not appear in the dialog box.

See [U] 20 Estimation and postestimation commands for more capabilities of estimation commands.

**Menu**

Statistics > Longitudinal/panel data > Frontier models

**Description**

xtfrontier fits stochastic production or cost frontier models for panel data. More precisely, xtfrontier estimates the parameters of a linear model with a disturbance generated by specific mixture distributions.

The disturbance term in a stochastic frontier model is assumed to have two components. One component is assumed to have a strictly nonnegative distribution, and the other component is assumed to have a symmetric distribution. In the econometrics literature, the nonnegative component is often referred to as the **inefficiency term**, and the component with the symmetric distribution as the **idiosyncratic error**. xtfrontier permits two different parameterizations of the inefficiency term: a time-invariant model and the Battese–Coelli (1992) parameterization of time effects. In the time-invariant model, the inefficiency term is assumed to have a truncated-normal distribution. In the Battese–Coelli (1992) parameterization of time effects, the inefficiency term is modeled as a truncated-normal random variable multiplied by a specific function of time. In both models, the
Idiosyncratic error term is assumed to have a normal distribution. The only panel-specific effect is the random inefficiency term.

See Kumbhakar and Lovell (2000) for a detailed introduction to frontier analysis.

**Options for time-invariant model**

- noconstant; see [R] estimation options.
- ti specifies that the parameters of the time-invariant technical inefficiency model be estimated.
- cost specifies that the frontier model be fit in terms of a cost function instead of a production function. By default, xtfrontier fits a production frontier model.
- constraints(constraints), collinear; see [R] estimation options.
- vce(vcetype) specifies the type of standard error reported, which includes types that are derived from asymptotic theory (oim) and that use bootstrap or jackknife methods (bootstrap, jackknife); see [XT] vce_options.
- level(#); see [R] estimation options.
- nocnsreport; see [R] estimation options.
- display_options: noomitted, vsquish, noemptycells, baselevels, allbaselevels, nofvlabel, fvwrap(#), fvwrapon(style), cformat(%fmt), pformat(%fmt), sformat(%fmt), and nolstretch; see [R] estimation options.
- maximize_options: difficult, technique(algorithm_spec) iterate(#), [no] log, trace, gradient, showstep, hessian, showtolerance, tolerance(#), ltolerance(#), nrtolerance(#), nonrtolerance, and from(init_specs); see [R] maximize. These options are seldom used.

The following option is available with xtfrontier but is not shown in the dialog box: coeflegend; see [R] estimation options.

**Options for time-varying decay model**

- noconstant; see [R] estimation options.
- tvd specifies that the parameters of the time-varying decay model be estimated.
- cost specifies that the frontier model be fit in terms of a cost function instead of a production function. By default, xtfrontier fits a production frontier model.
- constraints(constraints), collinear; see [R] estimation options.
vce(vcetype) specifies the type of standard error reported, which includes types that are derived from asymptotic theory (oim) and that use bootstrap or jackknife methods (bootstrap, jackknife); see [XT] vce_options.

Reporting

level(#) ; see [R] estimation options.
nocnsreport; see [R] estimation options.
display_options: noomitted, vsquish, noemptycells, baselevels, allbaselevels, nolabel, fvwrap(#), fvwrapon(style), cformat(%fmt), pformat(%fmt), sformat(%fmt), and nolstretch; see [R] estimation options.

Maximization

maximize_options: difficult, technique(algorithm_spec), iterate(#), [no]log, trace, gradient, showstep, hessian, showtolerance, tolerance(#), ltolerance(#), nrtolerance(#), nonrtolerance, and from(init_specs); see [R] maximize. These options are seldom used.

The following option is available with xt Frontier but is not shown in the dialog box:

coeflegend; see [R] estimation options.

Remarks and examples

Remarks are presented under the following headings:

Introduction
Time-invariant model
Time-varying decay model

Introduction

Stochastic production frontier models were introduced by Aigner, Lovell, and Schmidt (1977) and Meeusen and van den Broeck (1977). Since then, stochastic frontier models have become a popular subfield in econometrics; see Kumbhakar and Lovell (2000) for an introduction. xt Frontier fits two stochastic frontier models with distinct specifications of the inefficiency term and can fit both production- and cost-frontier models.

Let’s review the nature of the stochastic frontier problem. Suppose that a producer has a production function \( f(z_{it}, \beta) \). In a world without error or inefficiency, in time \( t \), the \( i \)th firm would produce

\[
q_{it} = f(z_{it}, \beta)
\]

A fundamental element of stochastic frontier analysis is that each firm potentially produces less than it might because of a degree of inefficiency. Specifically,

\[
q_{it} = f(z_{it}, \beta)\xi_{it}
\]
where $\xi_{it}$ is the level of efficiency for firm $i$ at time $t$; $\xi_i$ must be in the interval $(0, 1]$. If $\xi_{it} = 1$, the firm is achieving the optimal output with the technology embodied in the production function $f(z_{it}, \beta)$. When $\xi_{it} < 1$, the firm is not making the most of the inputs $z_{it}$ given the technology embodied in the production function $f(z_{it}, \beta)$. Because the output is assumed to be strictly positive (that is, $q_{it} > 0$), the degree of technical efficiency is assumed to be strictly positive (that is, $\xi_{it} > 0$).

Output is also assumed to be subject to random shocks, implying that

$$q_{it} = f(z_{it}, \beta)\xi_{it}\exp(v_{it})$$

Taking the natural log of both sides yields

$$\ln(q_{it}) = \ln\{f(z_{it}, \beta)\} + \ln(\xi_{it}) + v_{it}$$

Assuming that there are $k$ inputs and that the production function is linear in logs, defining $u_{it} = -\ln(\xi_{it})$ yields

$$\ln(q_{it}) = \beta_0 + \sum_{j=1}^{k} \beta_j \ln(z_{jit}) + v_{it} - u_{it} \tag{1}$$

Because $u_{it}$ is subtracted from $\ln(q_{it})$, restricting $u_{it} \geq 0$ implies that $0 < \xi_{it} \leq 1$, as specified above.

Kumbhakar and Lovell (2000) provide a detailed version of this derivation, and they show that performing an analogous derivation in the dual cost function problem allows us to specify the problem as

$$\ln(c_{it}) = \beta_0 + \beta_q \ln(q_{it}) + \sum_{j=1}^{k} \beta_j \ln(p_{jit}) + v_{it} - s u_{it} \tag{2}$$

where $q_{it}$ is output, the $z_{jit}$ are input quantities, $c_{it}$ is cost, the $p_{jit}$ are input prices, and

$$s = \begin{cases} 
1, & \text{for production functions} \\
-1, & \text{for cost functions}
\end{cases}$$

Intuitively, the inefficiency effect is required to lower output or raise expenditure, depending on the specification.

### Technical note

The model that xtfrontier actually fits has the form

$$y_{it} = \beta_0 + \sum_{j=1}^{k} \beta_j x_{jit} + v_{it} - s u_{it}$$

so in the context of the discussion above, $y_{it} = \ln(q_{it})$ and $x_{jit} = \ln(z_{jit})$ for a production function; for a cost function, $y_{it} = \ln(c_{it})$, the $x_{jit}$ are the $\ln(p_{jit})$, and $\ln(q_{it})$. You must perform the natural logarithm transformation of the data before estimation to interpret the estimation results correctly for a stochastic frontier production or cost model. xtfrontier does not perform any transformations on the data.
Equation (2) is a variant of a panel-data model in which \( v_{it} \) is the idiosyncratic error and \( u_{it} \) is a time-varying panel-level effect. Much of the literature on this model has focused on deriving estimators for different specifications of the \( u_{it} \) term. Kumbhakar and Lovell (2000) provide a survey of this literature.

`xtfrontier` provides estimators for two different specifications of \( u_{it} \). To facilitate the discussion, let \( N^+(\mu,\sigma^2) \) denote the truncated-normal distribution, which is truncated at zero with mean \( \mu \) and variance \( \sigma^2 \), and let \( \text{iid} \) stand for independently and identically distributed.

Consider the simplest specification in which \( u_{it} \) is a time-invariant truncated-normal random variable. In the time-invariant model, \( u_{it} = u_i, u_i \text{iid} \sim N^+(\mu,\sigma^2_u), v_{it} \text{iid} \sim N(0,\sigma^2_v) \), and \( u_i \) and \( v_{it} \) are distributed independently of each other and the covariates in the model. Specifying the `ti` option causes `xtfrontier` to estimate the parameters of this model.

In the time-varying decay specification,

\[
\begin{align*}
    u_{it} &= \exp\{-\eta(t - T_i)\} u_i
\end{align*}
\]

where \( T_i \) is the last period in the \( i \)th panel, \( \eta \) is the decay parameter, \( u_i \text{iid} \sim N^+(\mu,\sigma^2_u), v_{it} \text{iid} \sim N(0,\sigma^2_v) \), and \( u_i \) and \( v_{it} \) are distributed independently of each other and the covariates in the model. Specifying the `tvd` option causes `xtfrontier` to estimate the parameters of this model.

### Time-invariant model

#### Example 1

`xtfrontier`, `ti` provides maximum likelihood estimates for the parameters of the time-invariant decay model. In this model, the inefficiency effects are modeled as \( u_{it} = u_i, u_i \text{iid} \sim N^+(\mu,\sigma^2_u), v_{it} \text{iid} \sim N(0,\sigma^2_v) \), and \( u_i \) and \( v_{it} \) are distributed independently of each other and the covariates in the model. In this example, firms produce a product called a widget, using a constant-returns-to-scale technology. We have 948 observations—91 firms, with 6–14 observations per firm. Our dataset contains variables representing the quantity of widgets produced, the number of machine hours used in production, the number of labor hours used in production, and three additional variables that are the natural logarithm transformations of the three aforementioned variables.
We fit a time-invariant model using the transformed variables:

```
use http://www.stata-press.com/data/r13/xtfrontier1
xtfrontier lnnwidgets lnmachines lnnworkers, ti
```

Iteration 0: log likelihood = -1473.8703  
Iteration 1: log likelihood = -1473.0565  
Iteration 2: log likelihood = -1472.6155  
Iteration 3: log likelihood = -1472.6074  
Iteration 4: log likelihood = -1472.6069  

Time-invariant inefficiency model  
Number of obs = 948  
Group variable: id  
Number of groups = 91  
Obs per group: min = 6  
avg = 10.4  
max = 14  
Wald chi2(2) = 661.76  
Log likelihood = -1472.6069  
Prob > chi2 = 0.0000  

|       | Coef. | Std. Err. | z    | P>|z| | [95% Conf. Interval] |
|-------|-------|-----------|------|------|----------------------|
| lnnwidgets | .2904551 | .0164219 | 17.69 | 0.000 | .2582688 -.3226415 |
| lnmachines | .2943333 | .0154352 | 19.07 | 0.000 | .2640808 .3245858  |
| lnnworkers | 3.030983 | .1441022 | 21.03 | 0.000 | 2.748548 3.313418 |
| /mu | 1.125667 | .6479217 | 1.74  | 0.082 | -.144236 2.39557  |
| /lnsigma2 | 1.421979 | .2672745 | 5.32  | 0.000 | .898131 1.945828 |
| /ilgtgamma | 1.138685 | .3562642 | 3.20  | 0.001 | .4404204 1.83695 |
| sigma2 | 4.145318 | 1.107938 | 3.70  | 0.000 | 2.946445 5.34419 |
| gamma | .7574382 | .0654548 | 11.89 | 0.000 | .6283952 .8864812 |
| sigma_u2 | 3.139822 | 1.107938 | 2.87  | 0.004 | 1.140334 5.13931 |
| sigma_v2 | 1.005496 | .0484143 | 20.80 | 0.000 | .906055 1.104936 |

In addition to the coefficients, the output reports estimates for the parameters sigma_v2, sigma_u2, gamma, sigma2, ilgtgamma, lnsigma2, and mu. sigma_v2 is the estimate of $\sigma^2_v$. sigma_u2 is the estimate of $\sigma^2_u$. gamma is the estimate of $\gamma = \sigma^2_u/\sigma^2_S$. sigma2 is the estimate of $\sigma^2_S = \sigma^2_v + \sigma^2_u$. Because $\gamma$ must be between 0 and 1, the optimization is parameterized in terms of the inverse logit of $\gamma$, and this estimate is reported as ilgtgamma. Because $\sigma^2_S$ must be positive, the optimization is parameterized in terms of $\ln(\sigma^2_S)$, and this estimate is reported as lnsigma2. Finally, mu is the estimate of $\mu$.

Technique note

Our simulation results indicate that this estimator requires relatively large samples to achieve any reasonable degree of precision in the estimates of $\mu$ and $\sigma^2_u$.

Time-varying decay model

`xtfrontier, tvd` provides maximum likelihood estimates for the parameters of the time-varying decay model. In this model, the inefficiency effects are modeled as

$$ u_{it} = \exp\{-\eta(t - T_i)\}u_i $$

where $u_i \sim N^+(\mu, \sigma^2_u)$.
When \( \eta > 0 \), the degree of inefficiency decreases over time; when \( \eta < 0 \), the degree of inefficiency increases over time. Because \( t = T_i \) in the last period, the last period for firm \( i \) contains the base level of inefficiency for that firm. If \( \eta > 0 \), the level of inefficiency decays toward the base level. If \( \eta < 0 \), the level of inefficiency increases to the base level.

Example 2

When \( \eta = 0 \), the time-varying decay model reduces to the time-invariant model. The following example illustrates this property and demonstrates how to specify constraints and starting values in these models.

Let's begin by fitting the time-varying decay model on the same data that were used in the previous example for the time-invariant model.

\[
\text{. xtf} \text{rontier lnwidgets lnmachines lnworkers, tvd}
\]

Iteration 0: log likelihood = -1551.3798 (not concave)
Iteration 1: log likelihood = -1502.2637
Iteration 2: log likelihood = -1476.3093 (not concave)
Iteration 3: log likelihood = -1472.9845
Iteration 4: log likelihood = -1472.5365
Iteration 5: log likelihood = -1472.529
Iteration 6: log likelihood = -1472.5289

Time-varying decay inefficiency model

<table>
<thead>
<tr>
<th>Number of obs = 948</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group variable: id</td>
</tr>
<tr>
<td>Number of groups = 91</td>
</tr>
<tr>
<td>Time variable: t</td>
</tr>
<tr>
<td>Obs per group: min = 6</td>
</tr>
<tr>
<td>avg = 10.4</td>
</tr>
<tr>
<td>max = 14</td>
</tr>
<tr>
<td>Wald chi2(2) = 661.93</td>
</tr>
</tbody>
</table>

Log likelihood = -1472.5289

| Probs > chi2 = 0.0000 |

| lnwidgets | Coef.   | Std. Err. | z      | P>|z|  | [95% Conf. Interval] |
|-----------|---------|-----------|--------|------|----------------------|
| lnmachines | 0.2907555 | 0.0164376 | 17.69  | 0.000 | .2585384 -.3229725 |
| lnworkers  | 0.2942412 | 0.0154373 | 19.06  | 0.000 | .2639846 -.3244978 |
| _cons     | 3.028939  | 0.1436046 | 21.09  | 0.000 | 2.74748 3.310399 |

| /mu | 1.110831 | 0.6452809 | 1.72  | 0.085 | -.1538967 2.375558 |
| /eta| .0016764 | .00425   | 0.39  | 0.693 | -.0066535 .0100064 |
| /lnsigma2 | 1.410723 | .2679485 | 5.26  | 0.000 | .885554 1.935893 |
| /ilgtgamma | 1.123982 | .3584243 | 3.14  | 0.002 | .4214828 1.82648 |

| sigma2  | 4.098919  | 1.098299  | 2.424327 | 6.930228 |
| gamma   | .7547265  | .0663495  | .603838  | .8613419 |
| sigma_u2| 3.093563  | 1.097606  | .9422943 | 5.244832 |
| sigma_v2| 1.005356  | 0.0484079 | .9104785 | 1.100234 |

The estimate of \( \eta \) is close to zero, and the other estimates are not too far from those of the time-invariant model.

We can use constraint to constrain \( \eta = 0 \) and obtain the same results produced by the time-invariant model. Although there is only one statistical equation to be estimated in this model, the model fits five of Stata's [R] ml equations; see [R] ml or Gould, Pitblado, and Poi (2010). The equation names can be seen by listing the matrix of estimated coefficients.
To constrain a parameter to a particular value in any equation, except the first equation, you must specify both the equation name and the parameter name by using the syntax

```
constraint # [eqname]_b[varname] = value  or  constraint # [eqname]coefficient = value
```

where `eqname` is the equation name, `varname` is the name of the variable in a linear equation, and `coefficient` refers to any parameter that has been estimated. More elaborate specifications with expressions are possible; see the example with constant returns to scale below, and see \[ R \] `constraint` for general reference.

Suppose that we impose the constraint \( \eta = 0 \); we get the same results as those reported above for the time-invariant model, except for some minute differences attributable to an alternate convergence path in the optimization.

```
constraint 1 [eta]_cons = 0
xtfrontier lnwidgets lnmachines lnworkers, tvd constraints(1)
```

```
 Iteration 0:  log likelihood = -1540.7124  (not concave)
 Iteration 1:  log likelihood = -1515.7726
 Iteration 2:  log likelihood = -1473.0162
 Iteration 3:  log likelihood = -1472.9223
 Iteration 4:  log likelihood = -1472.6254
 Iteration 5:  log likelihood = -1472.607
 Iteration 6:  log likelihood = -1472.6069
```

```
Time-varying decay inefficiency model
Number of obs = 948
Group variable: id  Number of groups = 91
Time variable: t  Obs per group: min = 6
                   avg = 10.4
                   max = 14
Wald chi2(2) = 661.76
Log likelihood = -1472.6069
Prob > chi2 = 0.0000
( 1) [eta]_cons = 0
```

| Coefficient | Std. Err. | z   | P>|z| | [95% Conf. Interval] |
|-------------|-----------|-----|------|---------------------|
| lnwidgets   | .2904551  | .0164219 | 17.69 | 0.000 | .2582688 - .3226414 |
| lnmachines  | .2943332  | .0154352 | 19.07 | 0.000 | .2640807 - .3245857 |
| lnworkers   | 3.030963  | .1440995 | 21.03 | 0.000 | 2.748534 - 3.313393 |
| _cons       |           |         |      |        |                     |
| /mu         | 1.125507  | .6480444 | 1.74  | 0.082 | -.144639 - 2.39565 |
| /eta        | 0 (omitted)|         |       |        |                     |
| /lnsigma2   | 1.422039  | .2673128 | 5.32  | 0.000 | .8981155 - 1.945962 |
| /ilgtgamma  | 1.138764  | .3563076 | 3.20  | 0.001 | .4404135 - 1.837114 |
| sigma2      | 4.145565  | 1.108162 |       |        | 2.454972 - 7.000366 |
| gamma       | .7574526  | .0654602 |       |        | .6083575 - .862607 |
| sigma_u2    | 3.140068  | 1.107459 |       |        | .9694878 - 5.310649 |
| sigma_v2    | 1.005496  | .0484143 |       |        | .9106057 - 1.100386 |
Stored results

_xtfrontier_ stores the following in _e_():

Scalars

- _e(N)_ number of observations
- _e(N_g)_ number of groups
- _e(k)_ number of parameters
- _e(k_eq)_ number of equations in _e(b)_
- _e(k_eq_model)_ number of equations in overall model test
- _e(k_dv)_ number of dependent variables
- _e(df_m)_ model degrees of freedom
- _e(ll)_ log likelihood
- _e(g_min)_ minimum number of observations per group
- _e(g_avg)_ average number of observations per group
- _e(g_max)_ maximum number of observations per group
- _e(sigma2)_ sigma2
- _e(gamma)_ gamma
- _e(Tcon)_ 1 if panels balanced; 0 otherwise
- _e(sigma_u)_ standard deviation of technical inefficiency
- _e(sigma_v)_ standard deviation of random error
- _e(chi2)_ chi^2
- _e(p)_ model significance
- _e(rank)_ rank of _e(V)_
- _e(ic)_ number of iterations
- _e(rc)_ return code
- _e(converged)_ 1 if converged, 0 otherwise

Macros

- _e(cmd)_ _xtfrontier_
- _e(cmdline)_ command as typed
- _e(depvar)_ name of dependent variable
- _e(ivar)_ variable denoting groups
- _e(tvar)_ variable denoting time within groups
- _e(function)_ production or cost
- _e(model)_ `ti`, after time-invariant model; `tvd`, after time-varying decay model
- _e(wtype)_ weight type
- _e(wexp)_ weight expression
- _e(title)_ title in estimation output
- _e(chi2type)_ Wald; type of model chi^2 test
- _e(vce)_ vcetype specified in _vce()_ 
- _e(vcetype)_ title used to label Std. Err.
- _e(opt)_ type of optimization
- _e(which)_ max or min; whether optimizer is to perform maximization or minimization
- _e(ml_method)_ type of ml method
- _e(user)_ name of likelihood-evaluator program
- _e(technique)_ maximization technique
- _e(properties)_ b V
- _e(predict)_ program used to implement predict
- _e(asbalanced)_ factor variables fvset as asbalanced
- _e(asobserved)_ factor variables fvset as asobserved

Matrices

- _e(b)_ coefficient vector
- _e(Cns)_ constraints matrix
- _e(llog)_ iteration log (up to 20 iterations)
- _e(V)_ variance–covariance matrix of the estimators

Functions

- _e(sample)_ marks estimation sample
xtfrontier fits stochastic frontier models for panel data that can be expressed as

\[ y_{it} = \beta_0 + \sum_{j=1}^{k} \beta_j x_{jit} + v_{it} - su_{it} \]

where \( y_{it} \) is the natural logarithm of output, the \( x_{jit} \) are the natural logarithm of the input quantities for the production efficiency problem, \( y_{it} \) is the natural logarithm of costs, the \( x_{it} \) are the natural logarithm of input prices for the cost efficiency problem, and

\[ s = \begin{cases} 1, & \text{for production functions} \\ -1, & \text{for cost functions} \end{cases} \]

For the time-varying decay model, the log-likelihood function is derived as

\[
\ln L = -\frac{1}{2} \left\{ \sum_{i=1}^{N} T_i \right\} \left\{ \ln (2\pi) + \ln (\sigma^2_S) \right\} - \frac{1}{2} \sum_{i=1}^{N} (T_i - 1) \ln(1 - \gamma) \\
- \frac{1}{2} \sum_{i=1}^{N} \ln \left\{ 1 + \left( \sum_{t=1}^{T_i} \eta^2_{it} - 1 \right) \gamma \right\} - N \ln \left\{ 1 - \Phi (-\tilde{z}) \right\} - \frac{1}{2} N\tilde{z}^2 \\
+ \sum_{i=1}^{N} \ln \left\{ 1 - \Phi (-z_i^*) \right\} + \frac{1}{2} \sum_{i=1}^{N} z_i^{*2} - \frac{1}{2} \sum_{i=1}^{N} \sum_{t=1}^{T_i} \frac{\epsilon^2_{it}}{(1 - \gamma) \sigma^2_S} 
\]

where \( \sigma_S = (\sigma_u^2 + \sigma_v^2)^{1/2} \), \( \gamma = \sigma_u^2 / \sigma^2_S \), \( \epsilon_{it} = y_{it} - x_{it} \beta \), \( \eta_{it} = \exp\{-\eta(t - T_i)\} \), \( \tilde{z} = \mu / (\gamma \sigma^2_S)^{1/2} \), \( \Phi() \) is the cumulative distribution function of the standard normal distribution, and

\[
z_i^* = \frac{\mu (1 - \gamma) - s\gamma \sum_{t=1}^{T_i} \eta_{it} \epsilon_{it}}{\gamma (1 - \gamma) \sigma^2_S \left\{ 1 + \left( \sum_{t=1}^{T_i} \eta^2_{it} - 1 \right) \gamma \right\}^{1/2}} 
\]

Maximizing the above log likelihood estimates the coefficients \( \eta, \mu, \sigma_v, \) and \( \sigma_u \).

References


Also see

[XT] `xtfrontier postestimation` — Postestimation tools for `xtfrontier`

[XT] `xtset` — Declare data to be panel data

[R] `frontier` — Stochastic frontier models

[U] 20 Estimation and postestimation commands


**xtfrontier postestimation — Postestimation tools for xtfrontier**

### Description

The following postestimation commands are available after *xtfrontier*:

<table>
<thead>
<tr>
<th>Command</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>contrast</td>
<td>contrasts and ANOVA-style joint tests of estimates</td>
</tr>
<tr>
<td>estat ic</td>
<td>Akaike’s and Schwarz’s Bayesian information criteria (AIC and BIC)</td>
</tr>
<tr>
<td>estat summarize</td>
<td>summary statistics for the estimation sample</td>
</tr>
<tr>
<td>estat vce</td>
<td>variance–covariance matrix of the estimators (VCE)</td>
</tr>
<tr>
<td>estimates</td>
<td>cataloging estimation results</td>
</tr>
<tr>
<td>lincom</td>
<td>point estimates, standard errors, testing, and inference for linear combinations of coefficients</td>
</tr>
<tr>
<td>lrtest</td>
<td>likelihood-ratio test</td>
</tr>
<tr>
<td>margins</td>
<td>marginal means, predictive margins, marginal effects, and average marginal effects</td>
</tr>
<tr>
<td>marginsplot</td>
<td>graph the results from margins (profile plots, interaction plots, etc.)</td>
</tr>
<tr>
<td>nlcom</td>
<td>point estimates, standard errors, testing, and inference for nonlinear combinations of coefficients</td>
</tr>
<tr>
<td>predict</td>
<td>predictions, residuals, influence statistics, and other diagnostic measures</td>
</tr>
<tr>
<td>predictnl</td>
<td>point estimates, standard errors, testing, and inference for generalized predictions</td>
</tr>
<tr>
<td>pwcompare</td>
<td>pairwise comparisons of estimates</td>
</tr>
<tr>
<td>test</td>
<td>Wald tests of simple and composite linear hypotheses</td>
</tr>
<tr>
<td>testnl</td>
<td>Wald tests of nonlinear hypotheses</td>
</tr>
</tbody>
</table>

### Syntax for predict

```
predict [type] newvar [if] [in] [ , statistic ]
```

<table>
<thead>
<tr>
<th>statistic</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Main</td>
<td></td>
</tr>
<tr>
<td>xb</td>
<td>linear prediction; the default</td>
</tr>
<tr>
<td>stdp</td>
<td>standard error of the linear prediction</td>
</tr>
<tr>
<td>u</td>
<td>minus the natural log of the technical efficiency via $E(u_{it} \mid \epsilon_{it})$</td>
</tr>
<tr>
<td>m</td>
<td>minus the natural log of the technical efficiency via $M(u_{it} \mid \epsilon_{it})$</td>
</tr>
<tr>
<td>te</td>
<td>the technical efficiency via $E{\exp(-su_{it}) \mid \epsilon_{it}}$</td>
</tr>
</tbody>
</table>
where

\[
S = \begin{cases} 
1, & \text{for production functions} \\
-1, & \text{for cost functions}
\end{cases}
\]

### Menu for predict

Statistics > Postestimation > Predictions, residuals, etc.

### Options for predict

- **Main**
  - `xb`, the default, calculates the linear prediction.
  - `stdp` calculates the standard error of the linear prediction.
  - `u` produces estimates of minus the natural log of the technical efficiency via \(E(u_{it} | \epsilon_{it})\).
  - `m` produces estimates of minus the natural log of the technical efficiency via the mode, \(M(u_{it} | \epsilon_{it})\).
  - `te` produces estimates of the technical efficiency via \(E\{\exp(-su_{it}) | \epsilon_{it}\}\).

### Remarks and examples

#### Example 1

A production function exhibits constant returns to scale if doubling the amount of each input results in a doubling in the quantity produced. When the production function is linear in logs, constant returns to scale implies that the sum of the coefficients on the inputs is one. In example 2 of [XT] `xtfrontier`, we fit a time-varying decay model. Here we test whether the estimated production function exhibits constant returns:

```
. use http://www.stata-press.com/data/r13/xtfrontier1
. xtfrontier lnwidgets lnmachines lnworkers, tvd
(output omitted)
. test lnmachines + lnworkers = 1
```

```
( 1) [lnwidgets]lnmachines + [lnwidgets]lnworkers = 1
     chi2( 1) = 331.55
     Prob > chi2 = 0.0000
```

The test statistic is highly significant, so we reject the null hypothesis and conclude that this production function does not exhibit constant returns to scale.

The previous Wald \(\chi^2\) test indicated that the sum of the coefficients does not equal one. An alternative is to use `lincom` to compute the sum explicitly:

```
. lincom lnmachines + lnworkers
```

```
( 1) [lnwidgets]lnmachines + [lnwidgets]lnworkers = 0
    Coef.  Std. Err.    z  P>|z|     [95% Conf. Interval]
lnwidgets |     0.5849967    0.0227918  25.67    0.000    0.5403256    0.6296677
            |                  |          |          |                  |                  |
```

The test statistic is highly significant, so we reject the null hypothesis and conclude that this production function does not exhibit constant returns to scale.
The sum of the coefficients is significantly less than one, so this production function exhibits *decreasing returns to scale*. If we doubled the number of machines and workers, we would obtain less than twice as much output.

### Methods and formulas

Continuing from the *Methods and formulas* section of [XT] `xtfrontier`, estimates for $u_{it}$ can be obtained from the mean or the mode of the conditional distribution $f(u|\epsilon)$.

\[
E(u_{it} \mid \epsilon_{it}) = \tilde{\mu}_i + \tilde{\sigma}_i \left\{ \frac{\phi(-\tilde{\mu}_i/\tilde{\sigma}_i)}{1 - \Phi(-\tilde{\mu}_i/\tilde{\sigma}_i)} \right\}
\]

\[
M(u_{it} \mid \epsilon_{it}) = \begin{cases} -\tilde{\mu}_i, & \text{if } \tilde{\mu}_i \geq 0 \\ 0, & \text{otherwise} \end{cases}
\]

where

\[
\tilde{\mu}_i = \frac{\mu \sigma_v^2 - s \sum_{t=1}^{T_i} \eta_{it} \epsilon_{it} \sigma_u^2}{\sigma_v^2 + \sum_{t=1}^{T_i} \eta_{it}^2 \sigma_u^2}
\]

\[
\tilde{\sigma}_i^2 = \frac{\sigma_v^2 \sigma_u^2}{\sigma_v^2 + \sum_{t=1}^{T_i} \eta_{it}^2 \sigma_u^2}
\]

These estimates can be obtained from `predict newvar, u` and `predict newvar, m`, respectively, and are calculated by plugging in the estimated parameters.

`predict newvar, te` produces estimates of the technical-efficiency term. These estimates are obtained from

\[
E\{\exp(-su_{it}) \mid \epsilon_{it}\} = \left[ \frac{1 - \Phi\{s\eta_{it} \tilde{\sigma}_i - (\tilde{\mu}_i/\tilde{\sigma}_i)\}}{1 - \Phi(-\tilde{\mu}_i/\tilde{\sigma}_i)} \right] \exp \left( -s\eta_{it} \tilde{\mu}_i + \frac{1}{2} \eta_{it}^2 \tilde{\sigma}_i^2 \right)
\]

Replacing $\eta_{it} = 1$ and $\eta = 0$ in these formulas produces the formulas for the time-invariant models.

### Also see

[XT] `xtfrontier` — Stochastic frontier models for panel data

[U] 20 Estimation and postestimation commands
**xtgee — Fit population-averaged panel-data models by using GEE**

### Syntax

```bash
xtgee depvar [ indepvars ] [ if ] [ in ] [ weight ] [, options ]
```

### options

<table>
<thead>
<tr>
<th>Description</th>
<th>Syntax</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Model</strong></td>
<td></td>
</tr>
<tr>
<td><code>family(family)</code></td>
<td>distribution of <code>depvar</code></td>
</tr>
<tr>
<td><code>link(link)</code></td>
<td>link function</td>
</tr>
<tr>
<td><strong>Model 2</strong></td>
<td></td>
</tr>
<tr>
<td><code>exposure(varname)</code></td>
<td>include ln(<code>varname</code>) in model with coefficient constrained to 1</td>
</tr>
<tr>
<td><code>offset(varname)</code></td>
<td>include <code>varname</code> in model with coefficient constrained to 1</td>
</tr>
<tr>
<td><code>noconstant</code></td>
<td>suppress constant term</td>
</tr>
<tr>
<td><code>asis</code></td>
<td>retain perfect predictor variables</td>
</tr>
<tr>
<td><code>force</code></td>
<td>estimate even if observations unequally spaced in time</td>
</tr>
<tr>
<td><strong>Correlation</strong></td>
<td></td>
</tr>
<tr>
<td><code>corr(correlation)</code></td>
<td>within-group correlation structure</td>
</tr>
<tr>
<td><strong>SE/Robust</strong></td>
<td></td>
</tr>
<tr>
<td><code>vce(vcetype)</code></td>
<td><code>vcetype</code> may be <code>conventional</code>, <code>robust</code>, <code>bootstrap</code>, or <code>jackknife</code></td>
</tr>
<tr>
<td><code>rgf</code></td>
<td>multiply the robust variance estimate by <code>(N - 1)/(N - P)</code></td>
</tr>
<tr>
<td><code>nmp</code></td>
<td>use divisor <code>N - P</code> instead of the default <code>N</code></td>
</tr>
<tr>
<td><code>scale(parm)</code></td>
<td>overrides the default scale parameter; <code>parm</code> may be <code>x2</code>, <code>dev</code>, <code>phi</code>, or <code>#</code></td>
</tr>
<tr>
<td><strong>Reporting</strong></td>
<td></td>
</tr>
<tr>
<td><code>level(#)</code></td>
<td>set confidence level; default is <code>level(95)</code></td>
</tr>
<tr>
<td><code>eform</code></td>
<td>report exponentiated coefficients</td>
</tr>
<tr>
<td><code>display_options</code></td>
<td>control column formats, row spacing, line width, display of omitted</td>
</tr>
<tr>
<td></td>
<td>variables and base and empty cells, and factor-variable labeling</td>
</tr>
<tr>
<td><strong>Optimization</strong></td>
<td></td>
</tr>
<tr>
<td><code>optimize_options</code></td>
<td>control the optimization process; seldom used</td>
</tr>
<tr>
<td><code>nodisplay</code></td>
<td>suppress display of header and coefficients</td>
</tr>
<tr>
<td><code>coeflegend</code></td>
<td>display legend instead of statistics</td>
</tr>
</tbody>
</table>
A panel variable must be specified. Correlation structures other than exchangeable and independent require that a time variable also be specified. Use xtset; see [XT] xtset.

`indevars` may contain factor variables; see [U] 11.4.3 Factor variables. `depvar` and `indepvars` may contain time-series operators; see [U] 11.4 varlists.

by, mfp, mi estimate, and statsby are allowed; see [U] 11.1.10 Prefix commands. vce(bootstrap) and vce(jackknife) are not allowed with the mi estimate prefix; see [MI] mi estimate.
iweights, fweights, and pweights are allowed; see [U] 11.1.6 weight. Weights must be constant within panel. nodisplay and coeflegend do not appear in the dialog box.

See [U] 20 Estimation and postestimation commands for more capabilities of estimation commands.

<table>
<thead>
<tr>
<th>family</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>gaussian</td>
<td>Gaussian (normal); <code>family(normal)</code> is a synonym</td>
</tr>
<tr>
<td>igaussian</td>
<td>inverse Gaussian</td>
</tr>
<tr>
<td>binomial[#]</td>
<td>Bernoulli/binomial</td>
</tr>
<tr>
<td>poisson</td>
<td>Poisson</td>
</tr>
<tr>
<td>nbinomial[#]</td>
<td>negative binomial</td>
</tr>
<tr>
<td>gamma</td>
<td>gamma</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>link</th>
<th>Link function/definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>identity</td>
<td>identity; ( y = y )</td>
</tr>
<tr>
<td>log</td>
<td>log; ( \ln(y) )</td>
</tr>
<tr>
<td>logit</td>
<td>logit; ( \ln\left{\frac{y}{1-y}\right} ), natural log of the odds</td>
</tr>
<tr>
<td>probit</td>
<td>probit; ( \Phi^{-1}(y) ), where ( \Phi(\cdot) ) is the normal cumulative distribution</td>
</tr>
<tr>
<td>cloglog</td>
<td>cloglog; ( \ln{-\ln(1-y)} )</td>
</tr>
<tr>
<td>power[#]</td>
<td>power; ( y^k ) with ( k = # ); ( k = 1 ) if not specified</td>
</tr>
<tr>
<td>opower[#]</td>
<td>odds power; ( \left{\frac{y}{1-y}\right}^k - 1 )/( k ) with ( k = # ); ( k = 1 ) if not specified</td>
</tr>
<tr>
<td>nbinclog</td>
<td>negative binomial; ( \ln\left{\frac{y}{y+\alpha}\right} )</td>
</tr>
<tr>
<td>reciprocal</td>
<td>reciprocal; ( 1/y )</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>correlation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>exchangeable</td>
<td>exchangeable</td>
</tr>
<tr>
<td>independent</td>
<td>independent</td>
</tr>
<tr>
<td>unstructured</td>
<td>unstructured</td>
</tr>
<tr>
<td>fixed matname</td>
<td>user-specified</td>
</tr>
<tr>
<td>ar #</td>
<td>autoregressive of order #</td>
</tr>
<tr>
<td>stationary #</td>
<td>stationary of order #</td>
</tr>
<tr>
<td>nonstationary #</td>
<td>nonstationary of order #</td>
</tr>
</tbody>
</table>

**Menu**

Statistics > Longitudinal/panel data > Generalized estimating equations (GEE) > Generalized estimating equations (GEE)
Description

`xtgee` fits population-averaged panel-data models. In particular, `xtgee` fits generalized linear models and allows you to specify the within-group correlation structure for the panels.

See [R] logistic and [R] regress for lists of related estimation commands.

Options

```
Model

family(family) specifies the distribution of `depvar`; `family(gaussian)` is the default.
link(link) specifies the link function; the default is the canonical link for the `family()` specified (except for `family(nbinomial)`).

Model 2

exposure(varname) and offset(varname) are different ways of specifying the same thing. 
exposure() specifies a variable that reflects the amount of exposure over which the `depvar` events were observed for each observation; `ln(varname)` with coefficient constrained to be 1 is entered into the regression equation. offset() specifies a variable that is to be entered directly into the log-link function with its coefficient constrained to be 1; thus, exposure is assumed to be `e^{varname}`. If you were fitting a Poisson regression model, `family(poisson) link(log)`, for instance, you would account for exposure time by specifying offset() containing the log of exposure time.

noconstant specifies that the linear predictor has no intercept term, thus forcing it through the origin on the scale defined by the link function.
asis forces retention of perfect predictor variables and their associated, perfectly predicted observations and may produce instabilities in maximization; see [R] probit. This option is only allowed with option family(binomial) with a denominator of 1.
force specifies that estimation be forced even though the time variable is not equally spaced. This is relevant only for correlation structures that require knowledge of the time variable. These correlation structures require that observations be equally spaced so that calculations based on lags correspond to a constant time change. If you specify a time variable indicating that observations are not equally spaced, the (time dependent) model will not be fit. If you also specify force, the model will be fit, and it will be assumed that the lags based on the data ordered by the time variable are appropriate.

Correlation

corr(correlation) specifies the within-group correlation structure; the default corresponds to the equal-correlation model, `corr(exchangeable)`.

When you specify a correlation structure that requires a lag, you indicate the lag after the structure’s name with or without a blank; for example, `corr(ar 1)` or `corr(ar1)`.

If you specify the fixed correlation structure, you specify the name of the matrix containing the assumed correlations following the word fixed, for example, `corr(fixed myr)`.```
vce(vcetype) specifies the type of standard error reported, which includes types that are derived from asymptotic theory (conventional), that are robust to some kinds of misspecification (robust), and that use bootstrap or jackknife methods (bootstrap, jackknife); see [XT] vce_options.

vce(conventional), the default, uses the conventionally derived variance estimator for generalized least-squares regression.

vce(robust) specifies that the Huber/White/sandwich estimator of variance is to be used in place of the default conventional variance estimator (see Methods and formulas below). Use of this option causes xtgee to produce valid standard errors even if the correlations within group are not as hypothesized by the specified correlation structure. Under a noncanonical link, it does, however, require that the model correctly specifies the mean. The resulting standard errors are thus labeled “semirobust” instead of “robust” in this case. Although there is no vce(cluster clustvar) option, results are as if this option were included and you specified clustering on the panel variable.

rgf specifies that the robust variance estimate is multiplied by \((N - 1)/(N - P)\), where \(N\) is the total number of observations and \(P\) is the number of coefficients estimated. This option can be used only with family(gaussian) when vce(robust) is either specified or implied by the use of pweights. Using this option implies that the robust variance estimate is not invariant to the scale of any weights used.

scale(x2|dev|phi|#); see [XT] vce_options.

Reporting

level(#) ; see [R] estimation options.

eform displays the exponentiated coefficients and corresponding standard errors and confidence intervals as described in [R] maximize. For family(binomial) link(logit) (that is, logistic regression), exponentiation results in odds ratios; for family(poisson) link(log) (that is, Poisson regression), exponentiated coefficients are incidence-rate ratios.

display_options: noomitted, vsquish, noemptycells, baselevels, allbaselevels, nofvlabel, fvwrap(#) , fvwrapon(style), cformat(%fmt), pformat(%fmt), sformat(%fmt), and nolstretch ; see [R] estimation options.

Optimization

optimize_options control the iterative optimization process. These options are seldom used.

iterate(#) specifies the maximum number of iterations. When the number of iterations equals #, the optimization stops and presents the current results, even if convergence has not been reached. The default is iterate(100).

tolerance(#) specifies the tolerance for the coefficient vector. When the relative change in the coefficient vector from one iteration to the next is less than or equal to #, the optimization process is stopped. tolerance(1e-6) is the default.

nolog suppresses display of the iteration log.

trace specifies that the current estimates be printed at each iteration.

The following options are available with xtgee but are not shown in the dialog box: nodisplay is for programmers. It suppresses display of the header and coefficients. coeflegend; see [R] estimation options.
Remarks and examples

For a thorough introduction to GEE in the estimation of GLM, see Hardin and Hilbe (2013). More information on linear models is presented in Nelder and Wedderburn (1972). Finally, there have been several illuminating articles on various applications of GEE in Zeger, Liang, and Albert (1988); Zeger and Liang (1986), and Liang (1987). Pendergast et al. (1996) surveys the current methods for analyzing clustered data in regard to binary response data. Our implementation follows that of Liang and Zeger (1986).

_xtgee_ fits generalized linear models of \( y_{it} \) with covariates \( x_{it} \)

\[
\text{g} \left\{ \mathbb{E}(y_{it}) \right\} = x_{it}\beta, \quad y \sim F \text{ with parameters } \theta_{it}
\]

for \( i = 1, \ldots, m \) and \( t = 1, \ldots, n_i \), where there are \( n_i \) observations for each group identifier \( i \). \( g() \) is called the link function, and \( F \) is the distributional family. Substituting various definitions for \( g() \) and \( F \) results in a wide array of models. For instance, if \( y_{it} \) is distributed Gaussian (normal) and \( g() \) is the identity function, we have

\[
\mathbb{E}(y_{it}) = x_{it}\beta, \quad y \sim N()
\]

yielding linear regression, random-effects regression, or other regression-related models, depending on what we assume for the correlation structure.

If \( g() \) is the logit function and \( y_{it} \) is distributed Bernoulli (binomial), we have

\[
\text{logit} \left\{ \mathbb{E}(y_{it}) \right\} = x_{it}\beta, \quad y \sim \text{Bernoulli}
\]

or logistic regression. If \( g() \) is the natural log function and \( y_{it} \) is distributed Poisson, we have

\[
\text{ln} \left\{ \mathbb{E}(y_{it}) \right\} = x_{it}\beta, \quad y \sim \text{Poisson}
\]

or Poisson regression, also known as the log-linear model. Other combinations are possible.

You specify the link function with the link() option, the distributional family with family(), and the assumed within-group correlation structure with corr().

The binomial distribution can be specified as case 1 family(binomial), case 2 family(binomial #), or case 3 family(binomial varname). In case 2, # is the value of the binomial denominator \( N \), the number of trials. Specifying family(binomial 1) is the same as specifying family(binomial); both mean that \( y \) has the Bernoulli distribution with values 0 and 1 only. In case 3, varname is the variable containing the binomial denominator, thus allowing the number of trials to vary across observations.

The negative binomial distribution must be specified as family(nbinomial #), where # denotes the value of the parameter \( \alpha \) in the negative binomial distribution. The results will be conditional on this value.

You do not have to specify both family() and link(); the default link() is the canonical link for the specified family() (excluding family(nbinomial)):

<table>
<thead>
<tr>
<th>Family</th>
<th>Default link</th>
</tr>
</thead>
<tbody>
<tr>
<td>family(binomial)</td>
<td>link(logit)</td>
</tr>
<tr>
<td>family(gamma)</td>
<td>link(reciprocal)</td>
</tr>
<tr>
<td>family(gaussian)</td>
<td>link(identity)</td>
</tr>
<tr>
<td>family(igaussian)</td>
<td>link(power -2)</td>
</tr>
<tr>
<td>family(nbinomial)</td>
<td>link(log)</td>
</tr>
<tr>
<td>family(poisson)</td>
<td>link(log)</td>
</tr>
</tbody>
</table>
The canonical link for the negative binomial family is obtained by specifying link(nbinomial). If you specify both family() and link(), not all combinations make sense. You may choose among the following combinations:

<table>
<thead>
<tr>
<th>Gaussian</th>
<th>Inverse</th>
<th>Binomial</th>
<th>Poisson</th>
<th>Negative</th>
<th>Gamma</th>
</tr>
</thead>
<tbody>
<tr>
<td>Identity</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Log</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Logit</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Probit</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. log-log</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Power</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Odds</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Power</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neg. binom.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reciprocal</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
</tbody>
</table>

You specify the assumed within-group correlation structure with the corr() option.

For example, call \( R \) the working correlation matrix for modeling the within-group correlation, a square \( \max\{n_i\} \times \max\{n_i\} \) matrix. corr() specifies the structure of \( R \). Let \( R_{t,s} \) denote the \( t, s \) element.

The independent structure is defined as

\[
R_{t,s} = \begin{cases} 
1 & \text{if } t = s \\
0 & \text{otherwise}
\end{cases}
\]

The corr(exchangeable) structure (corresponding to equal-correlation models) is defined as

\[
R_{t,s} = \begin{cases} 
1 & \text{if } t = s \\
\rho & \text{otherwise}
\end{cases}
\]

The corr(ar g) structure is defined as the usual correlation matrix for an AR(g) model. This is sometimes called multiplicative correlation. For example, an AR(1) model is given by

\[
R_{t,s} = \begin{cases} 
1 & \text{if } t = s \\
\rho^{|t-s|} & \text{otherwise}
\end{cases}
\]

The corr(stationary g) structure is a stationary(g) model. For example, a stationary(1) model is given by

\[
R_{t,s} = \begin{cases} 
1 & \text{if } t = s \\
\rho & \text{if } |t - s| = 1 \\
0 & \text{otherwise}
\end{cases}
\]

The corr(nonstationary g) structure is a nonstationary(g) model that imposes only the constraints that the elements of the working correlation matrix along the diagonal be 1 and the elements outside the \( g \)th band be zero,

\[
R_{t,s} = \begin{cases} 
1 & \text{if } t = s \\
\rho_{ts} & \text{if } 0 < |t - s| \leq g, \rho_{ts} = \rho_{st} \\
0 & \text{otherwise}
\end{cases}
\]
corr(unstructured) imposes only the constraint that the diagonal elements of the working correlation matrix be 1.

\[ R_{t,s} = \begin{cases} 
1 & \text{if } t = s \\
\rho_{ts} & \text{otherwise, } \rho_{ts} = \rho_{st} 
\end{cases} \]

The corr(fixed matname) specification is taken from the user-supplied matrix, such that

\[ R = \text{matname} \]

Here the correlations are not estimated from the data. The user-supplied matrix must be a valid correlation matrix with 1s on the diagonal.

Full formulas for all the correlation structures are provided in the Methods and formulas below.

**Technical note**

Some family(), link(), and corr() combinations result in models already fit by Stata:

<table>
<thead>
<tr>
<th>family()</th>
<th>link()</th>
<th>corr()</th>
<th>Other Stata estimation command</th>
</tr>
</thead>
<tbody>
<tr>
<td>gaussian</td>
<td>identity</td>
<td>independent</td>
<td>regress</td>
</tr>
<tr>
<td>gaussian</td>
<td>identity</td>
<td>exchangeable</td>
<td>xtreg, re</td>
</tr>
<tr>
<td>gaussian</td>
<td>identity</td>
<td>exchangeable</td>
<td>xtreg, pa</td>
</tr>
<tr>
<td>binomial</td>
<td>cloglog</td>
<td>independent</td>
<td>cloglog (see note 1)</td>
</tr>
<tr>
<td>binomial</td>
<td>cloglog</td>
<td>exchangeable</td>
<td>xtcloglog, pa</td>
</tr>
<tr>
<td>binomial</td>
<td>logit</td>
<td>independent</td>
<td>logit or logistic</td>
</tr>
<tr>
<td>binomial</td>
<td>logit</td>
<td>exchangeable</td>
<td>xtlogit, pa</td>
</tr>
<tr>
<td>binomial</td>
<td>probit</td>
<td>independent</td>
<td>probit (see note 2)</td>
</tr>
<tr>
<td>binomial</td>
<td>probit</td>
<td>exchangeable</td>
<td>xtprob, pa</td>
</tr>
<tr>
<td>nbinomial</td>
<td>log</td>
<td>independent</td>
<td>nbreg (see note 3)</td>
</tr>
<tr>
<td>poisson</td>
<td>log</td>
<td>independent</td>
<td>poisson</td>
</tr>
<tr>
<td>poisson</td>
<td>log</td>
<td>exchangeable</td>
<td>xtpoisson, pa</td>
</tr>
<tr>
<td>gamma</td>
<td>log</td>
<td>independent</td>
<td>streg, dist(exp) nohr (see note 4)</td>
</tr>
<tr>
<td>family</td>
<td>link</td>
<td>independent</td>
<td>glm, irls (see note 5)</td>
</tr>
</tbody>
</table>

Notes:

1. For cloglog estimation, xtgee with corr(independent) and cloglog (see [R] cloglog) will produce the same coefficients, but the standard errors will be only asymptotically equivalent because cloglog is not the canonical link for the binomial family.

2. For probit estimation, xtgee with corr(independent) and probit will produce the same coefficients, but the standard errors will be only asymptotically equivalent because probit is not the canonical link for the binomial family. If the binomial denominator is not 1, the equivalent maximum-likelihood command is bprobit; see [R] probit and [R] glogit.

3. Fitting a negative binomial model by using xtgee (or using glm) will yield results conditional on the specified value of \( \alpha \). The nbreg command, however, estimates that parameter and provides unconditional estimates; see [R] nbreg.

4. xtgee with corr(independent) can be used to fit exponential regressions, but this requires specifying scale(1). As with probit, the xtgee-reported standard errors will be only asymptotically equivalent to those produced by streg, dist(exp) nohr (see [ST] streg) because log is not the canonical link for the gamma family. xtgee cannot be used to fit exponential regressions on censored data.

Using the independent correlation structure, the xtgee command will fit the same model fit with the glm, irls command if the family–link combination is the same.
If the `xtgee` command is equivalent to another command, using `corr(independent)` and the `vce(robust)` option with `xtgee` corresponds to using the `vce(cluster clustvar)` option in the equivalent command, where `clustvar` corresponds to the panel variable.

`xtgee` is a generalization of the `glm, irls` command and gives the same output when the same family and link are specified together with an independent correlation structure. What makes `xtgee` useful is:

- the number of statistical models that it generalizes for use with panel data, many of which are not otherwise available in Stata;
- the richer correlation structure `xtgee` allows, even when models are available through other `xt` commands; and
- the availability of robust standard errors (see [U] 20.21 Obtaining robust variance estimates), even when the model and correlation structure are available through other `xt` commands.

In the following examples, we illustrate the relationships of `xtgee` with other Stata estimation commands. Remember that, although `xtgee` generalizes many other commands, the computational algorithm is different; therefore, the answers you obtain will not be identical. The dataset we are using is a subset of the `nlswork` data (see [XT] `xt`); we are looking at observations before 1980.

**Example 1**

We can use `xtgee` to perform ordinary least squares by `regress`:

```
. use http://www.stata-press.com/data/r13/nlswork2
(National Longitudinal Survey. Young Women 14-26 years of age in 1968)
. regress ln_w grade age c.age#c.age
```

```
Source | SS df MS
------ | ---- ----
Model | 597.54468 3 199.18156
Residual | 2265.74584 16081 .14089583
Total | 2863.29052 16084 .178021047

Number of obs = 16085
F( 3, 16081) = 1413.68
Prob > F = 0.0000
R-squared = 0.2087
Adj R-squared = 0.2085
Root MSE = .37536

| ln_wage | Coef. | Std. Err. | t | P>|t| | [95% Conf. Interval] |
|--------|-------|-----------|---|-------|----------------------|
| grade  | .0724483 | .0014229 | 50.91 | 0.000 | .0696592 - .0752374 |
| age    | .1064874 | .0083644 | 12.73 | 0.000 | .0909022 - .1228825 |
| c.age#c.age | -.0016931 | .0001655 | -10.23 | 0.000 | -.0020174 - -.0013688 |
| _cons | -.8681487 | .1024896 | -8.47 | 0.000 | -1.06904 - .6672577 |
```
. xtgee ln_w grade age c.age#c.age, corr(indep) nmp
Iteration 1: tolerance = 1.285e-12
GEE population-averaged model
  Number of obs = 16085
  Group variable: idcode  Number of groups = 3913
  Link: identity  Obs per group: min = 1
  Family: Gaussian  avg = 4.1
  Correlation: independent  max = 9
Scale parameter: .1408958  Wald chi2(3) = 4241.04
Pearson chi2(16081): 2265.75  Deviance = 2265.75
Dispersion (Pearson): .1408958  Dispersion = .1408958

|         | Coef.  | Std. Err. | z     | P>|z|  | [95% Conf. Interval] |
|---------|--------|-----------|-------|------|----------------------|
| grade   | .0724483 | .0014229 | 50.91 | 0.000 | .0696594 .0752372   |
| age     | .1064874 | .0083644 | 12.73 | 0.000 | .0900935 .1228812   |
| c.age#c.age | -.0016931 | .0001655 | -10.23 | 0.000 | -.0020174 -.0013688 |
| _cons   | -.8681487 | .1024896 | -8.47 | 0.000 | -1.069025 -.6672728 |

When nmp is specified, the coefficients and the standard errors produced by the estimators are the same. Moreover, the scale parameter estimate from the xtgee command equals the MSE calculation from regress; both are estimates of the variance of the residuals.

> Example 2

The identity link and Gaussian family produce regression-type models. With the independent correlation structure, we reproduce ordinary least squares. With the exchangeable correlation structure, we produce an equal-correlation linear regression estimator.

xtgee, fam(gauss) link(ident) corr(exch) is asymptotically equivalent to the weighted-GLS estimator provided by xtreg, re and to the full maximum-likelihood estimator provided by xtreg, mle. In balanced data, xtgee, fam(gauss) link(ident) corr(exch) and xtreg, mle produce the same results. With unbalanced data, the results are close but differ because the two estimators handle unbalanced data differently. For both balanced and unbalanced data, the results produced by xtgee, fam(gauss) link(ident) corr(exch) and xtreg, mle differ from those produced by xtreg, re. Below we demonstrate the use of the three estimators with unbalanced data. We begin with xtgee; show the maximum likelihood estimator xtreg, mle; show the GLS estimator xtreg, re; and finally show xtgee with the vce(robust) option.
. xtgee ln_w grade age c.age#c.age, nolog

GEE population-averaged model  Number of obs =  16085
Group variable:     idcode  Number of groups =  3913
Link: identity  Obs per group: min =  1
Family: Gaussian  avg =  4.1
Correlation: exchangeable  max =  9
Wald chi2(3) =  2918.26
Scale parameter: .1416586  Prob > chi2 =  0.0000

| ln_wage   | Coef.    | Std. Err. | z     | P>|z|  | [95% Conf. Interval] |
|-----------|----------|-----------|-------|------|----------------------|
| grade     | .0717731 | .00211    | 34.02 | 0.000 | .0676737 -.0759086  |
| age       | .1077645 | .006885   | 15.65 | 0.000 | .0942701 .1212589   |
| c.age#c.age | -.0016381 | .0001362 | -12.03 | 0.000 | -.001905 -.0013712  |
| _cons     | -.9480449 | .0869277 | -10.91 | 0.000 | -1.11842 -.7776698  |

. xtreg ln_w grade age c.age#c.age, mle

Fitting constant-only model:
Iteration 0:  log likelihood = -6035.2751
Iteration 1:  log likelihood = -5870.6718
Iteration 2:  log likelihood = -5858.9478
Iteration 3:  log likelihood = -5858.8244
Iteration 4:  log likelihood = -5858.8244

Fitting full model:
Iteration 0:  log likelihood = -4591.9241
Iteration 1:  log likelihood = -4562.4406
Iteration 2:  log likelihood = -4562.3526
Iteration 3:  log likelihood = -4562.3525

Random-effects ML regression  Number of obs =  16085
Group variable:     idcode  Number of groups =  3913
Random effects u_i ~ Gaussian  Obs per group: min =  1
                                avg =  4.1
                                max =  9
LR chi2(3) =  2592.94
Log likelihood = -4562.3525  Prob > chi2 =  0.0000

| ln_wage     | Coef.    | Std. Err. | z     | P>|z|  | [95% Conf. Interval] |
|-------------|----------|-----------|-------|------|----------------------|
| grade       | .0717747 | .002142   | 33.51 | 0.000 | .0675765 .075973    |
| age         | .1077899 | .0068266  | 15.79 | 0.000 | .0944101 .1211697   |
| c.age#c.age | -.0016364 | .000135   | -12.12 | 0.000 | -.0019011 -.0013718 |
| _cons       | -.9500833 | .086384   | -11.00 | 0.000 | -1.119393 -.7807737 |
| /sigma_u    | .2689639 | .0040854  |       |      | .2610748 .2770915   |
| /sigma_e    | .2669944 | .0017113  |       |      | .2636613 .2703696   |
| rho         | .5036748 | .0086449  |       |      | .4867329 .52061     |

Likelihood-ratio test of sigma_u=0: chibar2(01) =  4996.22  Prob>=chibar2 =  0.000
. xtgee ln_w grade age c.age#c.age, re

Random-effects GLS regression                   Number of obs =  16085
Group variable: idcode                          Number of groups =  3913
R-sq: within = 0.0983                           Obs per group: min =  1
           between = 0.2946                        avg =  4.1
           overall = 0.2076                         max =  9
Wald chi2(3) = 2875.02                         corr(u_i, X) = 0 (assumed)
Prob > chi2 = 0.0000                           Prob > chi2 = 0.0000

|            | Coef.  | Std. Err. | z     | P>|z|  | [95% Conf. Interval] |
|------------|--------|-----------|-------|------|----------------------|
| ln_wage    | grade  | .0717757  | .0021666 | 33.13 | 0.000 | .0675294 - .0760221 |
|           | age    | .1078042  | .0068125 | 15.82 | 0.000 | .0944519 - .1215666 |
|           | c.age#c.age | -.0016355 | .0001347 | -12.14 | 0.000 | -.0018996 - -.0013714 |
|           | _cons  | -.9512118  | .0863139 | -11.02 | 0.000 | -1.120384 - -.7820397 |
| sigma_u   |        | .27383747  |        |       |         |
| sigma_e   |        | .26624266  |        |       |         |
| rho       |        | .51405959  |        |       |         |

Invariance assumption (fraction of variance due to u_i)

. xtgee ln_w grade age c.age#c.age, vce(robust) nolog

GEE population-averaged model                   Number of obs =  16085
Group variable: idcode                          Number of groups =  3913
Link: identity                                   Obs per group: min =  1
Family: Gaussian                                 avg =  4.1
Correlation: exchangeable                       max =  9
Wald chi2(3) = 2031.28                          Prob > chi2 = 0.0000
(Std. Err. adjusted for clustering on idcode)

|            | Coef.  | Std. Err. | z     | P>|z|  | [95% Conf. Interval] |
|------------|--------|-----------|-------|------|----------------------|
| ln_wage    | grade  | .0717731  | .0023341 | 30.75 | 0.000 | .0671983 - .0763479 |
|           | age    | .1077645  | .0098097 | 10.99 | 0.000 | .0985379 - .1269911 |
|           | c.age#c.age | -.0016381 | .0001964 | -8.34 | 0.000 | -.002023 - -.0012532 |
|           | _cons  | -.9480449  | .1195009 | -7.93 | 0.000 | -1.182262 - -.7138274 |

In [R] regress, regress, vce(cluster clustvar) may produce inefficient coefficient estimates with valid standard errors for random-effects models. These standard errors are robust to model misspecification. The vce(robust) option of xtgee, on the other hand, requires that the model correctly specify the mean and the link function when the noncanonical link is used.
xtgee stores the following in e():

Scalars
- e(N) number of observations
- e(N_g) number of groups
- e(df_m) model degrees of freedom
- e(chi2) $\chi^2$
- e(p) significance
- e(df_pear) degrees of freedom for Pearson $\chi^2$
- e(chi2_dev) $\chi^2$ test of deviance
- e(chi2_dis) $\chi^2$ test of deviance dispersion
- e(deviance) deviance
- e(dispers) deviance dispersion
- e(phi) scale parameter
- e(g_min) smallest group size
- e(g_avg) average group size
- e(g_max) largest group size
- e(tol) target tolerance
- e(dif) achieved tolerance
- e(rank) rank of e(V)
- e(rc) return code

Macros
- e(cmd) xtgee
- e(cmdline) command as typed
- e(depvar) name of dependent variable
- e(ivar) variable denoting groups
- e(tvar) variable denoting time within groups
- e(model) pa
- e(family) distribution family
- e(link) link function
- e(corr) correlation structure
- e(scale) x2, dev, phi, or #; scale parameter
- e(wtype) weight type
- e(wexp) weight expression
- e(offset) linear offset variable
- e(chi2type) Wald; type of model $\chi^2$ test
- e(vce) vcetype specified in vce()
- e(vcetype) title used to label Std. Err.
- e(properties) b V
- e(estat_cmd) program used to implement estat
- e(predict) program used to implement predict
- e(marginsnotok) predictions disallowed by margins
- e(asbalanced) factor variables fvset as asbalanced
- e(asobserved) factor variables fvset as asobserved

Matrices
- e(b) coefficient vector
- e(R) estimated working correlation matrix
- e(V) variance–covariance matrix of the estimators
- e(V_modelbased) model-based variance

Functions
- e(sample) marks estimation sample
Methods and formulas

Methods and formulas are presented under the following headings:

Introduction
Calculating GEE for GLM
Correlation structures
Nonstationary and unstructured

Introduction

`xtgee` fits generalized linear models for panel data with the GEE approach described in Liang and Zeger (1986). A related method, referred to as GEE2, is described in Zhao and Prentice (1990) and Prentice and Zhao (1991). The GEE2 method attempts to gain efficiency in the estimation of $\beta$ by specifying a parametric model for $\alpha$ and then assumes that the models for both the mean and dependency parameters are correct. Thus there is a tradeoff in robustness for efficiency. The preliminary work of Liang, Zeger, and Qaqish (1992), however, indicates that there is little efficiency gained with this alternative approach.

In the GLM approach (see McCullagh and Nelder [1989]), we assume that

$$h(\mu_{i,j}) = x_{i,j}^T \beta$$
$$\text{Var}(y_{i,j}) = g(\mu_{i,j}) \phi$$
$$\mu_i = E(y_i) = \{h^{-1}(x_{i,1}^T \beta), \ldots, h^{-1}(x_{i,n_i}^T \beta)\}^T$$
$$A_i = \text{diag}\{g(\mu_{i,1}), \ldots, g(\mu_{i,n_i})\}$$
$$\text{Cov}(y_i) = \phi A_i \text{ for independent observations.}$$

In the absence of a convenient likelihood function with which to work, we can rely on a multivariate analog of the quasiscore function introduced by Wedderburn (1974):

$$S_\beta(\beta, \alpha) = \sum_{i=1}^{m} \left( \frac{\partial \mu_i}{\partial \beta} \right)^T \text{Var}(y_i)^{-1}(y_i - \mu_i) = 0$$

We can solve for correlation parameters $\alpha$ by simultaneously solving

$$S_\alpha(\beta, \alpha) = \sum_{i=1}^{m} \left( \frac{\partial \eta_i}{\partial \alpha} \right)^T H_i^{-1}(W_i - \eta_i) = 0$$

In the GEE approach to GLM, we let $R_i(\alpha)$ be a “working” correlation matrix depending on the parameters in $\alpha$ (see the Correlation structures section for the number of parameters), and we estimate $\beta$ by solving the GEE,

$$U(\beta) = \sum_{i=1}^{m} \left( \frac{\partial \mu_i}{\partial \beta} \right)^T V_i^{-1}(\alpha)(y_i - \mu_i) = 0$$

where $V_i(\alpha) = A_i^{1/2} R_i(\alpha) A_i^{1/2}$
To solve this equation, we need only a crude approximation of the variance matrix, which we can obtain from a Taylor series expansion, where

$$\text{Cov}(y_i) = L_iZ_iD_iZ_i^T L_i + \phi A_i = \tilde{V}_i$$

$$L_i = \text{diag}\{\partial h^{-1}(u)/\partial u, u = x_{i,j}^T \beta, j = 1, \ldots, n_i\}$$

which allows that

$$\hat{D}_i \approx \left(Z_i^T Z_i\right)^{-1} Z_i \tilde{L}_i^{-1} \left\{ (y_i - \hat{\mu}_i)(y_i - \hat{\mu}_i)^T - \hat{\phi} A_i \right\} \tilde{L}_i^{-1} Z_i^T (Z_i^T Z_i)^{-1}$$

$$\hat{\phi} = \sum_{i=1}^m \sum_{j=1}^{n_i} \frac{(y_{i,j} - \hat{\mu}_{i,j})^2 - (\tilde{L}_{i,j})^2 Z_i^T \hat{D}_i Z_{i,j}}{g(\hat{\mu}_{i,j})}$$

### Calculating GEE for GLM

Using the notation from Liang and Zeger (1986), let $y_i = (y_{i,1}, \ldots, y_{i,n_i})^T$ be the $n_i \times 1$ vector of outcome values, and let $X_i = (x_{i,1}, \ldots, x_{i,n_i})^T$ be the $n_i \times p$ matrix of covariate values for the $i$th subject $i = 1, \ldots, m$. We assume that the marginal density for $y_{i,j}$ may be written in exponential family notation as

$$f(y_{i,j}) = \exp \left\{ \{y_{i,j} \theta_{i,j} - a(\theta_{i,j}) + b(y_{i,j})\} \phi \right\}$$

where $\theta_{i,j} = h(\eta_{i,j}), \eta_{i,j} = x_{i,j} \beta$. Under this formulation, the first two moments are given by

$$E(y_{i,j}) = a'(\theta_{i,j}), \quad \text{Var}(y_{i,j}) = a''(\theta_{i,j})/\phi$$

In what follows, we let $n_i = n$ without loss of generality. We define the quantities, assuming that we have an $n \times n$ working correlation matrix $R(\alpha)$,

$$\Delta_i = \text{diag}(d \theta_{i,j}/d \eta_{i,j}) \quad n \times n \text{ matrix}$$
$$A_i = \text{diag}\{a''(\theta_{i,j})\} \quad n \times n \text{ matrix}$$
$$S_i = y_i - a'(\theta_i) \quad n \times 1 \text{ matrix}$$
$$D_i = A_i \Delta_i X_i \quad n \times p \text{ matrix}$$
$$V_i = A_i^{1/2} R(\alpha) A_i^{1/2} \quad n \times n \text{ matrix}$$

such that the GEE becomes

$$\sum_{i=1}^m D_i^T V_i^{-1} S_i = 0$$

We then have that

$$\hat{\beta}_{j+1} = \hat{\beta}_j - \left\{ \sum_{i=1}^m D_i^T (\hat{\beta}_j) V_i^{-1} (\hat{\beta}_j) D_i (\hat{\beta}_j) \right\}^{-1} \left\{ \sum_{i=1}^m D_i^T (\hat{\beta}_j) V_i^{-1} (\hat{\beta}_j) S_i (\hat{\beta}_j) \right\}$$

where the term

$$\left\{ \sum_{i=1}^m D_i^T (\hat{\beta}_j) V_i^{-1} (\hat{\beta}_j) D_i (\hat{\beta}_j) \right\}^{-1}$$
is what we call the conventional variance estimate. It is used to calculate the standard errors if the \texttt{vce(robust)} option is not specified. This command supports the clustered version of the Huber/White/sandwich estimator of the variance with panels treated as clusters when \texttt{vce(robust)} is specified. See [P] \_robust, particularly \textit{Maximum likelihood estimators} and \textit{Methods and formulas}. Liang and Zeger (1986) also discuss the calculation of the robust variance estimator.

Define the following:

\[
D = (D_1^T, \ldots, D_m^T) \\
S = (S_1^T, \ldots, S_m^T)^T \\
\tilde{V} = nm \times nm \text{ block diagonal matrix with } \tilde{V}_i \\
Z = D\beta - S
\]

At a given iteration, the correlation parameters $\alpha$ and scale parameter $\phi$ can be estimated from the current Pearson residuals, defined by

\[
\hat{\tau}_{i,j} = \{y_{i,j} - a'(\hat{\theta}_{i,j})\}/\{a''(\hat{\theta}_{i,j})\}^{1/2}
\]

where $\hat{\theta}_{i,j}$ depends on the current value for $\hat{\beta}$. We can then estimate $\phi$ by

\[
\hat{\phi}^{-1} = \sum_{i=1}^{m} \sum_{j=1}^{n_i} \hat{\tau}_{i,j}^2 / (N - p)
\]

As this general derivation is complicated, let’s follow the derivation of the Gaussian family with the identity link (regression) to illustrate the generalization. After making appropriate substitutions, we will see a familiar updating equation. First, we rewrite the updating equation for $\beta$ as

\[
\hat{\beta}_{j+1} = \hat{\beta}_j - Z_1^{-1}Z_2
\]

and then derive $Z_1$ and $Z_2$.

\[
Z_1 = \sum_{i=1}^{m} D_i^T (\hat{\beta}_j) \tilde{V}_i^{-1}(\hat{\beta}_j) D_i (\hat{\beta}_j) = \sum_{i=1}^{m} X_i^T \Delta_i^T A_i^T \{A_i^{1/2} R(\alpha) A_i^{1/2}\}^{-1} A_i \Delta_i X_i
\]

\[
= \sum_{i=1}^{m} X_i^T \text{diag} \left( \frac{\partial \theta_{i,j}}{\partial (X\beta)} \right) \text{diag} \left( a''(\theta_{i,j}) \right) \left[ \text{diag} \left( a''(\theta_{i,j}) \right)^{1/2} R(\alpha) \text{diag} \left( a''(\theta_{i,j}) \right)^{1/2} \right]^{-1} \text{diag} \left( a''(\theta_{i,j}) \right) \text{diag} \left( \frac{\partial \theta_{i,j}}{\partial (X\beta)} \right) X_i
\]

\[
= \sum_{i=1}^{m} X_i^T \Pi(III)^{-1} \Pi X_i = \sum_{i=1}^{m} X_i^T X_i = X^T X
\]
\[ Z_2 = \sum_{i=1}^{m} D_i^T (\hat{\beta}_j) \tilde{V}_i^{-1} (\hat{\beta}_j) S_i (\hat{\beta}_j) = \sum_{i=1}^{m} X_i^T \Delta_i^T A_i^T \{ A_i^{1/2} R(\alpha) A_i^{1/2} \}^{-1} \left( y_i - X_i \hat{\beta}_j \right) \]

\[ = \sum_{i=1}^{m} X_i^T \text{diag} \left\{ \frac{\partial \theta_{i,j}}{\partial (X \beta)} \right\} \text{diag} \{ a''(\theta_{i,j}) \} \left[ \text{diag} \{ a''(\theta_{i,j}) \} \right]^{1/2} R(\alpha) \text{diag} \{ a''(\theta_{i,j}) \}^{1/2} \left( y_i - X_i \hat{\beta}_j \right) \]

\[ = \sum_{i=1}^{m} X_i^T \Pi(\Pi)^{-1} (y_i - X_i \hat{\beta}_j) = \sum_{i=1}^{m} X_i^T (y_i - X_i \hat{\beta}_j) = X^T \hat{\beta}_j \]

So, we may write the update formula as

\[ \hat{\beta}_{j+1} = \hat{\beta}_j - (X^TX)^{-1}X^T \hat{\beta}_j \]

which is the same formula for GLS in regression.

### Correlation structures

The working correlation matrix \( R \) is a function of \( \alpha \) and is more accurately written as \( R(\alpha) \). Depending on the assumed correlation structure, \( \alpha \) might be

- **Independent**: no parameters to estimate
- **Exchangeable**: \( \alpha \) is a scalar
- **Autoregressive**: \( \alpha \) is a vector
- **Stationary**: \( \alpha \) is a vector
- **Nonstationary**: \( \alpha \) is a matrix
- **Unstructured**: \( \alpha \) is a matrix

Also, throughout the estimation of a general unbalanced panel, it is more proper to discuss \( R_i \), which is the upper left \( n_i \times n_i \) submatrix of the ultimately stored matrix in \( e(R) \), \( \max\{n_i\} \times \max\{n_i\} \).

The only panels that enter into the estimation for a lag-dependent correlation structure are those with \( n_i > g \) (assuming a lag of \( g \)). \texttt{xtgee} drops panels with too few observations (and mentions when it does so).

### Independent

The working correlation matrix \( R \) is an identity matrix.

### Exchangeable

\[ \alpha = \frac{\sum_{j=1}^{n_i} \left( \sum_{k=1}^{n_i} \hat{r}_{i,j} \hat{r}_{i,k} - \sum_{j=1}^{n_i} \hat{r}_{i,j}^2 \right) \} \sum_{i=1}^{m} \left( \sum_{j=1}^{n_i} \hat{r}_{i,j}^2 \right)}{\sum_{i=1}^{m} \left\{ n_i (n_i - 1) \right\} \sum_{i=1}^{m} n_i} \]

and the working correlation matrix is given by

\[ R_{s,t} = \begin{cases} 1 & s = t \\ \alpha & \text{otherwise} \end{cases} \]
Autoregressive and stationary

These two structures require \( g \) parameters to be estimated so that \( \alpha \) is a vector of length \( g + 1 \) (the first element of \( \alpha \) is 1).

\[
\alpha = \sum_{i=1}^{m} \left( \frac{\sum_{j=1}^{n_i} \hat{r}_{i,j}^2}{n_i}, \frac{\sum_{j=1}^{n_i-1} \hat{r}_{i,j} \hat{r}_{i,j+1}}{n_i}, \ldots, \frac{\sum_{j=1}^{n_i-g} \hat{r}_{i,j} \hat{r}_{i,j+g}}{n_i} \right) / \left( \sum_{i=1}^{m} \frac{\sum_{j=1}^{n_i} \hat{r}_{i,j}^2}{n_i} \right)
\]

The working correlation matrix for the AR model is calculated as a function of Toeplitz matrices formed from the \( \alpha \) vector; see Newton (1988). The working correlation matrix for the stationary model is given by

\[
R_{s,t} = \begin{cases} 
\alpha_{1,|s-t|} & \text{if } |s - t| \leq g \\
0 & \text{otherwise}
\end{cases}
\]

Nonstationary and unstructured

These two correlation structures require a matrix of parameters. \( \alpha \) is estimated (where we replace \( \hat{r}_{i,j} = 0 \) whenever \( i > n_i \) or \( j > n_i \)) as

\[
\alpha = \sum_{i=1}^{m} \left( \frac{N_{1,1}^{-1} \hat{r}_{i,1}^2}{N_{1,1}}, \frac{N_{1,2}^{-1} \hat{r}_{i,1} \hat{r}_{i,2}}{N_{1,2}}, \ldots, \frac{N_{1,n}^{-1} \hat{r}_{i,1} \hat{r}_{i,n}}{N_{1,n}} \right) / \left( \sum_{i=1}^{m} \frac{\sum_{j=1}^{n_i} \hat{r}_{i,j}^2}{n_i} \right)
\]

where \( N_{p,q} = \sum_{i=1}^{m} I(i, p, q) \) and

\[
I(i, p, q) = \begin{cases} 
1 & \text{if panel } i \text{ has valid observations at times } p \text{ and } q \\
0 & \text{otherwise}
\end{cases}
\]

where \( N_{i,j} = \min(N_i, N_j) \), \( N_i = \text{number of panels observed at time } i \), and \( n = \max(n_1, n_2, \ldots, n_m) \).

The working correlation matrix for the nonstationary model is given by

\[
R_{s,t} = \begin{cases} 
1 & \text{if } s = t \\
\alpha_{s,t} & \text{if } 0 < |s - t| \leq g \\
0 & \text{otherwise}
\end{cases}
\]

The working correlation matrix for the unstructured model is given by

\[
R_{s,t} = \begin{cases} 
1 & \text{if } s = t \\
\alpha_{s,t} & \text{otherwise}
\end{cases}
\]

such that the unstructured model is equal to the nonstationary model at lag \( g = n - 1 \), where the panels are balanced with \( n_i = n \) for all \( i \).
References


Also see

- **xtgee postestimation** — Postestimation tools for xtgee
- **xtcloglog** — Random-effects and population-averaged cloglog models
- **xtlogit** — Fixed-effects, random-effects, and population-averaged logit models
- **xtnbreg** — Fixed-effects, random-effects, & population-averaged negative binomial models
- **xtpoisson** — Fixed-effects, random-effects, and population-averaged Poisson models
- **xtprobit** — Random-effects and population-averaged probit models
- **xtreg** — Fixed-, between-, and random-effects and population-averaged linear models
- **xtregar** — Fixed- and random-effects linear models with an AR(1) disturbance
- **xtset** — Declare data to be panel data
- **estimation** — Estimation commands for use with mi estimate
- **glm** — Generalized linear models
- **logistic** — Logistic regression, reporting odds ratios
- **regress** — Linear regression
- **20 Estimation and postestimation commands**
### Description

The following postestimation command is of special interest after `xtgee`:

<table>
<thead>
<tr>
<th>Command</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>estat wcorrelation</code></td>
<td>estimated matrix of the within-group correlations</td>
</tr>
</tbody>
</table>

The following standard postestimation commands are also available:

<table>
<thead>
<tr>
<th>Command</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>contrast</code></td>
<td>contrasts and ANOVA-style joint tests of estimates</td>
</tr>
<tr>
<td><code>estat summarize</code></td>
<td>summary statistics for the estimation sample</td>
</tr>
<tr>
<td><code>estat vce</code></td>
<td>variance–covariance matrix of the estimators (VCE)</td>
</tr>
<tr>
<td><code>estimates</code></td>
<td>cataloging estimation results</td>
</tr>
<tr>
<td><code>forecast</code>¹</td>
<td>dynamic forecasts and simulations</td>
</tr>
<tr>
<td><code>hausman</code></td>
<td>Hausman’s specification test</td>
</tr>
<tr>
<td><code>lincom</code></td>
<td>point estimates, standard errors, testing, and inference for linear combinations of coefficients</td>
</tr>
<tr>
<td><code>margins</code></td>
<td>marginal means, predictive margins, marginal effects, and average marginal effects</td>
</tr>
<tr>
<td><code>marginsplot</code></td>
<td>graph the results from margins (profile plots, interaction plots, etc.)</td>
</tr>
<tr>
<td><code>nlcom</code></td>
<td>point estimates, standard errors, testing, and inference for nonlinear combinations of coefficients</td>
</tr>
<tr>
<td><code>predict</code></td>
<td>predictions, residuals, influence statistics, and other diagnostic measures</td>
</tr>
<tr>
<td><code>predictnl</code></td>
<td>point estimates, standard errors, testing, and inference for generalized predictions</td>
</tr>
<tr>
<td><code>pwcompare</code></td>
<td>pairwise comparisons of estimates</td>
</tr>
<tr>
<td><code>test</code></td>
<td>Wald tests of simple and composite linear hypotheses</td>
</tr>
<tr>
<td><code>testnl</code></td>
<td>Wald tests of nonlinear hypotheses</td>
</tr>
</tbody>
</table>

¹ `forecast` is not appropriate with `mi` estimation results.
Special-interest postestimation commands

estat wcorrelation displays the estimated matrix of the within-group correlations.

Syntax for predict

```plaintext
predict [ type ] newvar [ if ] [ in ] [ , statistic nooffset ]
```

<table>
<thead>
<tr>
<th>statistic</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>mu</td>
<td>predicted value of depvar; considers the offset() or exposure(); the default</td>
</tr>
<tr>
<td>rate</td>
<td>predicted value of depvar</td>
</tr>
<tr>
<td>pr(n)</td>
<td>probability Pr(yj = n) for family(poisson) link(log)</td>
</tr>
<tr>
<td>pr(a,b)</td>
<td>probability Pr(a ≤ yj ≤ b) for family(poisson) link(log)</td>
</tr>
<tr>
<td>xb</td>
<td>linear prediction</td>
</tr>
<tr>
<td>stdp</td>
<td>standard error of the linear prediction</td>
</tr>
<tr>
<td>score</td>
<td>first derivative of the log likelihood with respect to xjβ</td>
</tr>
</tbody>
</table>

These statistics are available both in and out of sample; type predict ... if e(sample) ... if wanted only for the estimation sample.

Menu for predict

Statistics > Postestimation > Predictions, residuals, etc.

Options for predict

mu, the default, and rate calculate the predicted value of depvar. mu takes into account the offset() or exposure() together with the denominator if the family is binomial; rate ignores those adjustments. mu and rate are equivalent if you did not specify offset() or exposure() when you fit the xtgee model and you did not specify family(binomial #) or family(binomial varname), meaning the binomial family and a denominator not equal to one.

Thus mu and rate are the same for family(gaussian) link(identity).

mu and rate are not equivalent for family(binomial pop) link(logit). Then mu would predict the number of positive outcomes and rate would predict the probability of a positive outcome.

mu and rate are not equivalent for family(poisson) link(log) exposure(time). Then mu would predict the number of events given exposure time and rate would calculate the incidence rate—the number of events given an exposure time of 1.

pr(n) calculates the probability Pr(yj = n) for family(poisson) link(log), where n is a nonnegative integer that may be specified as a number or a variable.
pr(a, b) calculates the probability \( \text{Pr}(a \leq y_j \leq b) \) for family(poisson) link(log), where \( a \) and \( b \) are nonnegative integers that may be specified as numbers or variables;

- \( b \) missing (\( b \geq . \)) means \(+\infty\);
- \( \text{pr}(20,. \) ) calculates \( \text{Pr}(y_j \geq 20) \);
- \( \text{pr}(20,b \) ) calculates \( \text{Pr}(y_j \geq 20) \) in observations for which \( b \geq . \) and calculates \( \text{Pr}(20 \leq y_j \leq b) \) elsewhere.

\( \text{pr}(. , b \) ) produces a syntax error. A missing value in an observation of the variable \( a \) causes a missing value in that observation for \( \text{pr}(a,b) \).

\( \text{xb} \) calculates the linear prediction.

\( \text{stdp} \) calculates the standard error of the linear prediction.

\( \text{score} \) calculates the equation-level score, \( u_j = \frac{\partial \ln L_j(x_j\beta)}{\partial (x_j\beta)} \).

\( \text{nooffset} \) is relevant only if you specified \( \text{offset(varname)} \), \( \text{exposure(varname)} \), \( \text{family(binomial #)} \), or \( \text{family(binomial varname)} \) when you fit the model. It modifies the calculations made by \( \text{predict} \) so that they ignore the offset or exposure variable and the binomial denominator. Thus \( \text{predict ... , mu nooffset} \) produces the same results as \( \text{predict ... , rate} \).

### Syntax for estat wcorrelation

\[
\text{estat wcorrelation } [, \text{ compact format(\%fmt)}]\]

### Menu for estat

Statistics > Postestimation > Reports and statistics

### Options for estat wcorrelation

- \( \text{compact} \) specifies that only the parameters (alpha) of the estimated matrix of within-group correlations be displayed rather than the entire matrix.
- \( \text{format(\%fmt)} \) overrides the display format; see \([D]\ \text{format}\).

### Remarks and examples

#### Example 1

\( \text{xtgee} \) can estimate rich correlation structures. In \textbf{example 2} of \([XT]\ \text{xtgee}\), we fit the model

\[
\text{. use http://www.stata-press.com/data/r13/nlswork2} \\
\text{(National Longitudinal Survey. Young Women 14-26 years of age in 1968)} \\
\text{. xtgee ln_w grade age c.age#c.age} \\
\text{(output omitted)}
\]
After estimation, *estat wcorrelation* reports the working correlation matrix \( R \):

```plaintext
estat wcorrelation
Estimated within-idcode correlation matrix \( R \):

<table>
<thead>
<tr>
<th></th>
<th>c1</th>
<th>c2</th>
<th>c3</th>
<th>c4</th>
<th>c5</th>
<th>c6</th>
</tr>
</thead>
<tbody>
<tr>
<td>r1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>r2</td>
<td>.4851356</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>r3</td>
<td>.4851356</td>
<td>.4851356</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>r4</td>
<td>.4851356</td>
<td>.4851356</td>
<td>.4851356</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>r5</td>
<td>.4851356</td>
<td>.4851356</td>
<td>.4851356</td>
<td>.4851356</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>r6</td>
<td>.4851356</td>
<td>.4851356</td>
<td>.4851356</td>
<td>.4851356</td>
<td>.4851356</td>
<td>1</td>
</tr>
<tr>
<td>r7</td>
<td>.4851356</td>
<td>.4851356</td>
<td>.4851356</td>
<td>.4851356</td>
<td>.4851356</td>
<td>.4851356</td>
</tr>
<tr>
<td>r8</td>
<td>.4851356</td>
<td>.4851356</td>
<td>.4851356</td>
<td>.4851356</td>
<td>.4851356</td>
<td>.4851356</td>
</tr>
<tr>
<td>r9</td>
<td>.4851356</td>
<td>.4851356</td>
<td>.4851356</td>
<td>.4851356</td>
<td>.4851356</td>
<td>.4851356</td>
</tr>
</tbody>
</table>

The equal-correlation model corresponds to an exchangeable correlation structure, meaning that the correlation of observations within person is a constant. The working correlation estimated by *xtgee* is 0.4851. (*xtreg, re*, by comparison, reports 0.5141; see the *xtreg* command in example 2 of [*XT* *xtgee*].) We constrained the model to have this simple correlation structure. What if we relaxed the constraint? To go to the other extreme, let’s place no constraints on the matrix (other than its being symmetric). We do this by specifying *correlation(unstructured)*, although we can abbreviate the option.

```plaintext
xtgee ln_w grade age c.age#c.age, corr(unstr) nolog
```

```
GEE population-averaged model  Number of obs = 16085
Group and time vars: idcode year  Number of groups = 3913
Link: identity  Obs per group: min = 1
Family: Gaussian  avg = 4.1
Correlation: unstructured  max = 9
Wald chi2(3) = 2405.20
Scale parameter: .1418513  Prob > chi2 = 0.0000
```

| ln_wage       | Coef. | Std. Err. | z     | P>|z|  | [95% Conf. Interval] |
|---------------|-------|-----------|-------|------|---------------------|
| grade         | .0720684 | .002151 | 33.50 | 0.000 | .0678525 to .0762843 |
| age           | .1008095 | .0081471 | 12.37 | 0.000 | .0848416 to .1167775 |
| c.age#c.age   | -.0015104 | .0001617 | -.934 | 0.349 | -.0018272 to -.0011936 |
| _cons         | -.8645484 | .1009488 | -8.56 | 0.000 | -1.062404 to -.666923 |
```
. estat wcorrelation

Estimated within-idcode correlation matrix R:

<table>
<thead>
<tr>
<th></th>
<th>c1</th>
<th>c2</th>
<th>c3</th>
<th>c4</th>
<th>c5</th>
<th>c6</th>
</tr>
</thead>
<tbody>
<tr>
<td>r1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>r2</td>
<td>.4354838</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>r3</td>
<td>.4280248</td>
<td>.5597329</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>r4</td>
<td>.3772342</td>
<td>.5012129</td>
<td>.5475113</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>r5</td>
<td>.4031433</td>
<td>.5301403</td>
<td>.502668</td>
<td>.6216227</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>r6</td>
<td>.3663686</td>
<td>.4519138</td>
<td>.4783186</td>
<td>.5685009</td>
<td>.7306005</td>
<td>1</td>
</tr>
<tr>
<td>r7</td>
<td>.2819915</td>
<td>.3605743</td>
<td>.3918118</td>
<td>.4012104</td>
<td>.4642561</td>
<td>.50219</td>
</tr>
<tr>
<td>r8</td>
<td>.3162028</td>
<td>.3445668</td>
<td>.4285424</td>
<td>.4389241</td>
<td>.4696792</td>
<td>.522357</td>
</tr>
<tr>
<td>r9</td>
<td>.2148737</td>
<td>.3078491</td>
<td>.337292</td>
<td>.3584013</td>
<td>.4865802</td>
<td>.4613128</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>c7</th>
<th>c8</th>
<th>c9</th>
</tr>
</thead>
<tbody>
<tr>
<td>r7</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>r8</td>
<td>.6475654</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>r9</td>
<td>.5791417</td>
<td>.7386595</td>
<td>1</td>
</tr>
</tbody>
</table>

This correlation matrix looks different from the previously constrained one and shows, in particular, that the serial correlation of the residuals diminishes as the lag increases, although residuals separated by small lags are more correlated than, say, AR(1) would imply.

Example 2

In example 1 of [XT] xtprobit, we showed a random-effects model of unionization using the union data described in [XT] xt. We performed the estimation using xtprobit but said that we could have used xtgee as well. Here we fit a population-averaged (equal correlation) model for comparison:

. use http://www.stata-press.com/data/r13/union
(NLS Women 14–24 in 1968)
. xtgee union age grade i.not_smsa south##c.year, family(binomial) link(probit)

Iteration 1: tolerance = .12544249
Iteration 2: tolerance = .0034686
Iteration 3: tolerance = .00017448
Iteration 4: tolerance = 8.382e-06
Iteration 5: tolerance = 3.997e-07

GEE population-averaged model Number of obs = 26200
Group variable: idcode Number of groups = 4434
Link: probit Obs per group: min = 1
Family: binomial avg = 5.9
Correlation: exchangeable max = 12
Wald chi2(6) = 242.57
Scale parameter: 1 Prob > chi2 = 0.0000

|      | Coef. | Std. Err. | z     | P>|z| | [95% Conf. Interval] |
|------|-------|-----------|-------|------|---------------------|
| age  | .0089699 | .0053208 | 1.69  | 0.092 | -0.0014586 -.0193985 |
| grade| .0333174 | .0062352 | 5.34  | 0.000 | .0210966 .0455382  |
| 1.not_smsa | -.0715717 | .027543 | -2.60 | 0.009 | -.1255551 -.0175884 |
| 1.south | -1.017368 | .207931 | -4.89 | 0.000 | -1.424905 -.6098308 |
| year | -.0062708 | .0055314 | -1.13 | 0.257 | -.0171122 .0045706 |
| south#c.year | 1 | 0.006294 | .00258 | 3.34 | 0.001 | .0035727 .013686 |
| _cons | -.8670997 | .294771 | -2.94 | 0.003 | -1.44484 -.2893592 |
Let’s look at the correlation structure and then relax it:

```
. estat wcorrelation, format(%8.4f)
Estimated within-idcode correlation matrix R:
  | c1    | c2    | c3    | c4    | c5    | c6    | c7    |
---|-------|-------|-------|-------|-------|-------|-------|
r1 | 1.0000|       |       |       |       |       |       |
r2 | 0.4615| 1.0000|       |       |       |       |       |
r3 | 0.4615| 0.4615| 1.0000|       |       |       |       |
r4 | 0.4615| 0.4615| 0.4615| 1.0000|       |       |       |
r5 | 0.4615| 0.4615| 0.4615| 0.4615| 1.0000|       |       |
r6 | 0.4615| 0.4615| 0.4615| 0.4615| 0.4615| 1.0000|       |
r7 | 0.4615| 0.4615| 0.4615| 0.4615| 0.4615| 0.4615| 1.0000|
r8 | 0.4615| 0.4615| 0.4615| 0.4615| 0.4615| 0.4615| 0.4615|
r9 | 0.4615| 0.4615| 0.4615| 0.4615| 0.4615| 0.4615| 0.4615|
r10| 0.4615| 0.4615| 0.4615| 0.4615| 0.4615| 0.4615| 0.4615|
r11| 0.4615| 0.4615| 0.4615| 0.4615| 0.4615| 0.4615| 0.4615|
r12| 0.4615| 0.4615| 0.4615| 0.4615| 0.4615| 0.4615| 0.4615|
  | c8    | c9    | c10   | c11   | c12   |
---|-------|-------|-------|-------|-------|
r8 | 1.0000|       |       |       |       |
r9 | 0.4615| 1.0000|       |       |       |
r10| 0.4615| 0.4615| 1.0000|       |       |
r11| 0.4615| 0.4615| 0.4615| 1.0000|       |
r12| 0.4615| 0.4615| 0.4615| 0.4615| 1.0000|
```

We estimate the fixed correlation between observations within person to be 0.4615. We have many data (an average of 5.9 observations on 4,434 women), so estimating the full correlation matrix is feasible. Let’s do that and then examine the results:

```
. xtgee union age grade i.not_smsa south##c.year, family(binomial) link(probit)  
> corr(unstr) nolog
GEE population-averaged model Number of obs = 26200
Group and time vars: idcode year Number of groups = 4434
Link: probit Obs per group: min = 1
Family: binomial avg = 5.9
Correlation: unstructured max = 12
Wald chi2(6) = 198.45
Scale parameter: 1 Prob > chi2 = 0.0000
```

```
| union          | Coef. | Std. Err. | z     | P>|z|  | [95% Conf. Interval] |
|----------------|-------|-----------|-------|------|----------------------|
| age            | .0096612 | .0053366 | 1.81  | 0.070 | -.0007984 - .0201208 |
| grade          | .0352762 | .0065621 | 5.38  | 0.000 | .0224148 - .0481377 |
| 1.not_smsa     | -.093073 | .0291971 | -3.19 | 0.001 | -.1502983 - .0358478 |
| 1.south        | -1.028526 | .278802 | -3.69 | 0.000 | -1.574968 - -.4820839 |
| year           | -.0088187 | .005719 | -1.54 | 0.123 | -.0200278 - .0023904 |
| south##c.year  |       |           |       |      |                      |
| 1              | .0089824 | .0034865 | 2.58  | 0.010 | .002149 - .0158158   |
| _cons          | -.7306192 | .316757 | -2.31 | 0.021 | -1.351451 - -.109787 |
```
As before, we find that the correlation of residuals decreases as the lag increases, but more slowly than an AR(1) process.

Example 3

In this example, we examine injury incidents among 20 airlines in each of 4 years. The data are fictional, and, as a matter of fact, are really from a random-effects model.
Now there are not really enough data here to reliably estimate the correlation without any constraints of structure, but here is what happens if we try:

```
.xtgee i_cnt inprog, family(poisson) eform offset(lnpm) corr(unstr) nolog
```

GEE population-averaged model
Number of obs = 80
Group and time vars: airline time Number of groups = 20
Link: \(\log\) Obs per group: \(\min = 4\)
Family: Poisson \(\text{avg = 4.0}\)
Correlation: unstructured \(\text{max = 4}\)
Wald \(\chi^2(1) = 0.36\)
Scale parameter: 1 \(\text{Prob > chi2} = 0.5496\)

| i_cnt  | IRR   | Std. Err. | z     | P>|z|  | [95% Conf. Interval] |
|--------|-------|-----------|-------|-------|----------------------|
| inprog | .9791082 | .0345486 | -0.60 | 0.550 | .9136826 1.049219    |
| _cons  | .0078716 | .0002787 | -136.82 | 0.000 | .0073439 .0084373  |
| lnpm   | 1 (offset) |           |       |       |                      |

```
estat wcorrelation
```
Estimated within-airline correlation matrix \(R\):

\[
\begin{pmatrix}
1 & 0.5700298 & 0.716356 & 0.2383264 \\
0.5700298 & 1 & 0.4192126 & 0.3839863 \\
0.716356 & 0.4192126 & 1 & 0.3521287 \\
0.2383264 & 0.3839863 & 0.3521287 & 1
\end{pmatrix}
\]

There is no sensible pattern to the correlations.

We created this dataset from a random-effects Poisson model. We reran our data-creation program and this time had it create 400 airlines rather than 20, still with 4 years of data each. Here are the equal-correlation model and estimated correlation structure:

```
.use http://www.stata-press.com/data/r13/airacc2, clear
.xtgee i_cnt inprog, family(poisson) eform offset(lnpm) nolog
```

GEE population-averaged model
Number of obs = 1600
Group variable: airline Number of groups = 400
Link: \(\log\) Obs per group: \(\min = 4\)
Family: Poisson \(\text{avg = 4.0}\)
Correlation: exchangeable \(\text{max = 4}\)
Wald \(\chi^2(1) = 111.80\)
Scale parameter: 1 \(\text{Prob > chi2} = 0.0000\)

| i_cnt  | IRR   | Std. Err. | z     | P>|z|  | [95% Conf. Interval] |
|--------|-------|-----------|-------|-------|----------------------|
| inprog | .8915304 | .0096807 | -10.57 | 0.000 | .8727571 .9107076    |
| _cons  | .0071357 | .0000629 | -560.57 | 0.000 | .0070134 .0072601   |
| lnpm   | 1 (offset) |           |       |       |                      |

```
estat wcorrelation
```
Estimated within-airline correlation matrix \(R\):

\[
\begin{pmatrix}
1 & 0.5291707 & 0.5291707 & 0.5291707 \\
0.5291707 & 1 & 0.5291707 & 0.5291707 \\
0.5291707 & 0.5291707 & 1 & 0.5291707 \\
0.5291707 & 0.5291707 & 0.5291707 & 1
\end{pmatrix}
\]
The following estimation results assume unstructured correlation:

```
. xtgee i_cnt inprog, family(poisson) corr(unstr) eform offset(lnpm) nolog
```

GEE population-averaged model

| Estimate          | Std. Err. | z   | P>|z|  | [95% Conf. Interval] |
|-------------------|-----------|-----|------|----------------------|
| i_cnt             | .8914155  | .0096208 | -10.65 | 0.000 | .8727572 .9104728 |
| _cons             | .0071402  | .0000628 | -561.50 | 0.000 | .0070181 .0072645 |
| lnpm              | 1 (offset)|     |      |          |                     |

```

Estimated within-airline correlation matrix R:

<table>
<thead>
<tr>
<th></th>
<th>c1</th>
<th>c2</th>
<th>c3</th>
<th>c4</th>
</tr>
</thead>
<tbody>
<tr>
<td>r1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>r2</td>
<td>.4733189</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>r3</td>
<td>.5240576</td>
<td>.5748868</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>r4</td>
<td>.5139748</td>
<td>.5048895</td>
<td>.5840707</td>
<td>1</td>
</tr>
</tbody>
</table>

The equal-correlation model estimated a fixed correlation of 0.5292, and above we have correlations ranging between 0.4733 and 0.5841 with little pattern in their structure.

Also see

[XT] xtgee — Fit population-averaged panel-data models by using GEE

[U] 20 Estimation and postestimation commands
**xtgls — Fit panel-data models by using GLS**

### Syntax

```
xtgls depvar [indepvars] [if] [in] [weight] [ , options ]
```

### Options

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>noconstant</td>
<td>suppress constant term</td>
</tr>
<tr>
<td>panels(iid)</td>
<td>use i.i.d. error structure</td>
</tr>
<tr>
<td>panels(heteroskedastic)</td>
<td>use heteroskedastic but uncorrelated error structure</td>
</tr>
<tr>
<td>panels(correlated)</td>
<td>use heteroskedastic and correlated error structure</td>
</tr>
<tr>
<td>corr(independent)</td>
<td>use independent autocorrelation structure</td>
</tr>
<tr>
<td>corr(ar1)</td>
<td>use AR1 autocorrelation structure</td>
</tr>
<tr>
<td>corr(psar1)</td>
<td>use panel-specific AR1 autocorrelation structure</td>
</tr>
<tr>
<td>rho</td>
<td>specify method to compute autocorrelation parameter; see Options for details; seldom used</td>
</tr>
<tr>
<td>igls</td>
<td>use iterated GLS estimator instead of two-step GLS estimator</td>
</tr>
<tr>
<td>force</td>
<td>estimate even if observations unequally spaced in time</td>
</tr>
<tr>
<td>nmk</td>
<td>normalize standard error by ( N - k ) instead of ( N )</td>
</tr>
<tr>
<td>level(#)</td>
<td>set confidence level; default is level(95)</td>
</tr>
<tr>
<td>display_options</td>
<td>control column formats, row spacing, line width, display of omitted variables and base and empty cells, and factor-variable labeling</td>
</tr>
<tr>
<td>optimize_options</td>
<td>control the optimization process; seldom used</td>
</tr>
<tr>
<td>coeflegend</td>
<td>display legend instead of statistics</td>
</tr>
</tbody>
</table>

### Model

- **noconstant**: suppress constant term
- **panels(iid)**: use i.i.d. error structure
- **panels(heteroskedastic)**: use heteroskedastic but uncorrelated error structure
- **panels(correlated)**: use heteroskedastic and correlated error structure
- **corr(independent)**: use independent autocorrelation structure
- **corr(ar1)**: use AR1 autocorrelation structure
- **corr(psar1)**: use panel-specific AR1 autocorrelation structure
- **rho** specifies method to compute autocorrelation parameter; see Options for details; seldom used
- **igls** uses iterated GLS estimator instead of two-step GLS estimator
- **force** estimates even if observations unequally spaced in time
- **nmk** normalizes standard error by \( N - k \) instead of \( N \)
- **level(#)** sets confidence level; default is level(95)
- **display_options** control column formats, row spacing, line width, display of omitted variables and base and empty cells, and factor-variable labeling
- **optimize_options** control the optimization process; seldom used
- **coeflegend** displays legend instead of statistics

### Notes

- A panel variable must be specified. For correlation structures other than independent, a time variable must be specified. A time variable must also be specified if panels(correlated) is specified. Use `xtset`; see [XT] `xtset`.
- `indepvars` may contain factor variables; see [U] 11.4.3 Factor variables.
- `depvar` and `indepvars` may contain time-series operators; see [U] 11.4.4 Time-series varlists.
- by and `statsby` are allowed; see [U] 11.1.10 Prefix commands.
- `aweights` are allowed; see [U] 11.1.6 weight.
- `coeflegend` does not appear in the dialog box.

See [U] 20 Estimation and postestimation commands for more capabilities of estimation commands.
xtgls fits panel-data linear models by using feasible generalized least squares. This command allows estimation in the presence of AR(1) autocorrelation within panels and cross-sectional correlation and heteroskedasticity across panels.

Options

noconstant; see [R] estimation options.

panels(\textit{pdist}) specifies the error structure across panels.

- \texttt{panels(iid)} specifies a homoskedastic error structure with no cross-sectional correlation. This is the default.
- \texttt{panels(heteroskedastic)} specifies a heteroskedastic error structure with no cross-sectional correlation.
- \texttt{panels(correlated)} specifies a heteroskedastic error structure with cross-sectional correlation. If \texttt{p(c)} is specified, you must also specify a time variable (use \texttt{xtset}). The results will be based on a generalized inverse of a singular matrix unless \( T \geq m \) (the number of periods is greater than or equal to the number of panels).

\texttt{corr(\textit{corr})} specifies the assumed autocorrelation within panels.

- \texttt{corr(independent)} specifies that there is no autocorrelation. This is the default.
- \texttt{corr(ar1)} specifies that, within panels, there is AR(1) autocorrelation and that the coefficient of the AR(1) process is common to all the panels. If \texttt{c(ar1)} is specified, you must also specify a time variable (use \texttt{xtset}).
- \texttt{corr(psar1)} specifies that, within panels, there is AR(1) autocorrelation and that the coefficient of the AR(1) process is specific to each panel. \texttt{psar1} stands for panel-specific AR(1). If \texttt{c(psar1)} is specified, a time variable must also be specified; use \texttt{xtset}.

\texttt{rhotype(\textit{calc})} specifies the method to be used to calculate the autocorrelation parameter:

- \texttt{regress} regression using lags; the default
- \texttt{dw} Durbin–Watson calculation
- \texttt{freg} regression using leads
- \texttt{nagar} Nagar calculation
- \texttt{theil} Theil calculation
- \texttt{tscorr} time-series autocorrelation calculation

All the calculations are asymptotically equivalent and consistent; this is a rarely used option.

\texttt{igls} requests an iterated GLS estimator instead of the two-step GLS estimator for a nonautocorrelated model or instead of the three-step GLS estimator for an autocorrelated model. The iterated GLS estimator converges to the MLE for the \texttt{corr(independent)} models but does not for the other \texttt{corr()} models.
force specifies that estimation be forced even though the time variable is not equally spaced. This is relevant only for correlation structures that require knowledge of the time variable. These correlation structures require that observations be equally spaced so that calculations based on lags correspond to a constant time change. If you specify a time variable indicating that observations are not equally spaced, the (time dependent) model will not be fit. If you also specify force, the model will be fit, and it will be assumed that the lags based on the data ordered by the time variable are appropriate.

nwk specifies that standard errors be normalized by \( N-k \), where \( k \) is the number of parameters estimated, rather than \( N \), the number of observations. Different authors have used one or the other normalization. Greene (2012, 280) remarks that whether a degree-of-freedom correction improves the small-sample properties is an open question.

level(#)\n
display_options: noomitted, vsquish, noemptycells, baselevels, allbaselevels, nolvelabel, fwrap(#), fwrapon(style), cformat(%fmt), pformat(%fmt), sformat(%fmt), and nolstretch; see [R] estimation options.

Optimization\n
optimize_options control the iterative optimization process. These options are seldom used.

iterate(#) specifies the maximum number of iterations. When the number of iterations equals #, the optimization stops and presents the current results, even if convergence has not been reached. The default is iterate(100).

tolerance(#) specifies the tolerance for the coefficient vector. When the relative change in the coefficient vector from one iteration to the next is less than or equal to #, the optimization process is stopped. tolerance(1e-7) is the default.

nolog suppresses display of the iteration log.

The following option is available with xtgls but is not shown in the dialog box:

coefflegend; see [R] estimation options.

Remarks and examples\n
Remarks are presented under the following headings:

- Introduction
- Heteroskedasticity across panels
- Correlation across panels (cross-sectional correlation)
- Autocorrelation within panels
Introduction

Information on GLS can be found in Greene (2012), Maddala and Lahiri (2006), Davidson and MacKinnon (1993), and Judge et al. (1985).

If you have many panels relative to periods, see \texttt{xtreg} and \texttt{xtgee}. \texttt{xtgee}, in particular, provides capabilities similar to those of \texttt{xtgls} but does not allow cross-sectional correlation. On the other hand, \texttt{xtgee} allows a richer description of the correlation within panels as long as the same correlations apply to all panels. \texttt{xtgls} provides two unique features:

1. Cross-sectional correlation may be modeled (\texttt{panels(correlated)}).
2. Within panels, the AR(1) correlation coefficient may be unique (\texttt{corr(psar1)}).

\texttt{xtgls} allows models with heteroskedasticity and no cross-sectional correlation, but, strictly speaking, \texttt{xtgee} does not. \texttt{xtgee} with the \texttt{vce(robust)} option relaxes the assumption of equal variances, at least as far as the standard error calculation is concerned.

Also, \texttt{xtgls}, \texttt{panels(iid) corr(independent)} \texttt{nmk} is equivalent to \texttt{regress}. The \texttt{nmk} option uses \(n - k\) rather than \(n\) to normalize the variance calculation.

To fit a model with autocorrelated errors (\texttt{corr(ar1)} or \texttt{corr(psar1)}), the data must be equally spaced in time. To fit a model with cross-sectional correlation (\texttt{panels(correlated)}), panels must have the same number of observations (be balanced).

The equation from which the models are developed is given by

\[
y_{it} = x_{it}\beta + \epsilon_{it}
\]

where \(i = 1, \ldots, m\) is the number of units (or panels) and \(t = 1, \ldots, T_i\) is the number of observations for panel \(i\). This model can equally be written as

\[
\begin{bmatrix}
y_1 \\
y_2 \\
\vdots \\
y_m
\end{bmatrix} =
\begin{bmatrix}
x_{11} & x_{12} & \cdots & x_{1T} \\
x_{21} & x_{22} & \cdots & x_{2T} \\
\vdots & \vdots & \ddots & \vdots \\
x_{m1} & x_{m2} & \cdots & x_{mT}
\end{bmatrix} \begin{bmatrix}
\beta_1 \\
\beta_2 \\
\vdots \\
\beta_m
\end{bmatrix} +
\begin{bmatrix}
\epsilon_1 \\
\epsilon_2 \\
\vdots \\
\epsilon_m
\end{bmatrix}
\]

The variance matrix of the disturbance terms can be written as

\[
E[\epsilon\epsilon'] = \Omega =
\begin{bmatrix}
\sigma_{1,1}\Omega_{1,1} & \sigma_{1,2}\Omega_{1,2} & \cdots & \sigma_{1,m}\Omega_{1,m} \\
\sigma_{2,1}\Omega_{2,1} & \sigma_{2,2}\Omega_{2,2} & \cdots & \sigma_{2,m}\Omega_{2,m} \\
\vdots & \vdots & \ddots & \vdots \\
\sigma_{m,1}\Omega_{m,1} & \sigma_{m,2}\Omega_{m,2} & \cdots & \sigma_{m,m}\Omega_{m,m}
\end{bmatrix}
\]

For the \(\Omega_{i,j}\) matrices to be parameterized to model cross-sectional correlation, they must be square (balanced panels).

In these models, we assume that the coefficient vector \(\beta\) is the same for all panels and consider a variety of models by changing the assumptions on the structure of \(\Omega\).

For the classic OLS regression model, we have

\[
E[\epsilon_i] = 0 \\
\text{Var}[\epsilon_i] = \sigma^2 \\
\text{Cov}[\epsilon_i, \epsilon_j] = 0 \quad \text{if} \ t \neq s \text{ or } i \neq j
\]
This amounts to assuming that \( \Omega \) has the structure given by

\[
\Omega = \begin{bmatrix}
\sigma^2 I & 0 & \cdots & 0 \\
0 & \sigma^2 I & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & \sigma^2 I
\end{bmatrix}
\]

whether or not the panels are balanced (the \( 0 \) matrices may be rectangular). The classic OLS assumptions are the default \texttt{panels(iid)} and \texttt{corr(independent)} options for this command.

### Heteroskedasticity across panels

In many cross-sectional datasets, the variance for each of the panels differs. It is common to have data on countries, states, or other units that have variation of scale. The heteroskedastic model is specified by including the \texttt{panels(heteroskedastic)} option, which assumes that

\[
\Omega = \begin{bmatrix}
\sigma_1^2 I & 0 & \cdots & 0 \\
0 & \sigma_2^2 I & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & \sigma_m^2 I
\end{bmatrix}
\]

#### Example 1

Greene (2012, 1112) reprints data in a classic study of investment demand by Grunfeld and Griliches (1960). Below we allow the variances to differ for each of the five companies.

```
. use http://www.stata-press.com/data/r13/invest2
. xtgls invest market stock, panels(hetero)
```

Cross-sectional time-series FGLS regression

Coefficients: generalized least squares

Panels: heteroskedastic

Correlation: no autocorrelation

<table>
<thead>
<tr>
<th>Estimated covariances</th>
<th>Number of obs</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>100</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Estimated autocorrelations</th>
<th>Number of groups</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Estimated coefficients</th>
<th>Time periods</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>20</td>
</tr>
</tbody>
</table>

Wald chi2(2) = 865.38  Prob > chi2 = 0.0000

| invest | Coef. | Std. Err. | z     | P>|z|  | [95% Conf. Interval] |
|--------|-------|-----------|-------|------|----------------------|
| market | .0949905 | .007409  | 12.82 | 0.000 | .0804692 -.1095118  |
| stock  | .3378129 | .0302254 | 11.18 | 0.000 | .2785722 .3970535   |
| _cons  | -36.2537 | 6.124363 | -5.92 | 0.000 | -48.25723 -24.25017 |
Correlation across panels (cross-sectional correlation)

We may wish to assume that the error terms of panels are correlated, in addition to having different scale variances. The variance structure is specified by including the `panels(correlated)` option and is given by

\[
\Omega = \begin{bmatrix}
\sigma_1^2 I & \sigma_{1,2}^2 I & \cdots & \sigma_{1,m}^2 I \\
\sigma_{2,1}^2 I & \sigma_2^2 I & \cdots & \sigma_{2,m}^2 I \\
\vdots & \vdots & \ddots & \vdots \\
\sigma_{m,1}^2 I & \sigma_{m,2}^2 I & \cdots & \sigma_m^2 I
\end{bmatrix}
\]

Because we must estimate cross-sectional correlation in this model, the panels must be balanced (and \( T \geq m \) for valid results). A time variable must also be specified so that `xtgls` knows how the observations within panels are ordered. `xtset` shows us that this is true.

Example 2

```
. xtset panel variable: company (strongly balanced)
   time variable: time, 1 to 20
   delta: 1 unit
. xtgls invest market stock, panels(correlated)
```

Cross-sectional time-series FGLS regression

Coefficients: generalized least squares
Panels: heteroskedastic with cross-sectional correlation
Correlation: no autocorrelation

Estimated covariances = 15 Number of obs = 100
Estimated autocorrelations = 0 Number of groups = 5
Estimated coefficients = 3 Time periods = 20
Wald chi2(2) = 1285.19
Prob > chi2 = 0.0000

|     | Coef.  | Std. Err. | z     | P>|z| | [95% Conf. Interval] |
|-----|--------|-----------|-------|------|----------------------|
| invest | .0961894 | .0054752  | 17.57 | 0.000 | .0854583 .1069206    |
| market  | .3095321 | .0179851  | 17.21 | 0.000 | .2742819 .3447822    |
| stock   | -38.36128 | 5.344871  | -7.18 | 0.000 | -48.83703 -27.88552  |

| _cons  | -38.36128 | 5.344871  | -7.18 | 0.000 | -48.83703 -27.88552  |

The estimated cross-sectional covariances are stored in `e(Sigma)`.

```
. matrix list e(Sigma)
symmetric e(Sigma)[5,5]
   _ee | _ee 9410.9061   _ee2 -168.04631  _ee3 -1915.9538  _ee4 -1129.2896  _ee5 258.50132
   _ee2| 755.85077  _ee3 -4163.3434 | 34288.49
   _ee3| 755.85077  _ee4 -80.381742 | 2259.3242  633.42367
   _ee4| 755.85077  _ee5 -27898.235 | -1170.6801 33455.511
```

\[\text{e(Sigma)}\]
Example 3

We can obtain the MLE results by specifying the `igls` option, which iterates the GLS estimation technique to convergence:

```
.xtgls invest market stock, panels(correlated) igls
Iteration 1: tolerance = .2127384
Iteration 2: tolerance = .22817
(output omitted)
Iteration 1046: tolerance = 1.000e-07
```

Cross-sectional time-series FGLS regression
Coefficients: generalized least squares
Panels: heteroskedastic with cross-sectional correlation
Correlation: no autocorrelation

<table>
<thead>
<tr>
<th>Estimated covariances</th>
<th>15</th>
<th>Number of obs</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimated autocorrelations</td>
<td>0</td>
<td>Number of groups</td>
<td>5</td>
</tr>
<tr>
<td>Estimated coefficients</td>
<td>3</td>
<td>Time periods</td>
<td>20</td>
</tr>
<tr>
<td>Wald chi2(2)</td>
<td>558.51</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log likelihood</td>
<td>-515.4222</td>
<td>Prob &gt; chi2</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

|   | Coef.   | Std. Err. | z     | P>|z|   | [95% Conf. Interval] |
|---|---------|-----------|-------|-------|---------------------|
| invest |         |           |       |       |                     |
| market | .023631 | .004291   | 5.51  | 0.000 | 0.0152207  .0320413 |
| stock  | .1709472| .0152526  | 11.21 | 0.000 | .1410526  .2008417 |
| _cons  | -2.216508| 1.958845 | -1.13 | 0.258 | -6.055774  1.622759 |

Here the log likelihood is reported in the header of the output.

Automacorrelation within panels

The individual identity matrices along the diagonal of $\Omega$ may be replaced with more general structures to allow for serial correlation. `xtgls` allows three options so that you may assume a structure with `corr(independent)` (no autocorrelation); `corr(ar1)` (serial correlation where the correlation parameter is common for all panels); or `corr(psar1)` (serial correlation where the correlation parameter is unique for each panel).

The restriction of a common autocorrelation parameter is reasonable when the individual correlations are nearly equal and the time series are short.

If the restriction of a common autocorrelation parameter is reasonable, this allows us to use more information in estimating the autocorrelation parameter to produce a more reasonable estimate of the regression coefficients.

When you specify `corr(ar1)` or `corr(psar1)`, the iterated GLS estimator does not converge to the MLE.
Example 4

If `corr(ar1)` is specified, each group is assumed to have errors that follow the same AR(1) process; that is, the autocorrelation parameter is the same for all groups.

```
.xtreg invest market stock, panels(hetero) corr(ar1)
Cross-sectional time-series FGLS regression
Coefficients:  generalized least squares
Panels:  heteroskedastic
Correlation:  common AR(1) coefficient for all panels (0.8651)
```

| invest | Coef. | Std. Err. | z     | P>|z| | [95% Conf. Interval] |
|--------|-------|-----------|-------|-----|----------------------|
| market | .0744315 | .0097937 | 7.60  | 0.000 | .0552362 .0936268 |
| stock  | .2874294 | .0475391 | 6.05  | 0.000 | .1942545 .3806043 |
| _cons  | -18.96238 | 17.64943 | -1.07 | 0.283 | -53.55464 15.62987 |

Example 5

If `corr(psar1)` is specified, each group is assumed to have errors that follow a different AR(1) process.

```
.xtreg invest market stock, panels(iid) corr(psar1)
Cross-sectional time-series FGLS regression
Coefficients:  generalized least squares
Panels:  homoskedastic
Correlation:  panel-specific AR(1)
```

| invest | Coef. | Std. Err. | z     | P>|z| | [95% Conf. Interval] |
|--------|-------|-----------|-------|-----|----------------------|
| market | .0934343 | .0097783 | 9.56  | 0.000 | .0742693 .1125993 |
| stock  | .3838814 | .0416775 | 9.21  | 0.000 | .302195 .4655677 |
| _cons  | -10.1246 | 34.06675 | -0.30 | 0.766 | -76.8942 56.64499 |
 Stored results

`xtgls` stores the following in `e()`:

Scalars

- `e(N)`: number of observations
- `e(N_g)`: number of groups
- `e(N_t)`: number of periods
- `e(N_miss)`: number of missing observations
- `e(n_cf)`: number of estimated coefficients
- `e(n_cv)`: number of estimated covariances
- `e(n_cr)`: number of estimated correlations
- `e(df_pear)`: degrees of freedom for Pearson $\chi^2$
- `e(chi2)`: $\chi^2$
- `e(df)`: degrees of freedom
- `e(g_min)`: smallest group size
- `e(g_avg)`: average group size
- `e(g_max)`: largest group size
- `e(rank)`: rank of $e(V)$
- `e(rc)`: return code

Macros

- `e(cmd)`: `xtgls`
- `e(cmdline)`: command as typed
- `e(depvar)`: name of dependent variable
- `e(ivar)`: variable denoting groups
- `e(tvar)`: variable denoting time within groups
- `e(coeftype)`: estimation scheme
- `e(corr)`: correlation structure
- `e(vt)`: panel option
- `e(rhotype)`: type of estimated correlation
- `e(wtype)`: weight type
- `e(wexp)`: weight expression
- `e(title)`: title in estimation output
- `e(chi2type)`: Wald; type of model $\chi^2$ test
- `e(rho)`: $\rho$
- `e(properties)`: $b$ $V$
- `e(predict)`: program used to implement `predict`
- `e(asbalanced)`: factor variables `fvset` as `asbalanced`
- `e(asobserved)`: factor variables `fvset` as `asobserved`

Matrices

- `e(b)`: coefficient vector
- `e(Sigma)`: $\Sigma$ matrix
- `e(V)`: variance–covariance matrix of the estimators

Functions

- `e(sample)`: marks estimation sample

Methods and formulas

The GLS results are given by

$$\hat{\beta}_{GLS} = (X'\hat{\Omega}^{-1}X)^{-1}X'\hat{\Omega}^{-1}y$$

$$\text{Var}(\hat{\beta}_{GLS}) = (X'\hat{\Omega}^{-1}X)^{-1}$$

For all our models, the $\Omega$ matrix may be written in terms of the Kronecker product:

$$\Omega = \Sigma_{m \times m} \otimes I_{T_i \times T_i}$$
The estimated variance matrix is obtained by substituting the estimator $\hat{\Sigma}$ for $\Sigma$, where

$$
\hat{\Sigma}_{i,j} = \frac{\hat{e}_i \hat{e}_j'}{T}
$$

The residuals used in estimating $\Sigma$ are first obtained from OLS regression. If the estimation is iterated, residuals are obtained from the last fitted model.

Maximum likelihood estimates may be obtained by iterating the FGLS estimates to convergence for models with no autocorrelation, corr(independent).

The GLS estimates and their associated standard errors are calculated using $\hat{\Sigma}^{-1}$. As Beck and Katz (1995) point out, the $\Sigma$ matrix is of rank at most $\min(T,m)$ when you use the panels(correlated) option. For the GLS results to be valid (not based on a generalized inverse), $T$ must be at least as large as $m$, as you need at least as many period observations as there are panels.

Beck and Katz (1995) suggest using OLS parameter estimates with asymptotic standard errors that are corrected for correlation between the panels. This estimation can be performed with the xtpcse command; see [XT] xtpcse.

References


Also see

[XT] xtgls postestimation — Postestimation tools for xtgls

[XT] xtset — Declare data to be panel data

[XT] xtpcse — Linear regression with panel-corrected standard errors

[XT] xtreg — Fixed-, between-, and random-effects and population-averaged linear models

[XT] xtregar — Fixed- and random-effects linear models with an AR(1) disturbance

[R] regress — Linear regression

[TS] newey — Regression with Newey–West standard errors

[TS] prais — Prais–Winsten and Cochrane–Orcutt regression

[U] 20 Estimation and postestimation commands
The following postestimation commands are available after `xtgls`:

<table>
<thead>
<tr>
<th>Command</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>contrast</td>
<td>contrasts and ANOVA-style joint tests of estimates</td>
</tr>
<tr>
<td>*estat ic</td>
<td>Akaike’s and Schwarz’s Bayesian information criteria (AIC and BIC)</td>
</tr>
<tr>
<td>estat summarize</td>
<td>summary statistics for the estimation sample</td>
</tr>
<tr>
<td>estat vce</td>
<td>variance–covariance matrix of the estimators (VCE)</td>
</tr>
<tr>
<td>estimates</td>
<td>cataloging estimation results</td>
</tr>
<tr>
<td>forecast</td>
<td>dynamic forecasts and simulations</td>
</tr>
<tr>
<td>lincom</td>
<td>point estimates, standard errors, testing, and inference for linear combinations of coefficients</td>
</tr>
<tr>
<td>*lrtest</td>
<td>likelihood-ratio test</td>
</tr>
<tr>
<td>margins</td>
<td>marginal means, predictive margins, marginal effects, and average marginal effects</td>
</tr>
<tr>
<td>marginsplot</td>
<td>graph the results from margins (profile plots, interaction plots, etc.)</td>
</tr>
<tr>
<td>nlcom</td>
<td>point estimates, standard errors, testing, and inference for nonlinear combinations of coefficients</td>
</tr>
<tr>
<td>predict</td>
<td>predictions, residuals, influence statistics, and other diagnostic measures</td>
</tr>
<tr>
<td>predictnl</td>
<td>point estimates, standard errors, testing, and inference for generalized predictions</td>
</tr>
<tr>
<td>pwcompare</td>
<td>pairwise comparisons of estimates</td>
</tr>
<tr>
<td>test</td>
<td>Wald tests of simple and composite linear hypotheses</td>
</tr>
<tr>
<td>testnl</td>
<td>Wald tests of nonlinear hypotheses</td>
</tr>
</tbody>
</table>

* `estat ic` and `lrtest` are available only if `igls` and `corr(independent)` were specified at estimation.

### Syntax for predict

```
predict [type] newvar [if] [in] [, xb stdp]
```

These statistics are available both in and out of sample; type `predict ... if e(sample) ...` if wanted only for the estimation sample.

### Menu for predict

Statistics  >  Postestimation  >  Predictions, residuals, etc.
Options for predict

xb, the default, calculates the linear prediction.

stdp calculates the standard error of the linear prediction.

Also see

[XT] xtgls — Fit panel-data models by using GLS

[U] 20 Estimation and postestimation commands
**Syntax**

```
xthtaylor  depvar  indepvars  [if]  [in]  [weight],  endog(varlist)  [options]
```

**options**

**Model**

- `noconstant`  
  suppress constant term

- `endog(varlist)`  
  explanatory variables in `indepvars` to be treated as endogenous

- `constant(varlist)`_t_i
  independent variables that are constant within panel

- `varying(varlist)`_t_v
  independent variables that are time varying within panel

- `amacurdy`  
  fit model based on Amemiya and MaCurdy estimator

**SE**

- `vce(vcetype)`  
  `vcetype` may be `conventional`, `bootstrap`, or `jackknife`

**Reporting**

- `level(#)`  
  set confidence level; default is `level(95)`

- `small`  
  report small-sample statistics

*`endog(varlist)` is required.

A panel variable must be specified. For `xthtaylor`, `amacurdy`, a time variable must also be specified. Use `xtset`; see `[XT] xtset`.

`depvar`, `indepvars`, and all `varlists` may contain time-series operators; see [U] 11.4.4 Time-series varlists.

`by`, `statsby`, and `xi` are allowed; see [U] 11.1.10 Prefix commands.

`iweights` and `fweights` are allowed unless the `amacurdy` option is specified. Weights must be constant within panel; see [U] 11.1.6 weight.

See [U] 20 Estimation and postestimation commands for more capabilities of estimation commands.

**Menu**

Statistics > Longitudinal/panel data > Endogenous covariates > Hausman-Taylor regression (RE)

**Description**

`xthtaylor` fits panel-data random-effects models in which some of the covariates are correlated with the unobserved individual-level random effect. The estimators, originally proposed by Hausman and Taylor (1981) and by Amemiya and MaCurdy (1986), are based on instrumental variables. By default, `xthtaylor` uses the Hausman–Taylor estimator. When the `amacurdy` option is specified, `xthtaylor` uses the Amemiya–MaCurdy estimator.
Although the estimators implemented in xthtaylor and xtivreg (see [XT] xtivreg) use the method of instrumental variables, each command is designed for different problems. The estimators implemented in xtivreg assume that a subset of the explanatory variables in the model are correlated with the idiosyncratic error $\epsilon_{it}$. In contrast, the Hausman–Taylor and Amemiya–MaCurdy estimators that are implemented in xthtaylor assume that some of the explanatory variables are correlated with the individual-level random effects, $u_i$, but that none of the explanatory variables are correlated with the idiosyncratic error, $\epsilon_{it}$.

Options

- `noconstant;` see [R] estimation options.

- `endog(varlist)` specifies that a subset of explanatory variables in `indepvars` be treated as endogenous variables, that is, the explanatory variables that are assumed to be correlated with the unobserved random effect. `endog()` is required.

- `constant(varlist_t)` specifies the subset of variables in `indepvars` that are time invariant, that is, constant within panel. By using this option, you assert not only that the variables specified in `varlist_t` are time invariant but also that all other variables in `indepvars` are time varying. If this assertion is false, xthtaylor does not perform the estimation and will issue an error message. xthtaylor automatically detects which variables are time invariant and which are not. However, users may want to check their understanding of the data and specify which variables are time invariant and which are not.

- `varying(varlist_v)` specifies the subset of variables in `indepvars` that are time varying. By using this option, you assert not only that the variables specified in `varlist_v` are time varying but also that all other variables in `indepvars` are time invariant. If this assertion is false, xthtaylor does not perform the estimation and issues an error message. xthtaylor automatically detects which variables are time varying and which are not. However, users may want to check their understanding of the data and specify which variables are time varying and which are not.

- `amacurdy` specifies that the Amemiya–MaCurdy estimator be used. This estimator uses extra instruments to gain efficiency at the cost of additional assumptions on the data-generating process. This option may be specified only for samples containing balanced panels, and weights may not be specified. The panels must also have a common initial period.

- `vce(vcetype)` specifies the type of standard error reported, which includes types that are derived from asymptotic theory (`conventional`) and that use bootstrap or jackknife methods (`bootstrap, jackknife`); see [XT] vce_options.

  - `vce(conventional)`, the default, uses the conventionally derived variance estimator for this Hausman–Taylor model.

- `level(#);` see [R] estimation options.

  - `small` specifies that the $p$-values from the Wald tests in the output and all subsequent Wald tests obtained via `test` use $t$ and $F$ distributions instead of the large-sample normal and $\chi^2$ distributions. By default, the $p$-values are obtained using the normal and $\chi^2$ distributions.
Remarks and examples

If you have not read [XT] xt, please do so.

Consider a random-effects model of the form

\[ y_{it} = X_{1it} \beta_1 + X_{2it} \beta_2 + Z_{1i} \delta_1 + Z_{2i} \delta_2 + \mu_i + \epsilon_{it} \]

where

- \( X_{1it} \) is a 1 x \( k_1 \) vector of observations on exogenous, time-varying variables assumed to be uncorrelated with \( \mu_i \) and \( \epsilon_{it} \);
- \( X_{2it} \) is a 1 x \( k_2 \) vector of observations on endogenous, time-varying variables assumed to be (possibly) correlated with \( \mu_i \) but orthogonal to \( \epsilon_{it} \);
- \( Z_{1i} \) is a 1 x \( g_1 \) vector of observations on exogenous, time-invariant variables assumed to be uncorrelated with \( \mu_i \) and \( \epsilon_{it} \);
- \( Z_{2i} \) is a 1 x \( g_2 \) vector of observations on endogenous, time-invariant variables assumed to be (possibly) correlated \( \mu_i \) but orthogonal to \( \epsilon_{it} \);
- \( \mu_i \) is the unobserved, panel-level random effect that is assumed to have zero mean and finite variance \( \sigma^2_{\mu} \) and to be independently and identically distributed (i.i.d.) over the panels;
- \( \epsilon_{it} \) is the idiosyncratic error that is assumed to have zero mean and finite variance \( \sigma^2_{\epsilon} \) and to be i.i.d. over all the observations in the data;
- \( \beta_1, \beta_2, \delta_1, \) and \( \delta_2 \) are \( k_1 \times 1, k_2 \times 1, g_1 \times 1, \) and \( g_2 \times 1 \) coefficient vectors, respectively; and
- \( i = 1, \ldots, n, \) where \( n \) is the number of panels in the sample and, for each \( i, t = 1, \ldots, T_i. \)

Because \( X_{2it} \) and \( Z_{2i} \) may be correlated with \( \mu_i, \) the simple random-effects estimators—\( xtreg, \) \( re \) and \( xtreg, \) \( mle \)—are generally not consistent for the parameters in this model. Because the within estimator, \( xtreg, \) \( fe \), removes the \( \mu_i \) by mean-differencing the data before estimating \( \beta_1 \) and \( \beta_2, \) it is consistent for these parameters. However, in the process of removing the \( \mu_i, \) the within estimator also eliminates the \( Z_{1i} \) and the \( Z_{2i}. \) Thus it cannot estimate \( \delta_1 \) nor \( \delta_2. \) The Hausman–Taylor and Amemiya–MaCurdy estimators implemented in \( xtthaylor \) are designed to resolve this problem.

The within estimator consistently estimates \( \beta_1 \) and \( \beta_2. \) Using these estimates, we can obtain the within residuals, called \( \hat{d}_i. \) Intermediate, albeit consistent, estimates of \( \delta_1 \) and \( \delta_2—\text{called} \hat{\delta}_{11IV} \) and \( \hat{\delta}_{21IV}, \) respectively—are obtained by regressing the within residuals on \( Z_{1i} \) and \( Z_{2i}, \) using \( X_{1it} \) and \( Z_{1i} \) as instruments. The order condition for identification requires that the number of variables in \( X_{1it}, k_1, \) be at least as large as the number of elements in \( Z_{2i}, g_2 \) and that there be sufficient correlation between the instruments and \( Z_{2i} \) to avoid a weak-instrument problem.

The within estimates of \( \beta_1 \) and \( \beta_2 \) and the intermediate estimates \( \hat{\delta}_{11IV} \) and \( \hat{\delta}_{21IV} \) can be used to obtain sets of within and overall residuals. These two sets of residuals can be used to estimate the variance components (see Methods and formulas for details).

The estimated variance components can then be used to perform a GLS transform on each of the variables. For what follows, define the general notation \( \tilde{w}_{it} \) to represent the GLS transform of the variable \( w_{it}, \) \( \bar{w}_i \) to represent the within-panel mean of \( w_{it}, \) and \( \tilde{w}_{it} \) to represent the within transform of \( w_{it}. \) With this notational convention, the Hausman–Taylor (1981) estimator of the coefficients of interest can be obtained by the instrumental-variables regression

\[
\tilde{y}_{it} = \tilde{X}_{1it} \beta_1 + \tilde{X}_{2it} \beta_2 + \tilde{Z}_{1i} \delta_1 + \tilde{Z}_{2i} \delta_2 + \tilde{\mu}_i + \tilde{\epsilon}_{it} \tag{1}
\]

using \( \tilde{X}_{1it}, \tilde{X}_{2it}, \bar{X}_{1i}, \bar{X}_{2i}, \) and \( Z_{1i} \) as instruments.
For the instruments to be valid, this estimator requires that $X_{1i}$ and $Z_{1i}$ be uncorrelated with the random-effect $\mu_i$. More precisely, the instruments are valid when

$$\text{plim}_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} X_{1i} \mu_i = 0$$

and

$$\text{plim}_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} Z_{1i} \mu_i = 0$$

Amemiya and MaCurdy (1986) place stricter requirements on the instruments that vary within panels to obtain a more efficient estimator. Specifically, Amemiya and MaCurdy (1986) assume that $X_{1it}$ is orthogonal to $\mu_i$ in every period; that is, $\text{plim}_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} X_{1it} \mu_i = 0$ for $t = 1, \ldots, T$. With this restriction, they derive the Amemiya–MaCurdy estimator as the instrumental-variables regression of (1) using instruments $\tilde{X}_{1it}$, $\tilde{X}_{2it}$, $X_{1it}^\ast$, and $Z_{1i}$. The order condition for the Amemiya–MaCurdy estimator is now $Tk_1 > g_2$. \texttt{xthtaylor} uses the Amemiya–MaCurdy estimator when the \texttt{amacurdy} option is specified.

Example 1

This example replicates the results of Baltagi and Khanti-Akom (1990, table II, column HT) using 595 observations on individuals over 1976–1982 that were extracted from the Panel Study of Income Dynamics (PSID). In the model, the log-transformed wage $\text{lwage}$ is assumed to be a function of how long the person has worked for a firm, $\text{wks}$; binary variables indicating whether a person lives in a large metropolitan area or in the south, $\text{smsa}$ and $\text{south}$; marital status is $\text{ms}$; years of education, $\text{ed}$; a quadratic of work experience, $\text{exp}$ and $\text{exp}^2$; occupation, $\text{occ}$; a binary variable indicating employment in a manufacture industry, $\text{ind}$; a binary variable indicating that wages are set by a union contract, $\text{union}$; a binary variable indicating gender, $\text{fem}$; and a binary variable indicating whether the individual is African American, $\text{blk}$.

We suspect that the time-varying variables $\text{exp}$, $\text{exp}^2$, $\text{wks}$, $\text{ms}$, and $\text{union}$ are all correlated with the unobserved individual random effect. We can inspect these variables to see if they exhibit sufficient within-panel variation to serve as their own instruments.
. use http://www.stata-press.com/data/r13/psidextract
. xtsum exp exp2 wks ms union

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
<th>Observations</th>
</tr>
</thead>
<tbody>
<tr>
<td>exp</td>
<td>19.854</td>
<td>10.966</td>
<td>1</td>
<td>51</td>
<td>N = 4165</td>
</tr>
<tr>
<td></td>
<td>between</td>
<td>10.790</td>
<td>4</td>
<td>48</td>
<td>n = 595</td>
</tr>
<tr>
<td></td>
<td>within</td>
<td>2.00024</td>
<td>16.854</td>
<td>22.854</td>
<td>T = 7</td>
</tr>
<tr>
<td>exp2</td>
<td>514.405</td>
<td>496.996</td>
<td>1</td>
<td>2601</td>
<td>N = 4165</td>
</tr>
<tr>
<td></td>
<td>between</td>
<td>489.049</td>
<td>20</td>
<td>2308</td>
<td>n = 595</td>
</tr>
<tr>
<td></td>
<td>within</td>
<td>90.44581</td>
<td>231.405</td>
<td>807.405</td>
<td>T = 7</td>
</tr>
<tr>
<td>wks</td>
<td>46.812</td>
<td>5.129</td>
<td>5</td>
<td>52</td>
<td>N = 4165</td>
</tr>
<tr>
<td></td>
<td>between</td>
<td>3.284016</td>
<td>31.57143</td>
<td>51.57143</td>
<td>n = 595</td>
</tr>
<tr>
<td></td>
<td>within</td>
<td>3.941881</td>
<td>12.2401</td>
<td>63.66867</td>
<td>T = 7</td>
</tr>
<tr>
<td>ms</td>
<td>.8144058</td>
<td>.3888256</td>
<td>0</td>
<td>1</td>
<td>N = 4165</td>
</tr>
<tr>
<td></td>
<td>between</td>
<td>.3686109</td>
<td>0</td>
<td>1</td>
<td>n = 595</td>
</tr>
<tr>
<td></td>
<td>within</td>
<td>.1245274</td>
<td>-.0427371</td>
<td>1.671549</td>
<td>T = 7</td>
</tr>
<tr>
<td>union</td>
<td>.3639856</td>
<td>.4812023</td>
<td>0</td>
<td>1</td>
<td>N = 4165</td>
</tr>
<tr>
<td></td>
<td>between</td>
<td>.4543848</td>
<td>0</td>
<td>1</td>
<td>n = 595</td>
</tr>
<tr>
<td></td>
<td>within</td>
<td>.1593351</td>
<td>-.4931573</td>
<td>1.221128</td>
<td>T = 7</td>
</tr>
</tbody>
</table>

We are also going to assume that the exogenous variables occ, south, smsa, ind, fem, and blk are instruments for the endogenous, time-invariant variable ed. The output below indicates that although fem appears to be a weak instrument, the remaining instruments are probably sufficiently correlated to identify the coefficient on ed. (See Baltagi and Khanti-Akom [1990] for more discussion.)

. correlate fem blk occ south smsa ind ed
(obs=4165)

<table>
<thead>
<tr>
<th></th>
<th>fem</th>
<th>blk</th>
<th>occ</th>
<th>south</th>
<th>smsa</th>
<th>ind</th>
<th>ed</th>
</tr>
</thead>
<tbody>
<tr>
<td>fem</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>blk</td>
<td>0.2086</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>occ</td>
<td>-0.0847</td>
<td>0.0837</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>south</td>
<td>0.0516</td>
<td>0.1218</td>
<td>0.0413</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>smsa</td>
<td>0.1044</td>
<td>0.1154</td>
<td>-0.2018</td>
<td>-0.1350</td>
<td>1.0000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ind</td>
<td>-0.1778</td>
<td>-0.0475</td>
<td>0.2260</td>
<td>-0.0769</td>
<td>-0.0689</td>
<td>1.0000</td>
<td></td>
</tr>
<tr>
<td>ed</td>
<td>-0.0012</td>
<td>-0.1196</td>
<td>0.6194</td>
<td>-0.1216</td>
<td>0.1843</td>
<td>-0.2365</td>
<td>1.0000</td>
</tr>
</tbody>
</table>

We will assume that the correlations are strong enough and proceed with the estimation. The output below gives the Hausman–Taylor estimates for this model.
Hausman–Taylor estimator for error-components models

```
xthtaylor lwage occ south smsa ind exp exp2 wks ms union fem blk ed, > endog(exp exp2 wks ms union ed)
```

Hausman–Taylor estimation

<table>
<thead>
<tr>
<th></th>
<th>Number of obs = 4165</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group variable: id</td>
<td>Number of groups = 595</td>
</tr>
<tr>
<td>Obs per group: min = 7</td>
<td>avg = 7</td>
</tr>
<tr>
<td>max = 7</td>
<td></td>
</tr>
</tbody>
</table>

Random effects $u_i$ - i.i.d.

```
Wald chi2(12) = 6891.87
Prob > chi2 = 0.0000
```

|                | Coef. | Std. Err. | z     | P>|z| | [95% Conf. Interval] |
|----------------|-------|-----------|-------|------|----------------------|
| TVexogenous    |       |           |       |      |                      |
| occ            | -.0207047 | .0137809 | -1.50 | 0.133 | -.0477149 to .0063055 |
| south          | .0074398  | .031955  | 0.23  | 0.816 | -.0551908 to .0700705 |
| smsa           | -.0418334 | .0189581 | -2.21 | 0.027 | -.0789906 to -.0046761 |
| ind            | .0136039  | .0152374 | 0.89  | 0.372 | -.0162608 to .0434686 |
| TVendogenous   |       |           |       |      |                      |
| exp            | .1131328  | .002471  | 45.79 | 0.000 | .1082898 to .1179758 |
| exp2           | -.0004189 | .0000546 | -7.67 | 0.000 | -.0005259 to -.0003119 |
| wks            | .0008374  | .0005997 | 1.40  | 0.163 | -.0003381 to .00020129 |
| ms             | -.0298508 | .01898   | -1.57 | 0.116 | -.0670508 to .0073493 |
| union          | .0327714  | .0149084 | 2.20  | 0.028 | .0035514 to .0619914 |
| TIexogenous    |       |           |       |      |                      |
| fem            | -.1309236 | .126659  | -1.03 | 0.301 | -.3791707 to .1173234 |
| blk            | -.2857479 | .1557019 | -1.84 | 0.066 | -.5909179 to .0194221 |
| TIendogenous   |       |           |       |      |                      |
| ed             | .137944   | .0212485 | 6.49  | 0.000 | .0962977 to .1795902 |
| _cons          | 2.912726  | .2836522 | 10.27 | 0.000 | 2.356778 to 3.468674  |

```
sigma_u         | .94180304 |
sigma_e         | .15180273 |
rho             | .97467788  
```

(fraction of variance due to $u_i$)

Note: TV refers to time varying; TI refers to time invariant.

The estimated $\sigma_\mu$ and $\sigma_e$ are 0.9418 and 0.1518, respectively, indicating that a large fraction of the total error variance is attributed to $\mu_i$. The $z$ statistics indicate that several the coefficients may not be significantly different from zero. Whereas the coefficients on the time-invariant variables fem and blk have relatively large standard errors, the standard error for the coefficient on ed is relatively small.

Baltagi and Khanti-Akom (1990) also present evidence that the efficiency gains of the Amemiya–MaCurdy estimator over the Hausman–Taylor estimator are small for these data. This point is especially important given the additional restrictions that the estimator places on the data-generating process. The output below replicates the Baltagi and Khanti-Akom (1990) results from column AM of table II.
. xthtaylor lwage occ south smsa ind exp exp2 wks ms union fem blk ed, 
> endog(exp exp2 wks ms union ed) amacurdy

Amemiya–MaCurdy estimation  Number of obs  =  4165
Group variable: id  Number of groups  =  595
Time variable: t  Obs per group: min  =  7
avg  =  7
max  =  7

Random effects u_i ~ i.i.d.  Wald chi2(12)  =  6879.20
Prob > chi2  =  0.0000

|          | Coef. | Std. Err. | z     | P>|z| | [95% Conf. Interval] |
|----------|-------|-----------|-------|------|----------------------|
| lwage    |       |           |       |      |                      |
| TVexogenous |      |           |       |      |                      |
| occ      | -.0208498 | .0137653 | -1.51 | .130 | -.0478292 .0061297  |
| south    | .0072818  | .0319365  | 0.23  | .820 | -.0553126 .0698761  |
| smsa     | -.0419507  | .0189471  | -2.21 | .027 | -.0790864 -.0048149  |
| ind      | .0136289  | .015229   | 0.89  | .371 | -.0162194 .0434771  |
| TVendogenous |      |           |       |      |                      |
| exp      | .1129704  | .0024688  | 45.76 | .000 | .1081316 .1178093  |
| exp2     | -.0004214  | .0000546  | -7.72 | .000 | -.0005283 -.0003145 |
| wks      | .0008381   | .0005995  | 1.40  | .162 | -.0003368 .0003135  |
| ms       | -.0300894  | .0189674  | -1.59 | .113 | -.0672649 .0070861  |
| union    | .0324752   | .0148939  | 2.18  | .029 | .0032837 .0616667  |
| TIexogenous |      |           |       |      |                      |
| fem      | -.132008  | .1266039  | -1.04 | .297 | -.380147 .1161311  |
| blk      | -.2859004  | .1554857  | -1.84 | .066 | -.5906468 .0188459 |
| TIendogenous |      |           |       |      |                      |
| ed       | .1372049   | .0205695  | 6.67  | .000 | .0968894 .1775205  |
| _cons    | 2.927338   | .2751274  | 10.64 | .000 | 2.388098 3.466578 |
| sigma_u  | .94180304  |           |       |      |                      |
| sigma_e  | .15180273  |           |       |      |                      |
| rho      | .97467788  |           |       |      |                      |

Note: TV refers to time varying; TI refers to time invariant.

- Technical note

We mentioned earlier that insufficient correlation between an endogenous variable and the instruments can give rise to a weak-instrument problem. Suppose that we simulate data for a model of the form

\[ y = 3 + 3x_{1a} + 3x_{1b} + 3x_2 + 3z_1 + 3z_2 + u_i + e_{it} \]

and purposely construct the instruments so that they exhibit little correlation with the endogenous variable \( z_2 \).
. use http://www.stata-press.com/data/r13/xthtaylor1
. correlate ui z1 z2 x1a x1b x2 eit
   (obs=10000)

<table>
<thead>
<tr>
<th></th>
<th>ui</th>
<th>z1</th>
<th>z2</th>
<th>x1a</th>
<th>x1b</th>
<th>x2</th>
<th>eit</th>
</tr>
</thead>
<tbody>
<tr>
<td>ui</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>z1</td>
<td>0.0268</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>z2</td>
<td>0.8777</td>
<td>0.0286</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>x1a</td>
<td>0.0026</td>
<td>0.0079</td>
<td>0.0038</td>
<td>0.0030</td>
<td>1.0000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>x1b</td>
<td>0.8765</td>
<td>0.0191</td>
<td>0.7671</td>
<td>-0.0192</td>
<td>0.0037</td>
<td>1.0000</td>
<td></td>
</tr>
<tr>
<td>x2</td>
<td>0.0060</td>
<td>-0.0198</td>
<td>0.0123</td>
<td>-0.0100</td>
<td>-0.0138</td>
<td>0.0092</td>
<td>1.0000</td>
</tr>
<tr>
<td>eit</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In the output below, weak instruments have serious consequences on the estimates produced by xthtaylor. The estimate of the coefficient on z2 is three times larger than its true value, and its standard error is rather large. Without sufficient correlation between the endogenous variable and its instruments in a given sample, there is insufficient information for identifying the parameter. Also, given the results of Stock, Wright, and Yogo (2002), weak instruments will cause serious size distortions in any tests performed.

. xthtaylor yit x1a x1b x2 z1 z2, endog(x2 z2)

Hausman-Taylor estimation
Number of obs = 10000
Group variable: id
Number of groups = 1000
Obs per group: min = 10
avg = 10
max = 10

Random effects u_i ~ i.i.d.
Wald chi2(5) = 24172.91
Prob > chi2 = 0.0000

|           | Coef.  | Std. Err. | z     | P>|z| | [95% Conf. Interval] |
|-----------|--------|-----------|-------|-----|-----------------------|
| yit       |        |           |       |     |                       |
| TVexogenous |       |           |       |     |                       |
| x1a       | 2.959736| .0330233  | 89.63 | 0.000| 2.895011              | 3.02446 |
| x1b       | 2.953891| .0333051  | 88.69 | 0.000| 2.888614              | 3.019168|
| TVendogenous |       |           |       |     |                       |
| x2        | 3.022685| .033085   | 91.36 | 0.000| 2.957839              | 3.08753 |
| TIexogenous |       |           |       |     |                       |
| z1        | 2.709179| .587031   | 4.62  | 0.000| 1.55862               | 3.859739|
| TIendogenous |       |           |       |     |                       |
| z2        | 9.525973| 8.572966   | 1.11  | 0.266| -7.276732             | 26.32868 |
| _cons     | 2.837072| .4276595  | 6.63  | 0.000| 1.998875              | 3.675269 |

sigma_u  8.729479
sigma_e  3.1657492
rho  .88377062 (fraction of variance due to u_i)

Note: TV refers to time varying; TI refers to time invariant.

Example 2

Now let’s consider why we might want to specify the constant(varlist ti) option. For this example, we will use simulated data. In the output below, we fit a model over the full sample. Note the placement in the output of the coefficient on the exogenous variable x1c.
. use http://www.stata-press.com/data/r13/xthtaylor2
. xthtaylor yit x1a x1b x1c x2 z1 z2, endog(x2 z2)

Hausman-Taylor estimation
Number of obs = 10000
Group variable: id
Number of groups = 1000
Obs per group: min = 10
         avg = 10
         max = 10

Random effects u_i ~ i.i.d.
Wald chi2(6) = 10341.63
Prob > chi2 = 0.0000

|              | Coef.  | Std. Err. | z    | P>|z|   | [95% Conf. Interval] |
|--------------|--------|-----------|------|-------|----------------------|
| yit          |        |           |      |       |                      |
| TVexogenous  |        |           |      |       |                      |
| x1a          | 3.023647 | .0570274  | 53.02 | 0.000 | 2.911875 - 3.135418  |
| x1b          | 2.966666 | .0572659  | 51.81 | 0.000 | 2.854427 - 3.078905  |
| x1c          | .2355318 | .123502   | 1.91  | 0.057 | -.0065276 - .4775912 |
| TVendogenous |        |           |      |       |                      |
| x2           | 14.17476 | 3.128385  | 4.53  | 0.000 | 8.043234 - 20.30628  |
| TIIexogenous |        |           |      |       |                      |
| z1           | 1.741709 | .4280022  | 4.07  | 0.000 | .9028398 - 2.580578  |
| TIIendogenous|        |           |      |       |                      |
| z2           | 7.983849 | .6970903  | 11.45 | 0.000 | 6.617577 - 9.350121  |
| _cons        | 2.146038 | .3794179  | 5.66  | 0.000 | 1.402393 - 2.889684  |
| sigma_u      | 5.6787791 |            |      |       |                      |
| sigma_e      | 3.1806188 |            |      |       |                      |
| rho          | .76120931 |            |      |       | (fraction of variance due to u_i) |

Note: TV refers to time varying; TI refers to time invariant.

Now suppose that we want to fit the model using only the first eight periods. Below, x1c now appears under the TIIexogenous heading rather than the TVexogenous heading because x1c is time invariant in the subsample defined by t<9.
. xthtaylor yit xia xib xic x2 z1 z2 if t<9, endog(x2 z2)

Hausman–Taylor estimation
Number of obs = 8000
Group variable: id
Number of groups = 1000
Obs per group: min = 8
   avg = 8
   max = 8

Random effects u_i ~ i.i.d.
Wald chi2(6) = 15354.87
Prob > chi2 = 0.0000

|       | Coef.  | Std. Err. | z      | P>|z|  | [95% Conf. Interval] |
|-------|--------|-----------|--------|------|---------------------|
| yit   |        |           |        |      |                     |
| TVexogenous |   |           |        |      |                     |
| xia   | 3.051966 | 0.0367026 | 83.15  | 0.000| 2.98003 3.123901    |
| xib   | 2.967822 | 0.0368144 | 80.62  | 0.000| 2.895667 3.039977   |
| TVendogenous |   |           |        |      |                     |
| x2    | .7361217 | 3.199764  | 0.23   | 0.818| -5.5353 7.007543   |
| TIexogenous |   |           |        |      |                     |
| xic   | 3.215907 | .5657191  | 5.68   | 0.000| 2.107118 4.324696   |
| z1    | 3.347644 | .5819756  | 5.75   | 0.000| 2.206992 4.488295   |
| TIendogenous |   |           |        |      |                     |
| z2    | 2.010578 | 1.143982  | 1.76   | 0.079| -.231586 4.252742   |
| _cons | 3.257004 | .5295828  | 6.15   | 0.000| 2.219041 4.294967   |
| sigma_u | 15.445594 |           |       |      |                     |
| sigma_e | 3.175083 |           |       |      |                     |
| rho   | .95945606 |           |       |      | (fraction of variance due to u_i) |

Note: TV refers to time varying; TI refers to time invariant.

To prevent a variable from becoming time invariant, you can use either constant(varlist_t1) or varying(varlist_tv). constant(varlist_t1) specifies the subset of variables in varlist that are time invariant and requires the remaining variables in varlist to be time varying. If you specify constant(varlist_t1) and any of the variables contained in varlist_t1 are time varying, or if any of the variables not contained in varlist_t1 are time invariant, xthtaylor will not perform the estimation and will issue an error message.

. xthtaylor yit xia xib xic x2 z1 z2 if t<9, endog(x2 z2) constant(z1 z2)
   xic not included in -constant()-.
   r(198);

The same thing happens when you use the varying(varlist_tv) option.
Stored results

`xthtaylor` stores the following in `e()`:

Scalars
- `e(N)`: number of observations
- `e(_g)`: number of groups
- `e(df_m)`: model degrees of freedom
- `e(df_r)`: residual degrees of freedom (small only)
- `e(g_min)`: smallest group size
- `e(g_avg)`: average group size
- `e(g_max)`: largest group size
- `e(Tcon)`: 1 if panels balanced; 0 otherwise
- `e(sigma_u)`: panel-level standard deviation
- `e(chi2)`: standard deviation of $\epsilon_{it}$
- `e(rho)`: $\rho$
- `e(F)`: model $F$ (small only)
- `e(Tbar)`: harmonic mean of group sizes
- `e(rank)`: rank of `e(V)`

Macros
- `e(cmd)`: `xthtaylor`
- `e(cmdline)`: command as typed
- `e(depvar)`: name of dependent variable
- `e(ivar)`: variable denoting groups
- `e(tvar)`: variable denoting time within groups, `amacurdy` only
- `e(TVexogenous)`: exogenous time-varying variables
- `e(Tlexogenous)`: exogenous time-invariant variables
- `e(TVendogenous)`: endogenous time-varying variables
- `e(TIendogenous)`: endogenous time-invariant variables
- `e(wtype)`: weight type
- `e(wexp)`: weight expression
- `e(title)`: Hausman–Taylor or Amemiya–Macurdy
- `e(chi2type)`: Wald; type of model $\chi^2$ test
- `e(vce)`: vcetype specified in `vce()`
- `e(vctype)`: title used to label Std. Err.
- `e(properties)`: b V
- `e(predict)`: program used to implement `predict`

Matrices
- `e(b)`: coefficient vector
- `e(V)`: variance–covariance matrix of the estimators

Functions
- `e(sample)`: marks estimation sample

Methods and formulas

Consider an error-components model of the form

$$y_{it} = X_{1it}\beta_1 + X_{2it}\beta_2 + Z_{1i}\delta_1 + Z_{2i}\delta_2 + \mu_i + \epsilon_{it}$$

for $i = 1, \ldots, n$ and, for each $i$, $t = 1, \ldots, T_i$, of which $T_i$ periods are observed; $n$ is the number of panels in the sample. The covariates in $X$ are time varying, and the covariates in $Z$ are time invariant. Both $X$ and $Z$ are decomposed into two parts. The covariates in $X_1$ and $Z_1$ are assumed to be uncorrelated with $\mu_i$ and $\epsilon_{it}$, whereas the covariates in $X_2$ and $Z_2$ are allowed to be correlated with $\mu_i$ but not with $\epsilon_{it}$. Hausman and Taylor (1981) suggest an instrumental-variable estimator for this model.
For some variable $w$, the within transformation of $w$ is defined as

$$\tilde{w}_{it} = w_{it} - \overline{w}_i. \quad \overline{w}_i = \frac{1}{n} \sum_{t=1}^{T_i} w_{it}$$

Because the within estimator removes $Z$, the within transformation reduces the model to

$$\tilde{y}_{it} = \tilde{X}_{1it}\beta_1 + \tilde{X}_{2it}\beta_2 + \tilde{\epsilon}_{it}$$

The within estimators $\hat{\beta}_{1w}$ and $\hat{\beta}_{2w}$ are consistent for $\beta_1$ and $\beta_2$, but they may not be efficient. Also, note that the within estimator cannot estimate $\delta_1$ and $\delta_2$.

From the within estimator, we can obtain an estimate of the idiosyncratic error component, $\sigma^2_\epsilon$, as

$$\hat{\sigma}^2_\epsilon = \frac{\text{RSS}}{N - n}$$

where RSS is the residual sum of squares from the within regression and $N$ is the total number of observations in the sample.

Using the results of the within estimation, we can define

$$\tilde{d}_{it} = \tilde{y}_{it} - \overline{X}_{1it}\hat{\beta}_{1w} - \overline{X}_{2it}\hat{\beta}_{2w}$$

where $\tilde{y}_{it}$, $\overline{X}_{1it}$, and $\overline{X}_{2it}$ contain the panel level means of these variables in all observations.

Regressing $\tilde{d}_{it}$ on $Z_1$ and $Z_2$, using $X_1$ and $Z_1$ as instruments, provides intermediate, consistent estimates of $\delta_1$ and $\delta_2$, which we will call $\hat{\delta}_{11V}$ and $\hat{\delta}_{21V}$.

Using the within estimates, $\hat{\delta}_{11V}$, and $\hat{\delta}_{21V}$, we can obtain an estimate of the variance of the random effect, $\sigma^2_\mu$. First, let

$$\hat{\epsilon}_{it} = \left( y_{it} - X_{1it}\hat{\beta}_{1w} - X_{2it}\hat{\beta}_{2w} - Z_{1it}\hat{\delta}_{11V} - Z_{2it}\hat{\delta}_{21V} \right)$$

Then define

$$s^2 = \frac{1}{N} \sum_{i=1}^{n} \sum_{t=1}^{T_i} \left( \frac{1}{T_i} \sum_{t=1}^{T_i} \hat{\epsilon}_{it} \right)^2$$

Hausman and Taylor (1981) showed that, for balanced panels,

$$\text{plim}_{n \to \infty} s^2 = T\sigma^2_\mu + \sigma^2_\epsilon$$

For unbalanced panels,

$$\text{plim}_{n \to \infty} s^2 = \overline{T}\sigma^2_\mu + \sigma^2_\epsilon$$

where

$$\overline{T} = \frac{n}{\sum_{i=1}^{n} 1/T_i}$$

After we plug in $\hat{\sigma}^2_\epsilon$, our consistent estimate for $\sigma^2_\epsilon$, a little algebra suggests the estimate

$$\hat{\sigma}^2_\mu = \left( s^2 - \hat{\sigma}^2_\epsilon \right) (\overline{T})^{-1}$$
Define $\hat{\theta}_i$ as

$$\hat{\theta}_i = 1 - \left( \frac{\hat{\sigma}_e^2}{\hat{\sigma}_e^2 + T_i \hat{\sigma}_\mu^2} \right)^{\frac{1}{2}}$$

With $\hat{\theta}_i$ in hand, we can perform the standard random-effects GLS transform on each of the variables. The transform is given by

$$w_{it}^* = w_{it} - \hat{\theta}_i \bar{w}_i,$$

where $\bar{w}_i$ is the within-panel mean.

We can then obtain the Hausman–Taylor estimates of the coefficients in (2) and the conventional VCE by fitting an instrumental-variables regression of the GLS-transformed $y_{it}^*$ on $X_{it}^*$ and $Z_{it}^*$, with instruments $\tilde{X}_{it}$, $\bar{X}_{1i}$, and $Z_{1i}$.

We can obtain Amemiya–MaCurdy estimates of the coefficients in (2) and the conventional VCE by fitting an instrumental-variables regression of the GLS-transformed $y_{it}^*$ on $X_{it}^*$ and $Z_{it}^*$, using $\tilde{X}_{it}$, $\bar{X}_{1it}$, and $Z_{1i}$ as instruments, where $\bar{X}_{1it} = X_{1i1}, X_{1i2}, \ldots, X_{1iT_i}$. The order condition for the Amemiya–MaCurdy estimator is $Tk_1 > g_2$, and this estimator is available only for balanced panels.

**References**


**Also see**

[XT] **xthtaylor postestimation** — Postestimation tools for xthtaylor

[XT] **xtset** — Declare data to be panel data

[XT] **xtivreg** — Instrumental variables and two-stage least squares for panel-data models

[XT] **xtreg** — Fixed-, between-, and random-effects and population-averaged linear models

[U] 20 Estimation and postestimation commands
Description

The following postestimation commands are available after *xthtaylor*:

<table>
<thead>
<tr>
<th>Command</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>estat summarize</em></td>
<td>summary statistics for the estimation sample</td>
</tr>
<tr>
<td><em>estat vce</em></td>
<td>variance–covariance matrix of the estimators (VCE)</td>
</tr>
<tr>
<td><em>estimates</em></td>
<td>cataloging estimation results</td>
</tr>
<tr>
<td><em>forecast</em></td>
<td>dynamic forecasts and simulations</td>
</tr>
<tr>
<td><em>hausman</em></td>
<td>Hausman’s specification test</td>
</tr>
<tr>
<td><em>lincom</em></td>
<td>point estimates, standard errors, testing, and inference for linear combinations of coefficients</td>
</tr>
<tr>
<td><em>margins</em></td>
<td>marginal means, predictive margins, marginal effects, and average marginal effects</td>
</tr>
<tr>
<td><em>marginsplot</em></td>
<td>graph the results from margins (profile plots, interaction plots, etc.)</td>
</tr>
<tr>
<td><em>nlcom</em></td>
<td>point estimates, standard errors, testing, and inference for nonlinear combinations of coefficients</td>
</tr>
<tr>
<td><em>predict</em></td>
<td>predictions, residuals, influence statistics, and other diagnostic measures</td>
</tr>
<tr>
<td><em>predictnl</em></td>
<td>point estimates, standard errors, testing, and inference for generalized predictions</td>
</tr>
<tr>
<td><em>test</em></td>
<td>Wald tests of simple and composite linear hypotheses</td>
</tr>
<tr>
<td><em>testnl</em></td>
<td>Wald tests of nonlinear hypotheses</td>
</tr>
</tbody>
</table>

Syntax for predict

```
predict [ type ] newvar [ if ] [ in ] [ , statistic ]
```

<table>
<thead>
<tr>
<th>statistic</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Main</td>
<td></td>
</tr>
<tr>
<td><em>xb</em></td>
<td>$X_{it} \hat{\beta} + Z_i \hat{\delta}$, fitted values; the default</td>
</tr>
<tr>
<td><em>stdp</em></td>
<td>standard error of the fitted values</td>
</tr>
<tr>
<td><em>ue</em></td>
<td>$\hat{\mu}<em>i + \hat{\epsilon}</em>{it}$, the combined residual</td>
</tr>
<tr>
<td><em>xbu</em></td>
<td>$X_{it} \hat{\beta} + Z_i \hat{\delta} + \hat{\mu}_i$, prediction including effect</td>
</tr>
<tr>
<td><em>u</em></td>
<td>$\hat{\mu}_i$, the random-error component</td>
</tr>
<tr>
<td><em>e</em></td>
<td>$\hat{\epsilon}_{it}$, prediction of the idiosyncratic error component</td>
</tr>
</tbody>
</table>

Unstarred statistics are available both in and out of sample; type `predict ... if e(sample) ...` if wanted only for the estimation sample. Starred statistics are calculated only for the estimation sample, even when `if e(sample)` is not specified.
Menu for predict

Statistics > Postestimation > Predictions, residuals, etc.

Options for predict

Main

xb, the default, calculates the linear prediction, that is, \( X_{it} \hat{\beta} + Z_{it} \hat{\delta} \).

stdp calculates the standard error of the linear prediction.

ue calculates the prediction of \( \hat{\mu}_i + \hat{\epsilon}_{it} \).

xbu calculates the prediction of \( X_{it} \hat{\beta} + Z_{it} \hat{\delta} + \hat{\nu}_i \), the prediction including the random effect.

u calculates the prediction of \( \hat{\mu}_i \), the estimated random effect.

e calculates the prediction of \( \hat{\epsilon}_{it} \).

Remarks and examples

Example 1

Continuing with example 1 of [XT] xthtaylor, we use hausman to test whether we should use the Hausman–Taylor estimator instead of the fixed-effects estimator. We follow the empirical illustration in Baltagi (2013, sec. 7.5), but we fit the model without including the exp2 and wks variables.

We first fit the model with xthtaylor and then with xtreg, fe:

```
. use http://www.stata-press.com/data/r13/psidextract
. xthtaylor lwage occ south smsa ind exp ms union fem blk ed,
   > endog(exp ms union ed)
   (output omitted)
. estimates store eq_ht
. xtreg lwage occ south smsa ind exp ms union fem blk ed, fe
   (output omitted)
. estimates store eq_fe
```

We can now use hausman to compare the two estimators, but we need to specify the df() to indicate the degrees of freedom for the \( \chi^2 \) statistic, which would be determined by the overidentifying restrictions in the Hausman–Taylor estimation. In this case, there are three degrees of freedom because there are four time-varying exogenous variables (occ, south, smsa, ind) that can be used as instruments for only one time-invariant endogenous variable (ed).
. hausman eq_fe eq_ht, df(3)

<table>
<thead>
<tr>
<th></th>
<th>(b)</th>
<th>(B)</th>
<th>(b-B)</th>
<th>sqrt(diag(V_b-V_B))</th>
</tr>
</thead>
<tbody>
<tr>
<td>eq_fe</td>
<td>-.0239323</td>
<td>-.0231694</td>
<td>-.0007629</td>
<td>.0002395</td>
</tr>
<tr>
<td>south</td>
<td>-.0037282</td>
<td>.0062699</td>
<td>-.0099982</td>
<td>.0124188</td>
</tr>
<tr>
<td>smsa</td>
<td>-.0436251</td>
<td>-.0433518</td>
<td>-.0002733</td>
<td>.0042296</td>
</tr>
<tr>
<td>ind</td>
<td>.021184</td>
<td>.0156376</td>
<td>.0055465</td>
<td>.0025159</td>
</tr>
<tr>
<td>exp</td>
<td>.0965738</td>
<td>.0964748</td>
<td>.0000991</td>
<td>.000063</td>
</tr>
<tr>
<td>ms</td>
<td>-.0299908</td>
<td>-.0300703</td>
<td>.0000795</td>
<td>.000321</td>
</tr>
<tr>
<td>union</td>
<td>.0349156</td>
<td>.0348494</td>
<td>.0000662</td>
<td>.0006336</td>
</tr>
</tbody>
</table>

b = consistent under Ho and Ha; obtained from xtreg
B = inconsistent under Ha, efficient under Ho; obtained from xhtaylor

Test: Ho: difference in coefficients not systematic

\[ \chi^2(3) = (b-B)'[(V_{b-V_B})^{-1}](b-B) \]

= 5.22

Prob>\chi^2 = 0.1567

(V_{b-V_B} is not positive definite)

The \( p \)-value for the test provides evidence favoring the null hypothesis; therefore, in this case, the Hausman–Taylor estimation is adequate.

Notice that the variance–covariance matrix for the difference \( (b-B) \) is not positive definite. As Greene (2012, 237) points out, this kind of result is due to finite-sample conditions. He also states that Hausman considers it preferable to take the test statistic as zero and, therefore, not to reject the null hypothesis.

Example 2

We now want to determine whether the Amemiya–MaCurdy estimator produces significant efficiency gains with respect to the Hausman–Taylor estimator. We refit the two models, and we use the Hausman test again:

. use http://www.stata-press.com/data/r13/psidextract
. xhtaylor lwage occ south smsa ind exp ms union fem blk ed, endog(exp ms union ed)
(output omitted)
. estimates store eq_ht
. xhtaylor lwage occ south smsa ind exp ms union fem blk ed, endog(exp ms union ed) amacurdy
(output omitted)
. estimates store eq_am
. hausman eq_ht eq_am

<table>
<thead>
<tr>
<th></th>
<th>(b)</th>
<th>(B)</th>
<th>(b-B)</th>
<th>sqrt(diag(V_b-V_B))</th>
</tr>
</thead>
<tbody>
<tr>
<td>occ</td>
<td>-.0231694</td>
<td>-.023354</td>
<td>.0001846</td>
<td>.0006485</td>
</tr>
<tr>
<td>south</td>
<td>.0062699</td>
<td>.0060857</td>
<td>.0001842</td>
<td>.0010641</td>
</tr>
<tr>
<td>smsa</td>
<td>-.0433518</td>
<td>-.0434638</td>
<td>.0001121</td>
<td>.0006297</td>
</tr>
<tr>
<td>ind</td>
<td>.0156376</td>
<td>.0156602</td>
<td>-.0000226</td>
<td>.000492</td>
</tr>
<tr>
<td>exp</td>
<td>.0964748</td>
<td>.0962147</td>
<td>.00026</td>
<td>.000694</td>
</tr>
<tr>
<td>ms</td>
<td>-.0300703</td>
<td>-.0303139</td>
<td>.0002436</td>
<td>.0006735</td>
</tr>
<tr>
<td>union</td>
<td>.0348494</td>
<td>.0345742</td>
<td>.0002752</td>
<td>.0006471</td>
</tr>
<tr>
<td>fem</td>
<td>-.1277756</td>
<td>-.1287857</td>
<td>.0010101</td>
<td>.0036717</td>
</tr>
<tr>
<td>blk</td>
<td>-.2911574</td>
<td>-.291645</td>
<td>.0004876</td>
<td>.0082831</td>
</tr>
<tr>
<td>ed</td>
<td>.1390257</td>
<td>.1380699</td>
<td>.0009558</td>
<td>.005436</td>
</tr>
</tbody>
</table>

b = consistent under Ho and Ha; obtained from xthtaylor
B = inconsistent under Ha, efficient under Ho; obtained from xthtaylor

Test: Ho: difference in coefficients not systematic

\[
\text{chi2}(10) = (b-B)\left((V_b-V_B)^{-1}\right)(b-B) = 14.42
\]

Prob>chi2 = 0.1548

The result indicates that we should use the more efficient estimation produced by the Amemiya–MaCurdy estimator.

References


Also see

[XT] xthtaylor — Hausman–Taylor estimator for error-components models

[U] 20 Estimation and postestimation commands
xtintreg — Random-effects interval-data regression models

Syntax

```
xтинтreg depvar_{lower} depvar_{upper} [ indepvars ] [ if ] [ in ] [ weight ] [ , options ]
```

**options**

Model

```
noconstant
offset(varname)
collinear
```

suppress constant term
include varname in model with coefficient constrained to 1
keep collinear variables

SE

```
vce(vcetype)
```

vcetype may be oim, bootstrap, or jackknife

Reporting

```
level(#)
noskip
intreg
noconsreport
display_options
```

set confidence level; default is level(95)
perform overall model test as a likelihood-ratio test
perform likelihood-ratio test against pooled model
do not display constraints
control column formats, row spacing, line width, display of omitted variables and base and empty cells, and factor-variable labeling

Integration

```
intmethod(intmethod)
intpoints(#)
```

integration method; intmethod may be mvaghermite (the default) or ghermite
use # quadrature points; default is intpoints(12)

Maximization

```
maximize_options
coeflegend
```

control the maximization process; see [R] maximize
display legend instead of statistics

A panel variable must be specified; use xtset; see [XT] xtset.

indepvars may contain factor variables; see [U] 11.4.3 Factor variables.
depvar_{lower}, depvar_{upper}, and indepvars may contain time-series operators; see [U] 11.4.4 Time-series varlists.
by and stataby are allowed; see [U] 11.1.10 Prefix commands.
iweights are allowed; see [U] 11.1.6 weight. Weights must be constant within panel.
coeflegend does not appear in the dialog box.
See [U] 20 Estimation and postestimation commands for more capabilities of estimation commands.
xtintreg — Random-effects interval-data regression models

Menu
Statistics > Longitudinal/panel data > Censored outcomes > Interval regression (RE)

Description

xtintreg fits a random-effects regression model whose dependent variable may be measured as point data, interval data, left-censored data, or right-censored data. depvar\text{lower} and depvar\text{upper} represent how the dependent variable was measured.

The values in depvar\text{lower} and depvar\text{upper} should have the following form:

<table>
<thead>
<tr>
<th>Type of data</th>
<th>depvar\text{lower}</th>
<th>depvar\text{upper}</th>
</tr>
</thead>
<tbody>
<tr>
<td>point data</td>
<td>(a = [a, a])</td>
<td>(a)</td>
</tr>
<tr>
<td>interval data</td>
<td>([a, b])</td>
<td>(a)</td>
</tr>
<tr>
<td>left-censored data</td>
<td>((-\infty, b])</td>
<td>(.)</td>
</tr>
<tr>
<td>right-censored data</td>
<td>([a, +\infty))</td>
<td>(a)</td>
</tr>
</tbody>
</table>

Options

Model

\text{noconstant}, \text{offset}(\text{varname}), \text{constraints}(\text{constraints}), \text{collinear}; \text{see [R] estimation options.}

SE

\text{vce(vctype)} specifies the type of standard error reported, which includes types that are derived from asymptotic theory (\text{oim}) and that use bootstrap or jackknife methods (\text{bootstrap, jackknife}); \text{see [XT] vce\_options.}

Reporting

\text{level(#)}, \text{noskip}; \text{see [R] estimation options.}

\text{intreg} specifies that a likelihood-ratio test comparing the random-effects model with the pooled (intreg) model be included in the output.

\text{nocsnsreport}; \text{see [R] estimation options.}

\text{display\_options: noomitted, vsquish, noemptycells, baselevels, allbaselevels, nolabel, fvwrap(\#), fvwrapon(style), cformat(\%fmt), pformat(\%fmt), sformat(\%fmt)}, and \text{nolstretch}; \text{see [R] estimation options.}

Integration

\text{intmethod(intmethod)}, \text{intpoints(#)}; \text{see [R] estimation options.}

Maximization

\text{maximize\_options: difficult, technique(algorithm\_spec), iterate(\#), [no]log, trace, gradient, showstep, hessian, showtolerance, tolerance(\#), ltolerance(\#), nrtolerance(\#), nonrtolerance, and from(init\_specs)}; \text{see [R] maximize.} These options are seldom used.
The following option is available with \texttt{xtintreg} but is not shown in the dialog box: \texttt{coeflegend}; see \texttt{[R] estimation options}.

\section*{Remarks and examples}

Consider the linear regression model with panel-level random effects

\[ y_{it} = x_{it} \beta + \nu_i + \epsilon_{it} \]

for \( i = 1, \ldots, n \) panels, where \( t = 1, \ldots, n_i \). The random effects, \( \nu_i \), are i.i.d., \( N(0, \sigma^2_\nu) \), and \( \epsilon_{it} \) are i.i.d., \( N(0, \sigma^2_\epsilon) \) independently of \( \nu_i \). The observed data consist of the couples, \((y_{1it}, y_{2it})\), such that all that is known is that \( y_{1it} \leq y_{it} \leq y_{2it} \), where \( y_{1it} \) is possibly \(-\infty\) and \( y_{2it} \) is possibly \(+\infty\).

\section*{Example 1}

We begin with the \texttt{nlswork} dataset described in \texttt{[XT] xt} and create two fictional dependent variables, where the wages are instead reported sometimes as ranges. The wages have been adjusted to 1988 dollars and have further been recoded such that some of the observations are known exactly, some are left-censored, some are right-censored, and some are known only in an interval.

We wish to fit a random-effects interval regression model of adjusted (log) wages:

\begin{verbatim}
. use http://www.stata-press.com/data/r13/nlswork5
(National Longitudinal Survey. Young Women 14-26 years of age in 1968)
. xtintreg ln_wage1 ln_wage2 union age grade south##c.year occ_code, intreg
(output omitted)
\end{verbatim}

\begin{verbatim}
Random-effects interval regression
Number of obs = 19151
Group variable: idcode Number of groups = 4140
Random effects u_i ~ Gaussian
Obs per group: min = 1
avg = 4.6
max = 12
Integration method: mvaghermite
Integration points = 12
Wald chi2(7) = 2523.84
Log likelihood = -23174.355 Prob > chi2 = 0.0000
\end{verbatim}

\begin{verbatim}
Coef. Std. Err. z P>|z| [95% Conf. Interval]
union .1441844 .0094245 15.30 0.000 .1257128 .162656
age .0104083 .0018804 5.54 0.000 .0067228 .0140939
grade .0794958 .0023469 33.87 0.000 .074896 .0840955
1.south -.3778103 .0979415 -3.86 0.000 -.5697722 -.1858485
year .0013528 .0020176 0.67 0.503 -.0026016 .0053071
south#c.year 1 .0034385 .0012105 2.84 0.005 .0010659 .005811
occ_code -.0197912 .0014094 -14.04 0.000 -.0225535 -.0170289
_cons .3791078 .1136641 3.34 0.001 .1563303 .6018853
/\sigma_u .2987074 .0052697 56.68 0.000 .2883789 .309036
/\sigma_e .3528109 .0030935 114.05 0.000 .3467478 .358874
\end{verbatim}

Likelihood-ratio test of \( \sigma_u=0 \): chibar2(01)= 2516.85 Prob>=chibar2 = 0.000
The output includes the overall and panel-level variance components (labeled $\sigma_e$ and $\sigma_u$, respectively) together with $\rho$ (labeled rho),

$$\rho = \frac{\sigma^2_u}{\sigma^2_e + \sigma^2_u},$$

which is the proportion of the total variance contributed by the panel-level variance component.

When rho is zero, the panel-level variance component is unimportant, and the panel estimator is not different from the pooled estimator. A likelihood-ratio test of this is included at the bottom of the output. This test formally compares the pooled estimator (intreg) with the panel estimator.

---

Technical note

The random-effects model is calculated using quadrature, which is an approximation whose accuracy depends partially on the number of integration points used. We can use the quadchk command to see if changing the number of integration points affects the results. If the results change, the quadrature approximation is not accurate given the number of integration points. Try increasing the number of integration points using the `intpoints()` option and run quadchk again. Do not attempt to interpret the results of estimates when the coefficients reported by quadchk differ substantially. See [XT] quadchk for details and [XT] xtprob for an example.

Because the xtintreg likelihood function is calculated by Gauss–Hermite quadrature, on large problems the computations can be slow. Computation time is roughly proportional to the number of points used for the quadrature.
xtintreg stores the following in e():

Scalars
- e(N) number of observations
- e(N_\text{g}) number of groups
- e(N_{\text{unc}}) number of uncensored observations
- e(N_{\text{lc}}) number of left-censored observations
- e(N_{\text{rc}}) number of right-censored observations
- e(N_{\text{int}}) number of interval observations
- e(N_{\text{cd}}) number of completely determined observations
- e(k) number of parameters
- e(k_{\text{aux}}) number of auxiliary parameters
- e(k_{\text{eq}}) number of equations in e(b)
- e(k_{\text{eq, model}}) number of equations in overall model test
- e(k_{\text{dv}}) number of dependent variables
- e(df_{\text{m}}) model degrees of freedom
- e(ll) log likelihood
- e(ll_{0}) log likelihood, constant-only model
- e(chi2) $\chi^2$
- e(chi2_{\text{c}}) $\chi^2$ for comparison test
- e(rho) $\rho$
- e(sigma_{\text{u}}) panel-level standard deviation
- e(sigma_{\text{e}}) standard deviation of $\epsilon_{it}$
- e(n_{\text{quad}}) number of quadrature points
- e(g_{\text{min}}) smallest group size
- e(g_{\text{avg}}) average group size
- e(g_{\text{max}}) largest group size
- e(p) significance
- e(rank) rank of e(V)
- e(rank_{0}) rank of e(V) for constant-only model
- e(ic) number of iterations
- e(rc) return code
- e(converged) 1 if converged, 0 otherwise
xtintreg — Random-effects interval-data regression models

Macros

- `e(cmd)` xtintreg
- `e(cmdline)` command as typed
- `e(depvar)` names of dependent variables
- `e(ivar)` variable denoting groups
- `e(wtype)` weight type
- `e(wexp)` weight expression
- `e(title)` title in estimation output
- `e(offset1)` offset
- `e(chi2type)` Wald or LR; type of model \( \chi^2 \) test
- `e(chi2_ct)` Wald or LR; type of model \( \chi^2 \) test corresponding to `e(chi2_c)`
- `e(vce)` vcetype specified in `vce()`
- `e(vcetype)` title used to label Std. Err.
- `e(intmethod)` integration method
- `e(distrib)` Gaussian; the distribution of the random effect
- `e(opt)` type of optimization
- `e(which)` max or min; whether optimizer is to perform maximization or minimization
- `e(ml_method)` type of ml method
- `e(user)` name of likelihood-evaluator program
- `e(technique)` maximization technique
- `e(properties)` b V
- `e(predict)` program used to implement predict
- `e(asbalanced)` factor variables fvset as asbalanced
- `e(asobserved)` factor variables fvset as asobserved

Matrices

- `e(b)` coefficient vector
- `e(Cns)` constraints matrix
- `e(ilog)` iteration log
- `e(gradient)` gradient vector
- `e(V)` variance–covariance matrix of the estimators

Functions

- `e(sample)` marks estimation sample

Methods and formulas

Assuming a normal distribution, \( N(0, \sigma^2_\nu) \), for the random effects \( \nu_i \), we have the joint (unconditional of \( \nu_i \)) density of the observed data for the \( i \)th panel

\[
f \{ (y_{1i1}, y_{2i1}), \ldots, (y_{1in_i}, y_{2in_i}) | x_{1i}, \ldots, x_{in_i} \} =
\quad \int_{-\infty}^{\infty} \frac{e^{-\nu_i^2/2\sigma^2_\nu}}{\sqrt{2\pi\sigma_\nu}} \left\{ \prod_{t=1}^{n_i} F(y_{1it}, y_{2it}, x_{it}\beta + \nu_i) \right\} d\nu_i
\]
where

\[
F(y_{1it}, y_{2it}, \Delta_{it}) = \begin{cases} 
\left(\sqrt{2\pi\sigma_c}\right)^{-1} e^{-(y_{1it}-\Delta_{it})^2/(2\sigma_c^2)} & \text{if } (y_{1it}, y_{2it}) \in C \\
\Phi\left(\frac{y_{2it}-\Delta_{it}}{\sigma_c}\right) & \text{if } (y_{1it}, y_{2it}) \in L \\
1 - \Phi\left(\frac{y_{1it}-\Delta_{it}}{\sigma_c}\right) & \text{if } (y_{1it}, y_{2it}) \in R \\
\Phi\left(\frac{y_{2it}-\Delta_{it}}{\sigma_c}\right) - \Phi\left(\frac{y_{1it}-\Delta_{it}}{\sigma_c}\right) & \text{if } (y_{1it}, y_{2it}) \in I
\end{cases}
\]

where \(C\) is the set of noncensored observations \((y_{1it} = y_{2it} \text{ and both nonmissing})\), \(L\) is the set of left-censored observations \((y_{1it} \text{ missing and } y_{2it} \text{ nonmissing})\), \(R\) is the set of right-censored observations \((y_{1it} \text{ nonmissing and } y_{2it} \text{ missing})\), \(I\) is the set of interval observations \((y_{1it} < y_{2it} \text{ and both nonmissing})\), and \(\Phi()\) is the cumulative normal distribution.

The panel-level likelihood \(l_i\) is given by

\[
l_i = \int_{-\infty}^{\infty} e^{-\nu_i^2/2\sigma_i^2} \left\{ \prod_{t=1}^{n_i} F(y_{1it}, y_{2it}, x_{it}\beta + \nu_i) \right\} d\nu_i
\]

This integral can be approximated with \(M\)-point Gauss–Hermite quadrature

\[
\int_{-\infty}^{\infty} e^{-x^2} h(x) dx \approx \sum_{m=1}^{M} w_m^* h(a_m^*)
\]

This is equivalent to

\[
\int_{-\infty}^{\infty} f(x) dx \approx \sum_{m=1}^{M} w_m^* \exp \left\{ (a_m^*)^2 \right\} f(a_m^*)
\]

where the \(w_m^*\) denote the quadrature weights and the \(a_m^*\) denote the quadrature abscissas. The log likelihood, \(L_i\), is the sum of the logs of the panel-level likelihoods \(l_i\).

The default approximation of the log likelihood is by adaptive Gauss–Hermite quadrature, which approximates the panel-level likelihood with

\[
l_i \approx \sqrt{2\tilde{\sigma}_i} \sum_{m=1}^{M} w_m^* \exp \left\{ (a_m^*)^2 \right\} g(y_{1it}, y_{2it}, x_{it}, \sqrt{2\tilde{\sigma}_i}a_m^* + \hat{\mu}_i)
\]

where \(\tilde{\sigma}_i\) and \(\hat{\mu}_i\) are the adaptive parameters for panel \(i\). Therefore, using the definition of \(g(y_{1it}, y_{2it}, x_{it}, \nu_i)\), the total log likelihood is approximated by
\[ L \approx \sum_{i=1}^{n} w_i \log \left[ \sqrt{2\hat{\sigma}_i} \sum_{m=1}^{M} w_m^* \exp\left\{ a_m^* \right\} \exp\left\{ -(\sqrt{2\hat{\sigma}_i}a_m^* + \hat{\mu}_i)^2 / 2\sigma^2_{\nu} \right\} \right] \]

where \( w_i \) is the user-specified weight for panel \( i \); if no weights are specified, \( w_i = 1 \).

The default method of adaptive Gauss–Hermite quadrature is to calculate the posterior mean and variance and use those parameters for \( \hat{\mu}_i \) and \( \hat{\sigma}_i \) by following the method of Naylor and Smith (1982), further discussed in Skrondal and Rabe-Hesketh (2004). We start with \( \hat{\sigma}_{i,0} = 1 \) and \( \hat{\mu}_{i,0} = 0 \), and the posterior means and variances are updated in the \( k \)th iteration. That is, at the \( k \)th iteration of the optimization for \( l_i \) we use

\[ l_{i,k} \approx \sum_{m=1}^{M} \sqrt{2\hat{\sigma}_{i,k-1}} w_m^* \exp\left\{ a_m^* \right\} g(y_{1it}, y_{2it}, x_{it}, \sqrt{2\hat{\sigma}_{i,k-1}}a_m^* + \hat{\mu}_{i,k-1}) \]

Letting

\[ \tau_{i,m,k-1} = \sqrt{2\hat{\sigma}_{i,k-1}}a_m^* + \hat{\mu}_{i,k-1} \]

\[ \hat{\mu}_{i,k} = \sum_{m=1}^{M} \frac{\sqrt{2\hat{\sigma}_{i,k-1}} w_m^* \exp\left\{ (a_m^*)^2 \right\} g(y_{1it}, y_{2it}, x_{it}, \tau_{i,m,k-1})}{l_{i,k}} \]

and

\[ \hat{\sigma}_{i,k} = \sum_{m=1}^{M} \frac{(\tau_{i,m,k-1})^2 \sqrt{2\hat{\sigma}_{i,k-1}} w_m^* \exp\left\{ (a_m^*)^2 \right\} g(y_{1it}, y_{2it}, x_{it}, \tau_{i,m,k-1})}{l_{i,k}} - (\hat{\mu}_{i,k})^2 \]

and this is repeated until \( \hat{\mu}_{i,k} \) and \( \hat{\sigma}_{i,k} \) have converged for this iteration of the maximization algorithm. This adaptation is applied on every iteration until the log-likelihood change from the preceding iteration is less than a relative difference of 1e–6; after this, the quadrature parameters are fixed.

The log likelihood can also be calculated by nonadaptive Gauss–Hermite quadrature, the int-method(ghermite) option:

\[ L = \sum_{i=1}^{n} w_i \log f \left\{ (y_{1i1}, y_{2i1}), \ldots, (y_{1ini}, y_{2ini}) | x_{1i1}, \ldots, x_{ini} \right\} \]

\[ \approx \sum_{i=1}^{n} w_i \log \left\{ \frac{1}{\sqrt{n}} \sum_{m=1}^{M} w_m^* \prod_{t=1}^{n_i} F \left( y_{1it}, y_{2it}, x_{it} \beta + \sqrt{2\sigma_{\nu}}a_m^* \right) \right\} \]

Both quadrature formulas require that the integrated function be well approximated by a polynomial of degree equal to the number of quadrature points. The number of periods (panel size) can affect whether

\[ \prod_{t=1}^{n_i} F(y_{1it}, y_{2it}, x_{it} \beta + \nu_i) \]
is well approximated by a polynomial. As panel size and \( \rho \) increase, the quadrature approximation can become less accurate. For large \( \rho \), the random-effects model can also become unidentified. Adaptive quadrature gives better results for correlated data and large panels than nonadaptive quadrature; however, we recommend that you use the \texttt{quadchk} command (see \texttt{[XT] quadchk}) to verify the quadrature approximation used in this command, whichever approximation you choose.

References


Also see

- \texttt{[XT] xtintreg postestimation} — Postestimation tools for xtintreg
- \texttt{[XT] quadchk} — Check sensitivity of quadrature approximation
- \texttt{[XT] xtreg} — Fixed-, between-, and random-effects and population-averaged linear models
- \texttt{[XT] xtset} — Declare data to be panel data
- \texttt{[XT] xtobit} — Random-effects tobit models
- \texttt{[R] intreg} — Interval regression
- \texttt{[R] tobit} — Tobit regression
- \texttt{[U] 20 Estimation and postestimation commands}
Description

The following postestimation commands are available after xtintreg:

<table>
<thead>
<tr>
<th>Command</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>contrast</td>
<td>contrasts and ANOVA-style joint tests of estimates</td>
</tr>
<tr>
<td>estat ic</td>
<td>Akaike’s and Schwarz’s Bayesian information criteria (AIC and BIC)</td>
</tr>
<tr>
<td>estat summarize</td>
<td>summary statistics for the estimation sample</td>
</tr>
<tr>
<td>estat vce</td>
<td>variance–covariance matrix of the estimators (VCE)</td>
</tr>
<tr>
<td>estimates</td>
<td>cataloging estimation results</td>
</tr>
<tr>
<td>lincom</td>
<td>point estimates, standard errors, testing, and inference for linear combinations of coefficients</td>
</tr>
<tr>
<td>lrtest</td>
<td>likelihood-ratio test</td>
</tr>
<tr>
<td>margins</td>
<td>marginal means, predictive margins, marginal effects, and average marginal effects</td>
</tr>
<tr>
<td>marginsplot</td>
<td>graph the results from margins (profile plots, interaction plots, etc.)</td>
</tr>
<tr>
<td>nlcom</td>
<td>point estimates, standard errors, testing, and inference for nonlinear combinations of coefficients</td>
</tr>
<tr>
<td>predict</td>
<td>predictions, residuals, influence statistics, and other diagnostic measures</td>
</tr>
<tr>
<td>predictnl</td>
<td>point estimates, standard errors, testing, and inference for generalized predictions</td>
</tr>
<tr>
<td>pwcompare</td>
<td>pairwise comparisons of estimates</td>
</tr>
<tr>
<td>test</td>
<td>Wald tests of simple and composite linear hypotheses</td>
</tr>
<tr>
<td>testnl</td>
<td>Wald tests of nonlinear hypotheses</td>
</tr>
</tbody>
</table>

Syntax for predict

```
predict [type] newvar [if] [in] [, statistic nooffset]
```

<table>
<thead>
<tr>
<th>statistic</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Main</td>
<td>linear prediction assuming a zero random effect, the default</td>
</tr>
<tr>
<td>xb</td>
<td>standard error of the linear prediction</td>
</tr>
<tr>
<td>stdf</td>
<td>standard error of the linear forecast</td>
</tr>
<tr>
<td>pr0(a,b)</td>
<td>Pr(a &lt; y &lt; b) assuming a zero random effect</td>
</tr>
<tr>
<td>e0(a,b)</td>
<td>E(y</td>
</tr>
<tr>
<td>ystar0(a,b)</td>
<td>E(y*), y* = max{a, min(y_j, b)} assuming a zero random effect</td>
</tr>
</tbody>
</table>

These statistics are available both in and out of sample; type `predict ... if e(sample) ...` if wanted only for the estimation sample.
where \( a \) and \( b \) may be numbers or variables; \( a \) missing \((a \geq .)\) means \(-\infty\), and \( b \) missing \((b \geq .)\) means \(+\infty\); see \([U]\) 12.2.1 Missing values.

### Menu for predict

Statistics > Postestimation > Predictions, residuals, etc.

### Options for predict

- **xb**, the default, calculates the linear prediction.
- **stdp** calculates the standard error of the linear prediction. It can be thought of as the standard error of the predicted expected value or mean for the observation’s covariate pattern. The standard error of the prediction is also referred to as the standard error of the fitted value.
- **stdf** calculates the standard error of the linear forecast. This is the standard error of the point prediction for 1 observation. It is commonly referred to as the standard error of the future or forecast value. By construction, the standard errors produced by **stdf** are always larger than those produced by **stdp**; see Methods and formulas in \([R]\) **regress**.
- **pr0\((a, b)\)** calculates estimates of \( \Pr(a < y < b | x = x_{it}, \nu_i = 0) \), which is the probability that \( y \) would be observed in the interval \((a, b)\), given the current values of the predictors, \( x_{it} \), and given a zero random effect. In the discussion that follows, these two conditions are implied.
  - \( a \) and \( b \) may be specified as numbers or variable names; \( lb \) and \( ub \) are variable names;
  - \( pr0(20, 30) \) calculates \( \Pr(20 < y < 30) \);
  - \( pr0(lb, ub) \) calculates \( \Pr(lb < y < ub) \); and
  - \( pr0(20, ub) \) calculates \( \Pr(20 < y < ub) \).
  - \( a \) missing \((a \geq .)\) means \(-\infty\); \( pr0(., 30) \) calculates \( \Pr(-\infty < y < 30) \);
  - \( pr0(lb, 30) \) calculates \( \Pr(-\infty < y < 30) \) in observations for which \( lb \geq . \) (and calculates \( \Pr(lb < y < 30) \) elsewhere).
  - \( b \) missing \((b \geq .)\) means \(+\infty\); \( pr0(20, .) \) calculates \( \Pr(+\infty > y > 20) \);
  - \( pr0(20, ub) \) calculates \( \Pr(+\infty > y > 20) \) in observations for which \( ub \geq . \) (and calculates \( \Pr(20 < y < ub) \) elsewhere).
- **e0\((a, b)\)** calculates estimates of \( E(y | a < y < b, x = x_{it}, \nu_i = 0) \), which is the expected value of \( y \) conditional on \( y \) being in the interval \((a, b)\), meaning that \( y \) is truncated. \( a \) and \( b \) are specified as they are for **pr0()**.
- **ystar0\((a, b)\)** calculates estimates of \( E(y^* | x = x_{it}, \nu_i = 0) \), where \( y^* = a \) if \( y \leq a \), \( y^* = b \) if \( y \geq b \), and \( y^* = y \) otherwise, meaning that \( y^* \) is the censored version of \( y \). \( a \) and \( b \) are specified as they are for **pr0()**.
- **nooffset** is relevant only if you specified offset(varname) for xtintreg. It modifies the calculations made by predict so that they ignore the offset variable; the linear prediction is treated as \( x_{it}\beta \) rather than \( x_{it}\beta + \text{offset}_{it} \).
Remarks and examples

Example 1

In example 1 of [XT] xtintreg, we fit a random-effects model of wages. Say that we want to know how union membership status affects the probability that a worker’s wage will be “low”, where low means a log wage that is less than the 20th percentile of all observations in our dataset. First, we use centile to find the 20th percentile of ln_wage:

```
use http://www.stata-press.com/data/r13/nlswork5
(National Longitudinal Survey. Young Women 14-26 years of age in 1968)
xтинг рег ln_wage1 ln_wage2 i.union age grade south##c.year, intreg
(output omitted)
```

```
centile ln_wage, centile(20)
```

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>Percentile</th>
<th>Centile</th>
<th>[95% Conf. Interval]</th>
</tr>
</thead>
<tbody>
<tr>
<td>ln_wage</td>
<td>28534</td>
<td>20</td>
<td>1.301507</td>
<td>1.297063 1.308635</td>
</tr>
</tbody>
</table>

Now we use margins to obtain the effect of union status on the probability that the log of wages is in the bottom 20% of women. Given the results from centile that corresponds to the log of wages being below 1.30. We evaluate the effect for two groups: 1) women age 30 living in the south in 1988 who graduated high school, but had no more schooling, and 2) the same group of women, with the exception that they are college graduates (grade=16).

```
margins, dydx(union) predict(pr0(.,1.30))
> at(age=30 south=1 year=88 grade=12 union=0)
> at(age=30 south=1 year=88 grade=16 union=0)
```

```
Conditional marginal effects Number of obs = 19224
Model VCE : OIM
Expression : Pr(ln_wage1<1.30), predict(pr0(.,1.30))
dy/dx w.r.t. : 1.union
1._at : union = 0
   age = 30
   grade = 12
   south = 1
   year = 88
2._at : union = 0
   age = 30
   grade = 16
   south = 1
   year = 88

|                | dy/dx  | Std. Err. | z     | P>|z|   | [95% Conf. Interval] |
|----------------|--------|-----------|-------|-------|----------------------|
| 1.union _at    |        |           |       |       |                      |
| 1              | -.0787117 | .0060655  | -12.98 | 0.000 | -.0905999 -.0668235 |
| 2              | -.0378758 | .0035595  | -10.64 | 0.000 | -.0448523 -.0308993 |
```

Note: dy/dx for factor levels is the discrete change from the base level.

For the first group of women, according to our fitted model, being in a union lowers the probability of being classified as a low-wage worker by almost 7.9 percentage points. Being a college graduate attenuates this effect to just under 3.8 percentage points.
Also see

[XT] xtintreg — Random-effects interval-data regression models

[U] 20 Estimation and postestimation commands
xtivreg — Instrumental variables and two-stage least squares for panel-data models

Syntax

GLS random-effects (RE) model

xtivreg depvar [ varlist1 ] ( varlist2 = varlistiv ) [ if ] [ in ] [ , re RE_options ]

Between-effects (BE) model

xtivreg depvar [ varlist1 ] ( varlist2 = varlistiv ) [ if ] [ in ] , be [ BE_options ]

Fixed-effects (FE) model

xtivreg depvar [ varlist1 ] ( varlist2 = varlistiv ) [ if ] [ in ] , fe [ FE_options ]

First-differenced (FD) estimator

xtivreg depvar [ varlist1 ] ( varlist2 = varlistiv ) [ if ] [ in ] , fd [ FD_options ]

RE_options Description

Model
re use random-effects estimator; the default
ec2sls use Baltagi’s EC2SLS random-effects estimator
nosa use the Baltagi–Chang estimators of the variance components
regress treat covariates as exogenous and ignore instrumental variables

SE
vce(vcetype) vcetype may be conventional, bootstrap, or jackknife

Reporting
level(#) set confidence level; default is level(95)
first report first-stage estimates
small report t and F statistics instead of Z and χ² statistics
theta report θ
display_options control column formats, row spacing, line width, display of omitted variables and base and empty cells, and factor-variable labeling
coeflegend display legend instead of statistics

197
**BE** Options | Description
--- | ---
**Model** |  
*be* | use between-effects estimator  
*regress* | treat covariates as exogenous and ignore instrumental variables
**SE** |  
*vce(vcetype)* | *vcetype* may be *conventional, bootstrap, or jackknife*
**Reporting** |  
*level(#)* | set confidence level; default is *level(95)*  
*first* | report first-stage estimates  
*small* | report *t* and *F* statistics instead of *Z* and *χ²* statistics
*display_options* | control column formats, row spacing, line width, display of omitted variables and base and empty cells, and factor-variable labeling
*coeflegend* | display legend instead of statistics

**FE** Options | Description
--- | ---
**Model** |  
*fe* | use fixed-effects estimator  
*regress* | treat covariates as exogenous and ignore instrumental variables
**SE** |  
*vce(vcetype)* | *vcetype* may be *conventional, bootstrap, or jackknife*
**Reporting** |  
*level(#)* | set confidence level; default is *level(95)*  
*first* | report first-stage estimates  
*small* | report *t* and *F* statistics instead of *Z* and *χ²* statistics
*display_options* | control column formats, row spacing, line width, display of omitted variables and base and empty cells, and factor-variable labeling
*coeflegend* | display legend instead of statistics
xtivreg — Instrumental variables and two-stage least squares for panel-data models

### Description

*xtivreg* offers five different estimators for fitting panel-data models in which some of the right-hand-side covariates are endogenous. These estimators are two-stage least-squares generalizations of simple panel-data estimators for exogenous variables. *xtivreg* with the **be** option uses the two-stage least-squares between estimator. *xtivreg* with the **fe** option uses the two-stage least-squares within estimator. *xtivreg* with the **re** option uses a two-stage least-squares random-effects estimator. There are two implementations: G2SLS from Balestra and Varadharajan-Krishnakumar (1987) and EC2SLS from Baltagi. The Balestra and Varadharajan-Krishnakumar G2SLS is the default because it is computationally less expensive. Baltagi’s EC2SLS can be obtained by specifying the **ec2sls** option. *xtivreg* with the **fd** option requests the two-stage least-squares first-differenced estimator.

See Baltagi (2013) for an introduction to panel-data models with endogenous covariates. For the derivation and application of the first-differenced estimator, see Anderson and Hsiao (1981).

### FD_options

**Model**

- **noconstant** suppress constant term
- **fd** use first-differenced estimator
- **regress** treat covariates as exogenous and ignore instrumental variables

**SE**

- **vce(vcetype)** *vcetype* may be *conventional*, *bootstrap*, or *jackknife*

**Reporting**

- **level(#)** set confidence level; default is **level(95)**
- **first** report first-stage estimates
- **small** report *t* and *F* statistics instead of *Z* and *χ²* statistics
- **display_options** control column formats, row spacing, line width, and display of omitted variables

- **coeflegend** display legend instead of statistics

A panel variable must be specified. For *xtivreg*, **fd** a time variable must also be specified. Use *xtset*; see [XT] *xtset*.

`varlist₁` and `varlist₁v` may contain factor variables, except for the **fd** estimator; see [U] 11.4.3 Factor variables.

`depvar`, `varlist₁`, `varlist₂`, and `varlist₁v` may contain time-series operators; see [U] 11.4.4 Time-series varlists.

by and `statsby` are allowed; see [U] 11.1.10 Prefix commands.

`coeflegend` does not appear in the dialog box.

See [U] 20 Estimation and postestimation commands for more capabilities of estimation commands.

### Menu

Statistics > Longitudinal/panel data > Endogenous covariates > Instrumental-variables regression (FE, RE, BE, FD)
Options for RE model

re requests the G2SLS random-effects estimator. re is the default.

ec2sle requests Baltagi’s EC2SLS random-effects estimator instead of the default Balestra and Varadharajan-Krishnakumar estimator.

nosa specifies that the Baltagi–Chang estimators of the variance components be used instead of the default adapted Swamy–Arora estimators.

regress specifies that all the covariates be treated as exogenous and that the instrument list be ignored. Specifying regress causes xtitreg to fit the requested panel-data regression model of depvar on varlist₁ and varlist₂, ignoring varlistiv.

vce(vcetype) specifies the type of standard error reported, which includes types that are derived from asymptotic theory (conventional) and that use bootstrap or jackknife methods (bootstrap, jackknife); see [XT] vce_options.

vce(conventional), the default, uses the conventionally derived variance estimator for generalized least-squares regression.

level(#); see [R] estimation options.

first specifies that the first-stage regressions be displayed.

small specifies that t statistics be reported instead of Z statistics and that F statistics be reported instead of χ² statistics.

theta specifies that the output include the estimated value of θ used in combining the between and fixed estimators. For balanced data, this is a constant, and for unbalanced data, a summary of the values is presented in the header of the output.

display_options: noomitted, vsquish, noemptycells, baselevels, allbaselevels, nolabel, fvwrap(#), fvwrapon(style), cformat(%fmt), pformat(%fmt), sformat(%fmt), and nolstretch; see [R] estimation options.

The following option is available with xtitreg but is not shown in the dialog box:

coeflegend; see [R] estimation options.

Options for BE model

be requests the between regression estimator.

regress specifies that all the covariates are to be treated as exogenous and that the instrument list is to be ignored. Specifying regress causes xtitreg to fit the requested panel-data regression model of depvar on varlist₁ and varlist₂, ignoring varlistiv.
vce(vcetype) specifies the type of standard error reported, which includes types that are derived from asymptotic theory (conventional) and that use bootstrap or jackknife methods (bootstrap, jackknife); see [XT] vce_options.

vce(conventional), the default, uses the conventionally derived variance estimator for generalized least-squares regression.

Reporting

level(#) ; see [R] estimation options.

first specifies that the first-stage regressions be displayed.

small specifies that \( t \) statistics be reported instead of \( Z \) statistics and that \( F \) statistics be reported instead of \( \chi^2 \) statistics.

display_options: nomomitted, vsquish, noemptycells, baselevels, allbaselevels, nolabel, fvwrap(#), fvwrapon(style), cformat(%,fmt), pformat(%,fmt), sformat(%,fmt), and nolstretch; see [R] estimation options.

The following option is available with xtivreg but is not shown in the dialog box:

coefflegend; see [R] estimation options.

Options for FE model

fe requests the fixed-effects (within) regression estimator.

regress specifies that all the covariates are to be treated as exogenous and that the instrument list is to be ignored. Specifying regress causes xtivreg to fit the requested panel-data regression model of \( depvar \) on \( varlist_1 \) and \( varlist_2 \), ignoring \( varlist_4 \).

vce(vcetype) specifies the type of standard error reported, which includes types that are derived from asymptotic theory (conventional) and that use bootstrap or jackknife methods (bootstrap, jackknife); see [XT] vce_options.

vce(conventional), the default, uses the conventionally derived variance estimator for generalized least-squares regression.

Reporting

level(#) ; see [R] estimation options.

first specifies that the first-stage regressions be displayed.

small specifies that \( t \) statistics be reported instead of \( Z \) statistics and that \( F \) statistics be reported instead of \( \chi^2 \) statistics.

display_options: nomomitted, vsquish, noemptycells, baselevels, allbaselevels, nolabel, fvwrap(#), fvwrapon(style), cformat(%,fmt), pformat(%,fmt), sformat(%,fmt), and nolstretch; see [R] estimation options.
The following option is available with `xtivreg` but is not shown in the dialog box: `coeflegend`; see [R] estimation options.

**Options for FD model**

- **Model**
  - `noconstant`; see [R] estimation options.
  - `fd` requests the first-differenced regression estimator.
  - `regress` specifies that all the covariates are to be treated as exogenous and that the instrument list is to be ignored. Specifying `regress` causes `xtivreg` to fit the requested panel-data regression model of `depvar` on `varlist1` and `varlist2`, ignoring `varlistiv`.

- **SE**
  - `vce(vcetype)` specifies the type of standard error reported, which includes types that are derived from asymptotic theory (conventional) and that use bootstrap or jackknife methods (bootstrap, jackknife); see [XT] `vce_options`.
  - `vce(conventional)`, the default, uses the conventionally derived variance estimator for generalized least-squares regression.

- **Reporting**
  - `level(#);` see [R] estimation options.
  - `first` specifies that the first-stage regressions be displayed.
  - `small` specifies that $t$ statistics be reported instead of $Z$ statistics and that $F$ statistics be reported instead of $\chi^2$ statistics.

  `display_options: noomitted, vsquish, cformat(\%fmt), pformat(\%fmt), sformat(\%fmt), and nolstretch;` see [R] estimation options.

The following option is available with `xtivreg` but is not shown in the dialog box: `coeflegend`; see [R] estimation options.

**Remarks and examples**

If you have not read [XT] `xt`, please do so.

Consider an equation of the form

$$ y_{it} = Y_{it}\gamma + X_{1it}\beta + \mu_i + \nu_{it} = Z_{it}\delta + \mu_i + \nu_{it} \quad (1) $$

where

- $y_{it}$ is the dependent variable;
- $Y_{it}$ is an $1 \times g_2$ vector of observations on $g_2$ endogenous variables included as covariates, and these variables are allowed to be correlated with the $\nu_{it}$;
- $X_{1it}$ is an $1 \times k_1$ vector of observations on the exogenous variables included as covariates;
xtivreg — Instrumental variables and two-stage least squares for panel-data models 203

\[ Z_{it} = [Y_{it} X_{it}]; \]

\(\gamma\) is a \(g_2 \times 1\) vector of coefficients;

\(\beta\) is a \(k_1 \times 1\) vector of coefficients; and

\(\delta\) is a \(K \times 1\) vector of coefficients, where \(K = g_2 + k_1\).

Assume that there is a \(1 \times k_2\) vector of observations on the \(k_2\) instruments in \(X_{2it}\). The order condition is satisfied if \(k_2 \geq g_2\). Let \(X_{it} = [X_{1it} X_{2it}]\). xtivreg handles exogenously unbalanced panel data. Thus define \(T_i\) to be the number of observations on panel \(i\), \(n\) to be the number of panels and \(N\) to be the total number of observations; that is, \(N = \sum_{i=1}^{n} T_i\).

xtivreg offers five different estimators, which may be applied to models having the form of (1). The first-differenced estimator (FD2SLS) removes the \(\mu_i\) by fitting the model in first differences. The within estimator (FE2SLS) fits the model after sweeping out the \(\mu_i\) by removing the panel-level means from each variable. The between estimator (BE2SLS) models the panel averages. The two random-effects estimators, G2SLS and EC2SLS, treat the \(\mu_i\) as random variables that are independent and identically distributed (i.i.d.) over the panels. Except for (FD2SLS), all of these estimators are generalizations of estimators in xtreg. See [XT] xtreg for a discussion of these estimators for exogenous covariates.

Although the estimators allow for different assumptions about the \(\mu_i\), all the estimators assume that the idiosyncratic error term \(\nu_{it}\) has zero mean and is uncorrelated with the variables in \(X_{it}\). Just as when there are no endogenous covariates, as discussed in [XT] xtreg, there are various perspectives on what assumptions should be placed on the \(\mu_i\). If they are assumed to be fixed, the \(\mu_i\) may be correlated with the variables in \(X_{it}\), and the within estimator is efficient within a class of limited information estimators. Alternatively, if the \(\mu_i\) are assumed to be random, they are also assumed to be i.i.d. over the panels. If the \(\mu_i\) are assumed to be uncorrelated with the variables in \(X_{it}\), the GLS random-effects estimators are more efficient than the within estimator. However, if the \(\mu_i\) are correlated with the variables in \(X_{it}\), the random-effects estimators are inconsistent but the within estimator is consistent. The price of using the within estimator is that it is not possible to estimate coefficients on time-invariant variables, and all inference is conditional on the \(\mu_i\) in the sample. See Mundlak (1978) and Hsiao (2003) for discussions of this interpretation of the within estimator.

Example 1: Fixed-effects model

For the within estimator, consider another version of the wage equation discussed in [XT] xtreg. The data for this example come from an extract of women from the National Longitudinal Survey of Youth that was described in detail in [XT] xt. Restricting ourselves to only time-varying covariates, we might suppose that the log of the real wage was a function of the individual’s age, age\(^2\), her tenure in the observed place of employment, whether she belonged to union, whether she lives in metropolitan area, and whether she lives in the south. The variables for these are, respectively, age, c.age#c.age, tenure, union, not_smsa, and south. If we treat all the variables as exogenous, we can use the one-stage within estimator from xtreg, yielding
. use http://www.stata-press.com/data/r13/nlswork
(National Longitudinal Survey. Young Women 14-26 years of age in 1968)
. xtreg ln_w age c.age#c.age tenure not_smsa union south, fe

Fixed-effects (within) regression                       Number of obs   =    19007
Group variable: idcode                                  Number of groups =     4134
R-sq: within   = 0.1333                                Obs per group: min =     1
                     between = 0.2375                   avg =     4.6
                     overall = 0.2031                   max =     12

F(6, 14867) = 381.19  corr(u_i, Xb) = 0.2074
                  Prob > F = 0.0000                         

                      | Coef.     Std. Err.      t    P>|t|     [95% Conf. Interval]
----------------------|-----------|------------------|------|---------|---------------------------|
ln_wage               |           |                   |      |         |                           |
     age               |  .0311984 |   .0033902        |  9.20  |  0.000  |    .0245533 - .0378436    |
     c.age#c.age       |  -.0003457|   .0000543        | -6.37  |  0.000  |   -.0004522 - -.0002393  |
     tenure             |  .0176205 |   .0008099        |  21.76 |  0.000  |    .0160331  .0192079    |
     not_smsa           |  -.0972535|   .0125377        | -7.76  |  0.000  |   -.1218289 - -.072678   |
     union              |  .0975672 |   .0069844        |  13.97 |  0.000  |    .0838769  .1112576    |
     south              |  -.0620932|   .0133272        | -4.66  |  0.000  |   -.0882158 - -.0359706  |
     _cons              |  1.091612 |   .0523126        |  20.87 |  0.000  |    .9890729  1.194151    |

sigma_u: .3910683
sigma_e: .25545969
rho: .70091004 (fraction of variance due to u_i)

F test that all u_i=0: F(4133, 14867) = 8.31 Prob > F = 0.0000

All the coefficients are statistically significant and have the expected signs.

Now suppose that we wish to model tenure as a function of union and south and that we believe that
the errors in the two equations are correlated. Because we are still interested in the within estimates,
we now need a two-stage least-squares estimator. The following output shows the command and the
results from fitting this model:
Although all the coefficients still have the expected signs, the coefficients on age and not_smsa are no longer statistically significant. Given that these variables have been found to be important in many other studies, we might want to rethink our specification.

If we are willing to assume that the \( \mu_i \) are uncorrelated with the other covariates, we can fit a random-effects model. The model is frequently known as the variance-components or error-components model. \texttt{xtivreg} has estimators for two-stage least-squares one-way error-components models. In the one-way framework, there are two variance components to estimate, the variance of the \( \mu_i \) and the variance of the \( \nu_{it} \). Because the variance components are unknown, consistent estimates are required to implement feasible GLS. \texttt{xtivreg} offers two choices: a Swamy–Arora method and simple consistent estimators from Baltagi and Chang (2000).

Baltagi and Chang (1994) derived the Swamy–Arora estimators of the variance components for unbalanced panels. By default, \texttt{xtivreg} uses estimators that extend these unbalanced Swamy–Arora estimators to the case with instrumental variables. The default Swamy–Arora method contains a degree-of-freedom correction to improve its performance in small samples. Baltagi and Chang (2000) use variance-components estimators, which are based on the ideas of Amemiya (1971) and Swamy and Arora (1972), but they do not attempt to make small-sample adjustments. These consistent estimators of the variance components will be used if the nosa option is specified.

Using either estimator of the variance components, \texttt{xtivreg} offers two GLS estimators of the random-effects model. These two estimators differ only in how they construct the GLS instruments from the exogenous and instrumental variables contained in \( X_{it} = [X_{1it} \ X_{2it}] \). The default method, G2SLS, which is from Balestra and Varadharajan-Krishnakumar, uses the exogenous variables after they have been passed through the feasible GLS transform. In math, G2SLS uses \( X_{it}^* \) for the GLS instruments, where \( X_{it}^* \) is constructed by passing each variable in \( X_{it} \) through the GLS transform in (3) given in Methods and formulas. If the ec2sls option is specified, \texttt{xtivreg} performs Baltagi's
EC2SLS. In EC2SLS, the instruments are \( \tilde{X}_{it} \) and \( X_{it} \), where \( \tilde{X}_{it} \) is constructed by passing each of the variables in \( X_{it} \) through the within transform, and \( X_{it} \) is constructed by passing each variable through the between transform. The within and between transforms are given in the *Methods and formulas* section. Baltagi and Li (1992) show that, although the G2SLS instruments are a subset of those contained in EC2SLS, the extra instruments in EC2SLS are redundant in the sense of White (2001). Given the extra computational cost, G2SLS is the default.

Example 2: GLS random-effects model

Here is the output from applying the G2SLS estimator to this model:

```
xтивreg ln_w age c.age#c.age not_smsa 2.race (tenure = union birth south), re
G2SLS random-effects IV regression Number of obs = 19007
Group variable: idcode Number of groups = 4134
R-sq: within = 0.0664 Obs per group: min = 1
between = 0.2098 avg = 4.6
overall = 0.1463 max = 12
Wald chi2(5) = 1446.37
corr(u_i, X) = 0 (assumed) Prob > chi2 = 0.0000
```

| ln_wage  | Coef.   | Std. Err. | z     | P>|z|   | [95% Conf. Interval] |
|----------|---------|-----------|-------|-------|---------------------|
| tenure   | .1391798| .0078756  | 17.67 | 0.000 | .123744 .1546157    |
| age      | .0279649| .0054182  | 5.16  | 0.000 | .0173454 .0385843   |
| c.age#c.age | -.0008357| .0000871 | -9.60 | 0.000 | -.0010063 -.000665  |
| not_smsa | -.235103 | .011371 | -20.07 | 0.000 | -.2453386 -.2016821 |
| race     | -.2078613| .0125803 | -16.52 | 0.000 | -.2325183 -.1832044 |
| black    | 1.337684 | .0844988 | 15.83 | 0.000 | 1.172069 1.503299   |
| _cons    | .36582493|          |       |       |                     |
| sigma_u  | .63031479|          |       |       |                     |
| rho      | .25197078| (fraction of variance due to u_i) |

We have included two time-invariant covariates, *birth_yr* and *2.race*. All the coefficients are statistically significant and are of the expected sign.
Applying the EC2SLS estimator yields similar results:

```stata
.xtivreg ln_w age c.age#c.age not_smsa 2.race (tenure = union birth south), re
> ec2sls

EC2SLS random-effects IV regression
Number of obs = 19007
Group variable: idcode
Number of groups = 4134
R-sq: within = 0.0898
between = 0.2608
overall = 0.1926
Wald chi2(5) = 2721.92
corr(u_i, X) = 0 (assumed)

| Coef. Std. Err. | z    | P>|z|  | [95% Conf. Interval] |
|-----------------|------|------|----------------------|
| ln_wage tenure  | .064822 | .0025647 | 25.27 | 0.000 | .0597953 | .0698486 |
| ln_wage age     | .0380048 | .0039549 | 9.61  | 0.000 | .0302534 | .0457562 |
| ln_wage c.age#c.age | -.0006676 | .0000632 | -10.56 | 0.000 | -.0007915 | -.0005438 |
| ln_wage not_smsa| -.2298961 | .0082993 | -27.70 | 0.000 | -.2461625 | -.2136297 |
| ln_wage race black | -.1823627 | .0092005 | -19.82 | 0.000 | -.2003954 | -.16433 |
| ln_wage _cons   | 1.110564 | .0606538 | 18.31  | 0.000 | .9916849 | 1.229443 |

sigma_u .36582493
sigma_e .63031479
rho .25197078 (fraction of variance due to u_i)

Instrumented: tenure
Instruments: age c.age#c.age not_smsa 2.race union birth_yr south
```

Fitting the same model as above with the G2SLS estimator and the consistent variance components estimators yields
Example 3: First-differenced estimator

The two-stage least-squares first-differenced estimator (FD2SLS) has been used to fit both fixed-effect and random-effect models. If the $\mu_i$ are truly fixed-effects, the FD2SLS estimator is not as efficient as the two-stage least-squares within estimator for finite $T_i$. Similarly, if none of the endogenous variables are lagged dependent variables, the exogenous variables are all strictly exogenous, and the random effects are i.i.d. and independent of the $X_{it}$, the two-stage GLS estimators are more efficient than the FD2SLS estimator. However, the FD2SLS estimator has been used to obtain consistent estimates when one of these conditions fails. Anderson and Hsiao (1981) used a version of the FD2SLS estimator to fit a panel-data model with a lagged dependent variable. Arellano and Bond (1991) develop new one-step and two-step GMM estimators for dynamic panel data. See [XT] xtabond for a discussion of these estimators and Stata’s implementation of them. In their article, Arellano and Bond (1991) apply their new estimators to a model of dynamic labor demand that had previously been considered by Layard and Nickell (1986). They also compare the results of their estimators with those from the Anderson–Hsiao estimator using data from an unbalanced panel of firms from the United Kingdom. As is conventional, all variables are indexed over the firm $i$ and time $t$. In this dataset, $n_{it}$ is the log of employment in firm $i$ inside the United Kingdom at time $t$, $w_{it}$ is the natural log of the real product wage, $k_{it}$ is the natural log of the gross capital stock, and $y_{st}$ is the natural log of industry output. The model also includes time dummies $yr_{1980}$, $yr_{1981}$, $yr_{1982}$, $yr_{1983}$, and $yr_{1984}$. In Arellano and Bond (1991, table 5, column e), the authors present the results from applying one version of the Anderson–Hsiao estimator to these data. This example reproduces their results for the coefficients, though standard errors are different because Arellano and Bond are using robust standard errors.

### Example

#### First-differenced estimator

The two-stage least-squares first-differenced estimator (FD2SLS) has been used to fit both fixed-effect and random-effect models. If the $\mu_i$ are truly fixed-effects, the FD2SLS estimator is not as efficient as the two-stage least-squares within estimator for finite $T_i$. Similarly, if none of the endogenous variables are lagged dependent variables, the exogenous variables are all strictly exogenous, and the random effects are i.i.d. and independent of the $X_{it}$, the two-stage GLS estimators are more efficient than the FD2SLS estimator. However, the FD2SLS estimator has been used to obtain consistent estimates when one of these conditions fails. Anderson and Hsiao (1981) used a version of the FD2SLS estimator to fit a panel-data model with a lagged dependent variable.
. use http://www.stata-press.com/data/r13/abdata
. xtivreg n 12.n l(0/1).w l(0/2).ys yr1981-yr1984 (l.n = 13.n), fd

First-differenced IV regression
Group variable:  id  Number of obs =  471
Time variable:  year  Number of groups = 140
R-sq:  within = 0.0141  Obs per group:  min =  3
        between = 0.9165  avg =  3.4
        overall = 0.9892  max =  5
Wald chi2(14) = 122.53  corr(u_i, Xb) = 0.9239
Prob > chi2 = 0.0000

| Coef.    | Std. Err. | z     | P>|z|  | [95% Conf. Interval] |
|----------|-----------|-------|------|----------------------|
| n        | 1.422765  | 1.583053 | 0.90 | 0.369 | -1.679962 | 4.525493 |
| L2D.     | -0.1645517 | 1.647179 | -1.00 | 0.318 | -0.4873928 | 0.1582894 |
| w        | -0.7524675 | 0.1765733 | -4.26 | 0.000 | -1.098545 | -0.4063902 |
| D1.      | 0.9627611  | 1.086506 | 0.89 | 0.376 | -0.166752 | 3.092275 |
| LD.      | -0.3221686 | 0.1466086 | 2.20 | 0.028 | 0.0348211 | 0.6095161 |
| k        | 0.3221686  | 0.1466086 | 2.20 | 0.028 | 0.0348211 | 0.6095161 |
| D1.      | -0.3248778 | 0.1580599 | -2.07 | 0.038 | -0.6475037 | -0.002642 |
| LD.      | -0.953947 | 0.1960883 | -4.91 | 0.027 | -1.358757 | -0.549137 |
| L2D.     | 0.3212993  | 0.1580599 | 2.07 | 0.038 | -0.0432461 | 0.6868437 |
| yrs      | 1.361881   | 1.156835 | -1.18 | 0.239 | -3.629237 | 0.9054744 |
| D1.      | 0.3212993  | 0.1580599 | 2.07 | 0.038 | -0.0432461 | 0.6868437 |
| yr1981   | -0.0574197 | 0.0430158 | -1.33 | 0.182 | -0.1417291 | 0.0268896 |
| D1.      | -0.0882952 | 0.0706214 | -1.25 | 0.211 | -0.2267106 | 0.0501203 |
| yr1982   | -0.1063153 | 0.10861 | -0.99 | 0.328 | -0.319187 | 0.1065563 |
| D1.      | -0.1172108 | 0.15196 | -0.77 | 0.441 | -0.4150468 | 0.1806253 |
| _cons    | 0.0161204  | 0.0336264 | 0.48 | 0.632 | -0.0497861 | 0.082027 |
| sigma_u  | .29069213  |       |       |       |       |       |
| sigma_e  | .18855982  |       |       |       |       |       |
| rho      | .70384993  |       |       |       |       |       |

Instrumented:  L.n
xtivreg, re stores the following in e():

Scalars:
- e(N): number of observations
- e(N_g): number of groups
- e(df_m): model degrees of freedom
- e(df_rz): residual degrees of freedom
- e(g_min): smallest group size
- e(g_avg): average group size
- e(g_max): largest group size
- e(Tcon): 1 if panels balanced; 0 otherwise
- e(sigma): ancillary parameter (gamma, lnormal)
- e(sigma_u): panel-level standard deviation
- e(sigma_e): standard deviation of \( \epsilon_{it} \)
- e(r2_w): \( R^2 \) for within model
- e(r2_o): \( R^2 \) for overall model
- e(r2_b): \( R^2 \) for between model
- e(chi2): \( \chi^2 \)
- e(rho): \( \rho \)
- e(F): model \( F \) (small only)
- e(m_p): \( p \)-value from model test
- e(theta_min): minimum \( \theta \)
- e(theta_5): \( \theta \), 5th percentile
- e(theta_50): \( \theta \), 50th percentile
- e(theta_95): \( \theta \), 95th percentile
- e(theta_max): maximum \( \theta \)
- e(Tbar): harmonic mean of group sizes
- e(rank): rank of e(V)

Macros:
- e(cmd): xtivreg
- e(cmdline): command as typed
- e(depvar): name of dependent variable
- e(ivar): variable denoting groups
- e(tvar): variable denoting time within groups
- e(insts): instruments
- e(instd): instrumented variables
- e(model): g2sls or ec2sls
- e(small): small, if specified
- e(chi2type): Wald; type of model \( \chi^2 \) test
- e(vce): vcetype specified in vce()
- e(vcetype): title used to label Std. Err.
- e(properties): b V
- e(predict): program used to implement predict
- e(marginsok): predictions allowed by margins
- e(marginsnotok): predictions disallowed by margins
- e(asbalanced): factor variables fvset as asbalanced
- e(asobserved): factor variables fvset as asobserved
xtivreg — Instrumental variables and two-stage least squares for panel-data models

Matrices
- e(b)  coefficient vector
- e(V)  variance–covariance matrix of the estimators

Functions
- e(sample)  marks estimation sample

xtivreg, be stores the following in e():

Scalars
- e(N)  number of observations
- e(N_g)  number of groups
- e(mss)  model sum of squares
- e(df_m)  model degrees of freedom
- e(rss)  residual sum of squares
- e(df_r)  residual degrees of freedom
- e(df_rz)  residual degrees of freedom for the between-transformed regression
- e(g_min)  smallest group size
- e(g_avg)  average group size
- e(g_max)  largest group size
- e(rs_a)  adjusted $R^2$
- e(r2_w)  $R^2$ for within model
- e(r2_o)  $R^2$ for overall model
- e(r2_b)  $R^2$ for between model
- e(chi2)  model Wald
- e(chi2_p)  $p$-value for model $\chi^2$ test
- e(F)  $F$ statistic (small only)
- e(rmse)  root mean squared error
- e(rank)  rank of e(V)

Macros
- e(cmd)  xtivreg
- e(cmdline)  command as typed
- e(depvar)  name of dependent variable
- e(ivar)  variable denoting groups
- e(tvar)  variable denoting time within groups
- e(insts)  instruments
- e(instd)  instrumented variables
- e(model)  be
- e(small)  small, if specified
- e(vce)  vcetype specified in vce()
- e(vcetype)  title used to label Std. Err.
- e(properties)  b V
- e(predict)  program used to implement predict
- e(marginsok)  predictions allowed by margins
- e(marginsnotok)  predictions disallowed by margins
- e(asbalanced)  factor variables fvset as asbalanced
- e(asobserved)  factor variables fvset as asobserved

Matrices
- e(b)  coefficient vector
- e(V)  variance–covariance matrix of the estimators

Functions
- e(sample)  marks estimation sample
xtivreg, fe stores the following in e():

Scalars
- e(N): number of observations
- e(N_g): number of groups
- e(df_m): model degrees of freedom
- e(rss): residual sum of squares
- e(df_r): residual degrees of freedom (small only)
- e(df_rz): residual degrees of freedom for the within-transformed regression
- e(g_min): smallest group size
- e(g_avg): average group size
- e(g_max): largest group size
- e(sigma): ancillary parameter (gamma, lnormal)
- e(corr): corr(u_i, Xb)
- e(sigma_u): panel-level standard deviation
- e(sigma_e): standard deviation of $\epsilon_{it}$
- e(r2_w): $R^2$ for within model
- e(r2_o): $R^2$ for overall model
- e(r2_b): $R^2$ for between model
- e(chi2): model Wald (not small)
- e(df_b): degrees of freedom for $\chi^2$ statistic
- e(chi2_p): $p$-value for model $\chi^2$ statistic
- e(rho): $\rho$
- e(F): $F$ statistic (small only)
- e(F_f): $F$ for $H_0: u_i=0$
- e(F_fp): $p$-value for $F$ for $H_0: u_i=0$
- e(df_a): degrees of freedom for absorbed effect
- e(rank): rank of e(V)

Macros
- e(cmd): xtivreg
- e(cmdline): command as typed
- e(depvar): name of dependent variable
- e(ivar): variable denoting groups
- e(tvar): variable denoting time within groups
- e(insts): instruments
- e(instd): instrumented variables
- e(model): fe
- e(small): small, if specified
- e(vce): vcetype specified in vce()
- e(vcetype): title used to label Std. Err.
- e(properties): b V
- e(predict): program used to implement predict
- e(marginsok): predictions allowed by margins
- e(marginsnotok): predictions disallowed by margins
- e(asbalanced): factor variables fvset as asbalanced
- e(asobserved): factor variables fvset as asobserved

Matrices
- e(b): coefficient vector
- e(V): variance–covariance matrix of the estimators

Functions
- e(sample): marks estimation sample
xtivreg, fd stores the following in e():

Scalars

- _e(N)_ number of observations
- _e(N_g)_ number of groups
- _e(rss)_ residual sum of squares
- _e(df_r)_ residual degrees of freedom (small only)
- _e(df_rz)_ residual degrees of freedom for first-differenced regression
- _e(g_min)_ smallest group size
- _e(g_avg)_ average group size
- _e(g_max)_ largest group size
- _e(sigma)_ ancillary parameter (gamma, lnormal)
- _e(corr)_ cor(u_i, Xb)
- _e(sigma_u)_ panel-level standard deviation
- _e(sigma_e)_ standard deviation of \( \epsilon_{it} \)
- _e(r2_w)_ \( R^2 \)-squared for within model
- _e(r2_o)_ \( R^2 \)-squared for overall model
- _e(r2_b)_ \( R^2 \)-squared for between model
- _e(chi2)_ model Wald (not small)
- _e(chi2_p)_ \( p \)-value for model \( \chi^2 \) statistic
- _e(rho)_ \( \rho \)
- _e(F)_ \( F \) statistic (small only)
- _e(rank)_ rank of _e(V)_

Macros

- _e(cmd)_ xtivreg
- _e(cmdline)_ command as typed
- _e(depvar)_ name of dependent variable
- _e(ivar)_ variable denoting groups
- _e(tvar)_ variable denoting time within groups
- _e(insts)_ instruments
- _e(instd)_ instrumented variables
- _e(model)_ fd
- _e(small)_ small, if specified
- _e(vce)_ vcetype specified in vce()
- _e(vcetype)_ title used to label Std. Err.
- _e(properties)_ b V
- _e(predict)_ program used to implement predict
- _e(marginsok)_ predictions allowed by margins

Matrices

- _e(b)_ coefficient vector
- _e(V)_ variance–covariance matrix of the estimators

Functions

- _e(sample)_ marks estimation sample
Methods and formulas

Consider an equation of the form

\[ y_{it} = Y_{it} \gamma + X_{1it} \beta + \mu_i + \nu_{it} = Z_{it} \delta + \mu_i + \nu_{it} \]  

(2)

where

- \( y_{it} \) is the dependent variable;
- \( Y_{it} \) is an \( 1 \times g_2 \) vector of observations on \( g_2 \) endogenous variables included as covariates, and these variables are allowed to be correlated with the \( \nu_{it} \);
- \( X_{1it} \) is an \( 1 \times k_1 \) vector of observations on the exogenous variables included as covariates;
- \( Z_{it} = [Y_{it} \ X_{it}] \);
- \( \gamma \) is a \( g_2 \times 1 \) vector of coefficients;
- \( \beta \) is a \( k_1 \times 1 \) vector of coefficients; and
- \( \delta \) is a \( K \times 1 \) vector of coefficients, where \( K = g_2 + k_1 \).

Assume that there is a \( 1 \times k_2 \) vector of observations on the \( k_2 \) instruments in \( X_{2it} \). The order condition is satisfied if \( k_2 \geq g_2 \). Let \( X_{it} = [X_{1it} \ X_{2it}] \). \texttt{xtivreg} handles exogenously unbalanced panel data. Thus define \( T_i \) to be the number of observations on panel \( i \), \( n \) to be the number of panels, and \( N \) to be the total number of observations; that is, \( N = \sum_{i=1}^{n} T_i \).

Methods and formulas are presented under the following headings:

\texttt{xtivreg, fd}
\texttt{xtivreg, fe}
\texttt{xtivreg, be}
\texttt{xtivreg, re}

\texttt{xtivreg, fd}

As the name implies, this estimator obtains its estimates and conventional VCE from an instrumental-variables regression on the first-differenced data. Specifically, first differencing the data yields

\[ y_{it} - y_{it-1} = (Z_{it} - Z_{i,t-1}) \delta + \nu_{it} - \nu_{i,t-1} \]

With the \( \mu_i \) removed by differencing, we can obtain the estimated coefficients and their estimated variance–covariance matrix from a standard two-stage least-squares regression of \( \Delta y_{it} \) on \( \Delta Z_{it} \) with instruments \( \Delta X_{it} \).

\[ R^2 \, \text{within is reported as} \quad \left[ \text{corr}\left\{ (Z_{it} - \overline{Z}_i) \hat{\delta}, y_{it} - \overline{y}_i \right\} \right]^2. \]

\[ R^2 \, \text{between is reported as} \quad \left\{ \text{corr}(Z_i \hat{\delta}, \overline{y}_i) \right\}^2. \]

\[ R^2 \, \text{overall is reported as} \quad \left\{ \text{corr}(Z_{it} \hat{\delta}, y_{it}) \right\}^2. \]

\texttt{xtivreg, fe}

At the heart of this model is the within transformation. The within transform of a variable \( w \) is

\[ \tilde{w}_{it} = w_{it} - \overline{w}_i + \overline{w} \]
where
\[
\bar{w}_i = \frac{1}{n} \sum_{t=1}^{T_i} w_{it}
\]
\[
\bar{w} = \frac{1}{N} \sum_{i=1}^{n} \sum_{t=1}^{T_i} w_{it}
\]
and \(n\) is the number of groups and \(N\) is the total number of observations on the variable.

The within transform of (2) is
\[
\tilde{y}_{it} = \tilde{Z}_{it} + \tilde{\nu}_{it}
\]
The within transform has removed the \(\mu_i\). With the \(\mu_i\) gone, the within 2SLS estimator can be obtained from a two-stage least-squares regression of \(\tilde{y}_{it}\) on \(\tilde{Z}_{it}\) with instruments \(\tilde{X}_{it}\).

Suppose that there are \(K\) variables in \(Z_{it}\), including the mandatory constant. There are \(K + n - 1\) parameters estimated in the model, and the conventional VCE for the within estimator is
\[
\frac{N - K}{N - n - K + 1} V_{IV}
\]
where \(V_{IV}\) is the VCE from the above two-stage least-squares regression.

From the estimate of \(\hat{\delta}\), estimates \(\hat{\mu}_i\) of \(\mu_i\) are obtained as \(\hat{\mu}_i = \bar{y}_i - \bar{Z}_i \hat{\delta}\). Reported from the calculated \(\hat{\mu}_i\) is its standard deviation and its correlation with \(\bar{Z}_i \hat{\delta}\). Reported as the standard deviation of \(\nu_{it}\) is the regression’s estimated root mean squared error, \(s^2\), which is adjusted (as previously stated) for the \(n - 1\) estimated means.

\(R^2\) within is reported as the \(R^2\) from the mean-deviated regression.

\(R^2\) between is reported as \(\left\{\text{corr}(\bar{Z}_i \hat{\delta}, \bar{y}_i)\right\}^2\).

\(R^2\) overall is reported as \(\left\{\text{corr}(\bar{Z}_{it} \hat{\delta}, y_{it})\right\}^2\).

At the bottom of the output, an \(F\) statistics against the null hypothesis that all the \(\mu_i\) are zero is reported. This \(F\) statistic is an application of the results in Wooldridge (1990).

**xtivreg, be**

After passing (2) through the between transform, we are left with
\[
\bar{y}_i = \alpha + \bar{Z}_i \delta + \mu_i + \bar{\nu}_i
\]
where
\[
\bar{w}_i = \frac{1}{T_i} \sum_{t=1}^{T_i} w_{it} \quad \text{for } w \in \{y, Z, \nu\}
\]
Similarly, define \(\bar{X}_i\) as the matrix of instruments \(X_{it}\) after they have been passed through the between transform.
The BE2SLS estimator of (3) obtains its coefficient estimates and its conventional VCE, a two-stage least-squares regression of \( \bar{y}_i \) on \( \bar{Z}_i \) with instruments \( \bar{X}_i \) in which each average appears \( T_i \) times.

\( R^2 \) between is reported as the \( R^2 \) from the fitted regression.

\( R^2 \) within is reported as \[ \left( \text{corr}\left( (Z_{it} - \bar{Z}_i) \hat{\delta}, y_{it} - \bar{y}_i \right) \right)^2. \]

\( R^2 \) overall is reported as \[ \left( \text{corr}(Z_{it} \hat{\delta}, y_{it}) \right)^2. \]

**xtivreg, re**

Per Baltagi and Chang (2000), let

\[ u = \mu_i + \nu_{it} \]

be the \( N \times 1 \) vector of combined errors. Then under the assumptions of the random-effects model,

\[ E(uu') = \sigma^2_\nu \text{diag} \left( I_{T_i} - \frac{1}{T_i} \nu_T \nu_T' \right) + \text{diag} \left( w_i \frac{1}{T_i} \nu_T \nu_T' \right) \]

where

\[ \omega_i = T_i \sigma^2_\nu + \sigma^2_\nu \]

and \( \nu_T \) is a vector of ones of dimension \( T_i \).

Because the variance components are unknown, consistent estimates are required to implement feasible GLS. **xtivreg** offers two choices. The default is a simple extension of the Swamy–Arora method for unbalanced panels.

Let

\[ u_{it}^w = \tilde{y}_{it} - \tilde{Z}_{it} \hat{\delta}_w \]

be the combined residuals from the within estimator. Let \( \tilde{u}_{it} \) be the within-transformed \( u_{it} \). Then

\[ \hat{\sigma}_\nu = \frac{\sum_{i=1}^{n} \sum_{t=1}^{T_i} \tilde{u}_{it}^2}{N - n - K + 1} \]

Let

\[ u_{it}^b = y_{it} - Z_{it} \delta_b \]

be the combined residual from the between estimator. Let \( \bar{u}_{i.}^b \) be the between residuals after they have been passed through the between transform. Then

\[ \hat{\sigma}_\mu^2 = \frac{\sum_{i=1}^{n} \sum_{t=1}^{T_i} \bar{u}_{it}^2 - (n - K) \hat{\sigma}_\nu^2}{N - r} \]

where

\[ r = \text{trace} \left\{ \left( Z_i' Z_i \right)^{-1} Z_i' Z_{i.} Z_{i.}' Z_i \right\} \]

where

\[ Z_{i.} = \text{diag} \left( \nu_{T_i} \nu_{T_i}' \right) \]
If the nosa option is specified, the consistent estimators described in Baltagi and Chang (2000) are used. These are given by

\[
\hat{\sigma}_\nu = \frac{\sum_{i=1}^{n} \sum_{t=1}^{T_i} \tilde{u}_{it}^2}{N - n}
\]

and

\[
\hat{\sigma}_\mu^2 = \frac{\sum_{i=1}^{n} \sum_{t=1}^{T_i} \bar{u}_{it}^2 - n \hat{\sigma}_\nu^2}{N}
\]

The default Swamy–Arora method contains a degree-of-freedom correction to improve its performance in small samples.

Given estimates of the variance components, \(\hat{\sigma}_\nu^2\) and \(\hat{\sigma}_\mu^2\), the feasible GLS transform of a variable \(w\) is

\[
w^* = w_{it} - \hat{\theta}_{it} \bar{w}_i.
\]

where

\[
\bar{w}_i = \frac{1}{T_i} \sum_{t=1}^{T_i} w_{it}
\]

\[
\hat{\theta}_{it} = 1 - \left( \frac{\hat{\sigma}_\nu^2}{\hat{\omega}_i} \right)^{-\frac{1}{2}}
\]

and

\[
\hat{\omega}_i = T_i \hat{\sigma}_\mu^2 + \hat{\sigma}_\nu^2
\]

Using either estimator of the variance components, \texttt{xtivreg} contains two GLS estimators of the random-effects model. These two estimators differ only in how they construct the GLS instruments from the exogenous and instrumental variables contained in \(X_{it} = [X_{1it} \, X_{2it}]\). The default method, G2SLS, which is from Balestra and Varadharajan-Krishnakumar, uses the exogenous variables after they have been passed through the feasible GLS transform. Mathematically, G2SLS uses \(X^*\) for the GLS instruments, where \(X^*\) is constructed by passing each variable in \(X\) through the GLS transform in (4). The G2SLS estimator obtains its coefficient estimates and conventional VCE from an instrumental variable regression of \(y^*_{it}\) on \(Z^*_{it}\) with instruments \(X^*_{it}\).

If the ec2sls option is specified, \texttt{xtivreg} performs Baltagi’s EC2SLS. In EC2SLS, the instruments are \(\bar{X}_{it}\) and \(\bar{X}_{it}\), where \(\bar{X}_{it}\) is constructed by each of the variables in \(X_{it}\) throughout the GLS transform in (4), and \(\bar{X}_{it}\) is made of the group means of each variable in \(X_{it}\). The EC2SLS estimator obtains its coefficient estimates and its VCE from an instrumental variables regression of \(y^*_{it}\) on \(Z^*_{it}\) with instruments \(\bar{X}_{it}\) and \(\bar{X}_{it}\).

Baltagi and Li (1992) show that although the G2SLS instruments are a subset of those in EC2SLS, the extra instruments in EC2SLS are redundant in the sense of White (2001). Given the extra computational cost, G2SLS is the default.

The standard deviation of \(\mu_i + \nu_{it}\) is calculated as \(\sqrt{\hat{\sigma}_\mu^2 + \hat{\sigma}_\nu^2}\).
\[ R^2 \text{ between is reported as } \left\{ \text{corr}(\hat{Z}_i \delta, \bar{y}_i) \right\}^2. \]

\[ R^2 \text{ within is reported as } \left[ \text{corr}\{ (Z_{it} - \bar{Z}_i) \delta, y_{it} - \bar{y}_i \} \right]^2. \]

\[ R^2 \text{ overall is reported as } \left\{ \text{corr}(\hat{Z}_{it} \delta, y_{it}) \right\}^2. \]

**Acknowledgment**

We thank Mead Over of the Center for Global Development, who wrote an early implementation of xtivreg.

**References**


**Also see**

[XT] **xtivreg postestimation** — Postestimation tools for xtivreg
[XT] **xtset** — Declare data to be panel data
[XT] **xtabond** — Arellano–Bond linear dynamic panel-data estimation
[XT] **xthtaylor** — Hausman–Taylor estimator for error-components models
[XT] **xtreg** — Fixed-, between-, and random-effects and population-averaged linear models
[R] **ivregress** — Single-equation instrumental-variables regression
[U] 20 Estimation and postestimation commands
xtivreg postestimation — Postestimation tools for xtivreg

Description

The following postestimation commands are available after xtivreg:

<table>
<thead>
<tr>
<th>Command</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>contrast</td>
<td>contrasts and ANOVA-style joint tests of estimates</td>
</tr>
<tr>
<td>estat summarize</td>
<td>summary statistics for the estimation sample</td>
</tr>
<tr>
<td>estat vce</td>
<td>variance–covariance matrix of the estimators (VCE)</td>
</tr>
<tr>
<td>estimates</td>
<td>cataloging estimation results</td>
</tr>
<tr>
<td>forecast</td>
<td>dynamic forecasts and simulations</td>
</tr>
<tr>
<td>hausman</td>
<td>Hausman’s specification test</td>
</tr>
<tr>
<td>lincom</td>
<td>point estimates, standard errors, testing, and inference for linear combinations of coefficients</td>
</tr>
<tr>
<td>margins</td>
<td>marginal means, predictive margins, marginal effects, and average marginal effects</td>
</tr>
<tr>
<td>marginsplot</td>
<td>graph the results from margins (profile plots, interaction plots, etc.)</td>
</tr>
<tr>
<td>nlcom</td>
<td>point estimates, standard errors, testing, and inference for nonlinear combinations of coefficients</td>
</tr>
<tr>
<td>predict</td>
<td>predictions, residuals, influence statistics, and other diagnostic measures</td>
</tr>
<tr>
<td>predictnl</td>
<td>point estimates, standard errors, testing, and inference for generalized predictions</td>
</tr>
<tr>
<td>pwcompare</td>
<td>pairwise comparisons of estimates</td>
</tr>
<tr>
<td>test</td>
<td>Wald tests of simple and composite linear hypotheses</td>
</tr>
<tr>
<td>testnl</td>
<td>Wald tests of nonlinear hypotheses</td>
</tr>
</tbody>
</table>

Syntax for predict

For all but the first-differenced estimator

```
predict [ type ] newvar [ if ] [ in ] [ , statistic ]
```

First-differenced estimator

```
predict [ type ] newvar [ if ] [ in ] [ , FD_statistic ]
```
### Postestimation tools for xtivreg

#### Description

**Main**

<table>
<thead>
<tr>
<th>statistic</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>xb</td>
<td>$Z_{it}\hat{\delta}$, fitted values; the default</td>
</tr>
<tr>
<td>ue</td>
<td>$\hat{\mu}<em>i + \hat{\nu}</em>{it}$, the combined residual</td>
</tr>
<tr>
<td>*xbu</td>
<td>$Z_{it}\hat{\delta} + \hat{\mu}_i$, prediction including effect</td>
</tr>
<tr>
<td>*u</td>
<td>$\hat{\mu}_i$, the fixed- or random-error component</td>
</tr>
<tr>
<td>*e</td>
<td>$\hat{\nu}_{it}$, the overall error component</td>
</tr>
</tbody>
</table>

Unstarred statistics are available both in and out of sample; type `predict ... if e(sample) ...` if wanted only for the estimation sample. Starred statistics are calculated only for the estimation sample, even when `if e(sample)` is not specified.

#### FD_statistics

<table>
<thead>
<tr>
<th>statistic</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>xb</td>
<td>$x_jb$, fitted values for the first-differenced model; the default</td>
</tr>
<tr>
<td>e</td>
<td>$e_{it} - e_{it-1}$, the first-differenced overall error component</td>
</tr>
</tbody>
</table>

These statistics are available both in and out of sample; type `predict ... if e(sample) ...` if wanted only for the estimation sample.

---

### Menu for predict

Statistics > Postestimation > Predictions, residuals, etc.

### Options for predict

- **xb**, the default, calculates the linear prediction, that is, $Z_{it}\hat{\delta}$.
- **ue** calculates the prediction of $\hat{\mu}_i + \hat{\nu}_{it}$. This is not available after the first-differenced model.
- **xbu** calculates the prediction of $Z_{it}\hat{\delta} + \hat{\mu}_i$, the prediction including the fixed or random component. This is not available after the first-differenced model.
- **u** calculates the prediction of $\hat{\mu}_i$, the estimated fixed or random effect. This is not available after the first-differenced model.
- **e** calculates the prediction of $\hat{\nu}_{it}$.

### Also see

[XT] xtivreg — Instrumental variables and two-stage least squares for panel-data models

[U] 20 Estimation and postestimation commands
xtline — Panel-data line plots

Syntax

Options for graph by panel

Menu

Options for overlaid panels

Also see

Remarks and examples

Syntax

Graph by panel

xtline varlist [if] [in] [ , panel_options ]

Overlaid panels

xtline varname [if] [in] , overlay [ overlaid_options ]

panel_options

<table>
<thead>
<tr>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Main</td>
</tr>
<tr>
<td>i(varname_i) use varname_i as the panel ID variable</td>
</tr>
<tr>
<td>t(varname_t) use varname_t as the time variable</td>
</tr>
<tr>
<td>Plot</td>
</tr>
<tr>
<td>cline_options affect rendition of the plotted points connected by lines</td>
</tr>
<tr>
<td>Add plots</td>
</tr>
<tr>
<td>addplot(plot) add other plots to the generated graph</td>
</tr>
</tbody>
</table>

Overlaid_options

<table>
<thead>
<tr>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Main</td>
</tr>
<tr>
<td>overlay overlay each panel on the same graph</td>
</tr>
<tr>
<td>i(varname_i) use varname_i as the panel ID variable</td>
</tr>
<tr>
<td>t(varname_t) use varname_t as the time variable</td>
</tr>
<tr>
<td>Plots</td>
</tr>
<tr>
<td>plot#opts(cline_options) affect rendition of the # panel line</td>
</tr>
<tr>
<td>Add plots</td>
</tr>
<tr>
<td>addplot(plot) add other plots to the generated graph</td>
</tr>
</tbody>
</table>

A panel variable and a time variable must be specified. Use xtset (see XT xtset) or specify the i() and t() options. The t() option allows noninteger values for the time variable, whereas xtset does not. 221
Menu
Statistics > Longitudinal/panel data > Line plots

Description
xtline draws line plots for panel data.

Options for graph by panel

i(varname_i) and t(varname_t) override the panel settings from xtset; see [XT] xtset. varname_i is allowed to be a string variable. varname_t can take on noninteger values and have repeated values within panel. That is to say, it can be any numeric variable that you would like to specify for the x-dimension of the graph. It is an error to specify i() without t() and vice versa.

cline_options affect the rendition of the plotted points connected by lines; see [G-3] cline_options.

addplot(plot) provides a way to add other plots to the generated graph; see [G-3] addplot_option.

twoway_options are any of the options documented in [G-3] twoway_options, excluding by(). These include options for titling the graph (see [G-3] title_options) and for saving the graph to disk (see [G-3] saving_option).

byopts(byopts) allows all the options documented in [G-3] by_option. These options affect the appearance of the by-graph. byopts() may not be combined with overlay.

Options for overlaid panels

overlay causes the plot from each panel to be overlaid on the same graph. The default is to generate plots by panel. This option may not be combined with byopts() or be specified when there are multiple variables in varlist.

i(varname_i) and t(varname_t) override the panel settings from xtset; see [XT] xtset. varname_i is allowed to be a string variable. varname_t can take on noninteger values and have repeated values within panel. That is to say, it can be any numeric variable that you would like to specify for the x-dimension of the graph. It is an error to specify i() without t() and vice versa.

plot#opts(cline_options) affect the rendition of the #th panel (in sorted order). The cline_options can affect whether and how the points are connected; see [G-3] cline_options.

addplot(plot) provides a way to add other plots to the generated graph; see [G-3] addplot_option.
Remarks and examples

Example 1

Suppose that Tess, Sam, and Arnold kept a calorie log for an entire calendar year. At the end of the year, if they pooled their data together, they would have a dataset (for example, xtline1.dta) that contains the number of calories each of them consumed for 365 days. They could then use xtset to identify the date variable and treat each person as a panel and use xtline to plot the calories versus time for each person separately.

```
. use http://www.stata-press.com/data/r13/xtline1
. xtset person day
   panel variable:  person (strongly balanced)
   time variable:   day, 01jan2002 to 31dec2002
   delta:           1 day
. xtline calories, tlabel(#3)
```

Specify the overlay option so that the values are plotted on the same graph to provide a better comparison among Tess, Sam, and Arnold.
. xtline calories, overlay

Also see

[XT] xtset — Declare data to be panel data

[G-2] graph twoway — Twoway graphs

[TS] tsline — Plot time-series data
Title

**xtlogit — Fixed-effects, random-effects, and population-averaged logit models**

<table>
<thead>
<tr>
<th>Syntax</th>
<th>Menu</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Options for RE model</td>
<td>Options for FE model</td>
<td>Options for PA model</td>
</tr>
<tr>
<td>Remarks and examples</td>
<td>Stored results</td>
<td>Methods and formulas</td>
</tr>
<tr>
<td>References</td>
<td>Also see</td>
<td></td>
</tr>
</tbody>
</table>

**Syntax**

**Random-effects (RE) model**

```
xtlogit depvar [indepvars] [if] [in] [weight] [, re RE_options]
```

**Conditional fixed-effects (FE) model**

```
xtlogit depvar [indepvars] [if] [in] [weight] , fe [FE_options]
```

**Population-averaged (PA) model**

```
xtlogit depvar [indepvars] [if] [in] [weight] , pa [PA_options]
```
### Model

- **noconstant**
  - Description: suppress constant term
- **re**
  - Description: use random-effects estimator; the default
- **offset(varname)**
  - Description: include `varname` in model with coefficient constrained to 1
- **constraints(constraints)**
  - Description: apply specified linear constraints
- **collinear**
  - Description: keep collinear variables
- **asis**
  - Description: retain perfect predictor variables

### SE/Robust

- **vce(vcetype)**
  - Description: `vcetype` may be `oim`, `robust`, `cluster clustvar`, `bootstrap`, or `jackknife`

### Reporting

- **level(#)***
  - Description: set confidence level; default is `level(95)`
- **orskip**
  - Description: report odds ratios
- **nocnsreport**
  - Description: perform overall model test as a likelihood-ratio test
- **display_options**
  - Description: control column formats, row spacing, line width, display of omitted variables and base and empty cells, and factor-variable labeling

### Integration

- **intmethod(intmethod)**
  - Description: integration method; `intmethod` may be `mvaghermite` (the default) or `ghermite`
- **intpoints(#)***
  - Description: use `#` quadrature points; default is `intpoints(12)`

### Maximization

- **maximize_options**
  - Description: control the maximization process; seldom used
- **nodisplay**
  - Description: suppress display of header and coefficients
- **coeflegend**
  - Description: display legend instead of statistics
### FE_options Description

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Model</strong></td>
<td></td>
</tr>
<tr>
<td><code>fe</code></td>
<td>use fixed-effects estimator</td>
</tr>
<tr>
<td><code>offset(varname)</code></td>
<td>include <code>varname</code> in model with coefficient constrained to 1</td>
</tr>
<tr>
<td><code>constraints(constraints)</code></td>
<td>apply specified linear constraints</td>
</tr>
<tr>
<td><code>collinear</code></td>
<td>keep collinear variables</td>
</tr>
<tr>
<td><strong>SE</strong></td>
<td></td>
</tr>
<tr>
<td><code>vce(vcetype)</code></td>
<td><code>vcetype</code> may be <code>oim</code>, <code>bootstrap</code>, or <code>jackknife</code></td>
</tr>
<tr>
<td><strong>Reporting</strong></td>
<td></td>
</tr>
<tr>
<td><code>level(#)</code></td>
<td>set confidence level; default is <code>level(95)</code></td>
</tr>
<tr>
<td>or <code>norskip</code></td>
<td>report odds ratios</td>
</tr>
<tr>
<td><code>norsnsreport</code></td>
<td>perform overall model test as a likelihood-ratio test</td>
</tr>
<tr>
<td><code>display_options</code></td>
<td>control column formats, row spacing, line width, display of omitted variables and base and empty cells, and factor-variable labeling</td>
</tr>
<tr>
<td><strong>Maximization</strong></td>
<td></td>
</tr>
<tr>
<td><code>maximize_options</code></td>
<td>control the maximization process; seldom used</td>
</tr>
<tr>
<td><code>nodisplay</code></td>
<td>suppress display of header and coefficients</td>
</tr>
<tr>
<td><code>coeflegend</code></td>
<td>display legend instead of statistics</td>
</tr>
</tbody>
</table>
### Model

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>noconstant</code></td>
<td>suppress constant term</td>
</tr>
<tr>
<td><code>pa</code></td>
<td>use population-averaged estimator</td>
</tr>
<tr>
<td><code>offset(vname)</code></td>
<td>include <code>vname</code> in model with coefficient constrained to 1</td>
</tr>
<tr>
<td><code>asis</code></td>
<td>retain perfect predictor variables</td>
</tr>
</tbody>
</table>

### Correlation

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>corr(correlation)</code></td>
<td>within-panel correlation structure</td>
</tr>
<tr>
<td><code>force</code></td>
<td>estimate even if observations unequally spaced in time</td>
</tr>
</tbody>
</table>

### SE/Robust

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>vce(vcetype)</code></td>
<td><code>vcetype</code> may be conventional, robust, bootstrap, or jackknife</td>
</tr>
<tr>
<td><code>nmp</code></td>
<td>use divisor $N - P$ instead of the default $N$</td>
</tr>
<tr>
<td><code>scale(parm)</code></td>
<td>overrides the default scale parameter; <code>parm</code> may be x2, dev, phi, or #</td>
</tr>
</tbody>
</table>

### Reporting

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>level(#)</code></td>
<td>set confidence level; default is <code>level(95)</code></td>
</tr>
<tr>
<td><code>display_options</code></td>
<td>control column formats, row spacing, line width, display of omitted variables and base and empty cells, and factor-variable labeling</td>
</tr>
</tbody>
</table>

### Optimization

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>optimize_options</code></td>
<td>control the optimization process; seldom used</td>
</tr>
<tr>
<td><code>nodisplay</code></td>
<td>do not display the header and coefficients</td>
</tr>
<tr>
<td><code>coeflegend</code></td>
<td>display legend instead of statistics</td>
</tr>
</tbody>
</table>

### correlation

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>exchangeable</code></td>
<td>exchangeable</td>
</tr>
<tr>
<td><code>independent</code></td>
<td>independent</td>
</tr>
<tr>
<td><code>unstructured</code></td>
<td>unstructured</td>
</tr>
<tr>
<td><code>fixed matname</code></td>
<td>user-specified</td>
</tr>
<tr>
<td><code>ar #</code></td>
<td>autoregressive of order #</td>
</tr>
<tr>
<td><code>stationary #</code></td>
<td>stationary of order #</td>
</tr>
<tr>
<td><code>nonstationary #</code></td>
<td>nonstationary of order #</td>
</tr>
</tbody>
</table>
A panel variable must be specified. For *xtlogit*, the correlation structures other than exchangeable and independent require that a time variable also be specified. Use *xtset*; see [XT] *xtset*.

*indepvars* may contain factor variables; see [U] 11.4.3 Factor variables.

*depvar* and *indepvars* may contain time-series operators; see [U] 11.4.4 Time-series varlists.

by, mi estimate, and statsby are allowed; see [U] 11.1.10 Prefix commands. *fp* is allowed for the random-effects and fixed-effects models.

*vce(bootstrap)* and *vce(jackknife)* are not allowed with the mi estimate prefix; see [MI] *mi estimate*. *iweights*, *fweights*, and *pweights* are allowed for the population-averaged model, and *iweights* are allowed for the fixed-effects and random-effects models; see [U] 11.1.6 weight. Weights must be constant within panel.

*nodisplay* and *coeflegend* do not appear in the dialog box.

See [U] 20 Estimation and postestimation commands for more capabilities of estimation commands.

---

**Menu**

Statistics > Longitudinal/panel data > Binary outcomes > Logistic regression (FE, RE, PA)

**Description**

*xtlogit* fits random-effects, conditional fixed-effects, and population-averaged logit models. Whenever we refer to a fixed-effects model, we mean the conditional fixed-effects model. *depvar* equal to nonzero and nonmissing (typically *depvar* equal to one) indicates a positive outcome, whereas *depvar* equal to zero indicates a negative outcome.

By default, the population-averaged model is an equal-correlation model; *xtlogit*, pa assumes *corr(exchangeable)*. See [XT] *xtgee* for information on how to fit other population-averaged models.

See [R] *logistic* for a list of related estimation commands.

---

**Options for RE model**

- **Model**
  - *noconstant*; see [R] estimation options.
  - *re* requests the random-effects estimator, which is the default.
  - *offset(varname)* *constraints(constraints)*, *collinear*; see [R] estimation options.
  - *asis* forces retention of perfect predictor variables and their associated, perfectly predicted observations and may produce instabilities in maximization; see [R] *probit*.

- **SE/Robust**
  - *vce(vcetype)* specifies the type of standard error reported, which includes types that are derived from asymptotic theory (oim), that are robust to some kinds of misspecification (robust), that allow for intragroup correlation (cluster *clustvar*), and that use bootstrap or jackknife methods (bootstrap, jackknife); see [XT] *vce_options*.

    Specifying *vce(robust)* is equivalent to specifying *vce(cluster panelvar)*; see *xtlogit, re and the robust VCE estimator* in Methods and formulas.

- **Reporting**
  - *level(#)*; see [R] estimation options.
or reports the estimated coefficients transformed to odds ratios, that is, \( e^b \) rather than \( b \). Standard errors and confidence intervals are similarly transformed. This option affects how results are displayed, not how they are estimated. or may be specified at estimation or when replaying previously estimated results.

noskip; see [R] estimation options.
nocnsreport; see [R] estimation options.

display_options: noomitted, vsquish, noemptycells, baselevels, alllbaselevels, nofvlabel, fvwrap(#), fvwrapon(style), cformat(%fmt), pformat(%fmt), sformat(%fmt), and nolstretch; see [R] estimation options.

Integration

intmethod(intmethod), intpoints(#); see [R] estimation options.

Maximization

maximize_options: difficult, technique(algorithm_spec), iterate(#), [no]log, trace, gradient, showstep, hessian, showtolerance, tolerance(#), ltolerance(#), nrtolerance(#), nonrtolerance, and from(init_specs); see [R] maximize. These options are seldom used.

The following options are available with xtlogit but are not shown in the dialog box: nodisplay is for programmers. It suppresses the display of the header and the coefficients. coeflegend; see [R] estimation options.

Options for FE model

fe requests the fixed-effects estimator.

offset(varname), constraints(constraints), collinear; see [R] estimation options.

vce(vcetype) specifies the type of standard error reported, which includes types that are derived from asymptotic theory (oim) and that use bootstrap or jackknife methods (bootstrap, jackknife); see [XT] vce_options.

level(#); see [R] estimation options.

or reports the estimated coefficients transformed to odds ratios, that is, \( e^b \) rather than \( b \). Standard errors and confidence intervals are similarly transformed. This option affects how results are displayed, not how they are estimated. or may be specified at estimation or when replaying previously estimated results.

noskip; see [R] estimation options.
nocnsreport; see [R] estimation options.
display_options: noomitted, vsquish, noemptycells, baselevels, allbaselevels, nobelabel, fvwrap(#), fvwrapon(style), cformat(%,fmt), pformat(%,fmt), sformat(%,fmt), and nolstretch; see [R] estimation options.

maximize_options: difficult, technique(algorithm_spec), iterate(#), [no]log, trace, gradient, showstep, hessian, showtolerance, tolerance(#), ltolerance(#), nrtolerance(#), nonrtolerance, and from(init_specs); see [R] maximize. These options are seldom used.

The following options are available with xtlogit but are not shown in the dialog box:

nolog is for programmers. It suppresses the display of the header and the coefficients.

coeflegend; see [R] estimation options.

Options for PA model

noconstant; see [R] estimation options.

pa requests the population-averaged estimator.

offset(varname); see [R] estimation options.

asis forces retention of perfect predictor variables and their associated, perfectly predicted observations and may produce instabilities in maximization; see [R] probit.

corr(correlation) specifies the within-panel correlation structure; the default corresponds to the equal-correlation model, corr(exchangeable).

When you specify a correlation structure that requires a lag, you indicate the lag after the structure’s name with or without a blank; for example, corr(ar 1) or corr(ar1).

If you specify the fixed correlation structure, you specify the name of the matrix containing the assumed correlations following the word fixed, for example, corr(fixed myr).

force specifies that estimation be forced even though the time variable is not equally spaced.

This is relevant only for correlation structures that require knowledge of the time variable. These correlation structures require that observations be equally spaced so that calculations based on lags correspond to a constant time change. If you specify a time variable indicating that observations are not equally spaced, the (time dependent) model will not be fit. If you also specify force, the model will be fit, and it will be assumed that the lags based on the data ordered by the time variable are appropriate.

vce(vcetype) specifies the type of standard error reported, which includes types that are derived from asymptotic theory (conventional), that are robust to some kinds of misspecification (robust), and that use bootstrap or jackknife methods (bootstrap, jackknife); see [XT] vce_options.

vce(conventional), the default, uses the conventionally derived variance estimator for generalized least-squares regression.

nmp, scale(x2|dev|phi|#); see [XT] vce_options.
or reports the estimated coefficients transformed to odds ratios, that is, \( e^b \) rather than \( b \). Standard errors
and confidence intervals are similarly transformed. This option affects how results are displayed, not how they are estimated. or may be specified at estimation or when replaying previously
estimated results.

\[
\text{display_options: noomitted, vsquish, noemptycells, baselevels, allbaselevels,nofvlabel, fwrap(\#), fwrapon(style), cformat(\%fmt), pformat(\%fmt), sformat(\%fmt), and nolstretch; see [R] estimation options.}
\]

Optimization

\[
\text{optimize_options control the iterative optimization process. These options are seldom used.}
\]

\[
\text{iterate(\#) specifies the maximum number of iterations. When the number of iterations equals \#,
the optimization stops and presents the current results, even if convergence has not been reached.}
\]

\[
\text{The default is iterate(100).}
\]

\[
\text{tolerance(\#) specifies the tolerance for the coefficient vector. When the relative change in the}
\]

\[
\text{coefficient vector from one iteration to the next is less than or equal to \#, the optimization process}
\]

\[
\text{is stopped. tolerance(1e-6) is the default.}
\]

\[
\text{nolog suppresses display of the iteration log.}
\]

\[
\text{trace specifies that the current estimates be printed at each iteration.}
\]

The following options are available with xtlogit but are not shown in the dialog box:

\[
\text{nodisplay is for programmers. It suppresses the display of the header and the coefficients.}
\]

\[
\text{coeflegend; see [R] estimation options.}
\]

Remarks and examples

\[
\text{xtlogit is a convenience command if you want the population-averaged model. Typing}
\]

\[
\text{. xtlogit ... , pa ...}
\]

\[
\text{is equivalent to typing}
\]

\[
\text{. xttgee ... , ... family(binomial) link(logit) corr(exchangeable)}
\]

\[
\text{It is also a convenience command if you want the fixed-effects model. Typing}
\]

\[
\text{. xtlogit ... , fe ...}
\]

\[
\text{is equivalent to typing}
\]

\[
\text{. clogit ... , group(varname_i) ...}
\]

\[
\text{See also [XT] xttgee and [R] clogit for information about xtlogit.}
\]

By default or when re is specified, xtlogit fits via maximum likelihood the random-effects model

\[
\Pr(y_{it} \neq 0|x_{it}) = P(x_{it}\beta + \nu_{i})
\]

for \( i = 1, \ldots, n \) panels, where \( t = 1, \ldots, n_i, \nu_i \) are i.i.d., \( N(0, \sigma^2_{\nu}) \), and \( P(z) = \{1 + \exp(-z)\}^{-1} \).
Underlying this model is the variance components model

\[ y_{it} \neq 0 \iff x_{it}\beta + \nu_i + \epsilon_{it} > 0 \]

where \( \epsilon_{it} \) are i.i.d. logistic distributed with mean zero and variance \( \sigma^2_\epsilon = \pi^2/3 \), independently of \( \nu_i \).

Example 1

We are studying unionization of women in the United States and are using the union dataset; see [XT] xt. We wish to fit a random-effects model of union membership:

\[
\]

\[
\text{. xtlogit union age grade not_smsa south##c.year (output omitted)}
\]

Random-effects logistic regression

<table>
<thead>
<tr>
<th></th>
<th>Number of obs = 26200</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group variable: idcode</td>
<td>Number of groups = 4434</td>
</tr>
<tr>
<td>Random effects u_i ~ Gaussian</td>
<td>Obs per group: min = 1, avg = 5.9, max = 12</td>
</tr>
<tr>
<td>Integration method: mvaghermite</td>
<td>Integration points = 12</td>
</tr>
</tbody>
</table>

Log likelihood = -10540.274

<table>
<thead>
<tr>
<th></th>
<th>Wald ch(\hat{2})(6) = 227.46</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Prob &gt; ch(\hat{2}) = 0.0000</td>
</tr>
</tbody>
</table>

|                | Coef.  | Std. Err. | z     | P>|z| | [95% Conf. Interval] |
|----------------|--------|-----------|-------|-------|----------------------|
| union          | .0156732 | .0149895  | 1.05  | 0.296 | -.0137056 .045052   |
| age            | .0870851 | .0176476  | 4.93  | 0.000 | .0524965 .1216738   |
| grade          | -.2511884 | .0823508  | -3.05 | 0.002 | -.4125929 -.0897839 |
| not_smsa       | -2.839112 | .6413116  | -4.43 | 0.000 | -4.096059 -1.582164 |
| south          | -.0068604 | .0156575  | -0.44 | 0.661 | -.0375486 .0238277  |
| year           | .0238506 | .0079732  | 2.99  | 0.003 | .0082235 .0394777   |
| south#c.year 1 | .0309365 | .8414963  | -3.58 | 0.000 | -4.658667 -1.360062 |
| _cons          | -3.009365 | .8414963  | -3.58 | 0.000 | -4.658667 -1.360062 |

\[ /lnsig2u \]

|                | Coef.  | Std. Err. | z     | P>|z| | [95% Conf. Interval] |
|----------------|--------|-----------|-------|-------|----------------------|
| sigma_u        | 2.398116 | .0563577  | 4.20  | 0.000 | 2.290162 2.511158   |
| rho            | .6361098 | .0108797  | 58.85 | 0.000 | .6145307 .6571548   |

Likelihood-ratio test of rho=0: chibar2(01) = 6004.43 Prob > = chibar2 = 0.000

The output includes the additional panel-level variance component. This is parameterized as the log of the variance \( \ln(\sigma^2_\nu) \) (labeled \( lnsig2u \) in the output). The standard deviation \( \sigma_\nu \) is also included in the output and labeled \( sigma_u \) together with \( \rho \) (labeled \( rho \)),

\[
\rho = \frac{\sigma^2_\nu}{\sigma^2_\nu + \sigma^2_\epsilon}
\]

which is the proportion of the total variance contributed by the panel-level variance component.

When \( \rho \) is zero, the panel-level variance component is unimportant, and the panel estimator is no different from the pooled estimator. A likelihood-ratio test of this is included at the bottom of the output. This test formally compares the pooled estimator (logit) with the panel estimator.

As an alternative to the random-effects specification, we might want to fit an equal-correlation logit model:
. xtlogit union age grade not_smsa south##c.year, pa

Iteration 1: tolerance = .1487877
Iteration 2: tolerance = .00949342
Iteration 3: tolerance = .00040606
Iteration 4: tolerance = .0001602
Iteration 5: tolerance = 6.628e-07

GEE population-averaged model
Number of obs = 26200
Group variable: idcode Number of groups = 4434
Link: logit Obs per group: min = 1
Family: binomial avg = 5.9
Correlation: exchangeable max = 12

Wald chi2(6) = 235.08
Scale parameter: 1 Prob > chi2 = 0.0000

|             | Coef. | Std. Err. | z     | P>|z|    | [95% Conf. Interval] |
|-------------|-------|-----------|-------|--------|----------------------|
| union       |       |           |       |        |                      |
| age         | .0165893 | .0092229  | 1.80  | 0.072  | -.0014873            |
| grade       | .0600669 | .0108343  | 5.54  | 0.000  | .0388321             |
| not_smsa    | -.1215445 | .0483713  | -2.51 | 0.012  | -.2163505            |
| 1.south     | -1.857094 | .372967   | -4.98 | 0.000  | -2.588096            |
| year        | -.0121168 | .0095707  | -1.27 | 0.205  | -.030875             |
| south#c.year 1 | .0160193 | .0046076  | 3.48  | 0.001  | .0069886             |
| _cons       | -1.39755  | .5089508  | -2.75 | 0.006  | -2.395075            |

Example 2

xtlogit with the pa option allows a vce(robust) option, so we can obtain the population-averaged logit estimator with the robust variance calculation by typing

. xtlogit union age grade not_smsa south##c.year, pa vce(robust) nolog

(Std. Err. adjusted for clustering on idcode)

|             | Coef. | Std. Err. | z     | P>|z|    | [95% Conf. Interval] |
|-------------|-------|-----------|-------|--------|----------------------|
| union       |       |           |       |        |                      |
| age         | .0165893 | .008951   | 1.85  | 0.064  | -.0009543            |
| grade       | .0600669 | .0133193  | 4.51  | 0.000  | .0339616             |
| not_smsa    | -.1215445 | .0613803 | -1.98 | 0.048  | -.2418477            |
| 1.south     | -1.857094 | .5389238 | -3.45 | 0.001  | -2.913366            |
| year        | -.0121168 | .0096998 | -1.25 | 0.212  | -.0311282            |
| south#c.year 1 | .0160193 | .0067217  | 2.38  | 0.017  | .002845              |
| _cons       | -1.39755  | .5603767  | -2.49 | 0.013  | -2.495868            |

These standard errors are somewhat larger than those obtained without the vce(robust) option.
Finally, we can also fit a fixed-effects model to these data (see also [R] clogit for details):

```
. xtlogit union age grade not_smsa south##c.year, fe
  note: multiple positive outcomes within groups encountered.
  note: 2744 groups (14165 obs) dropped because of all positive or
        all negative outcomes.
Iteration 0: log likelihood = -4516.5881
Iteration 1: log likelihood = -4510.8906
Iteration 2: log likelihood = -4510.888
Iteration 3: log likelihood = -4510.888

Conditional fixed-effects logistic regression

Number of obs = 12035
Group variable: idcode
Number of groups = 1690
Obs per group: min = 2
              avg = 7.1
              max = 12

LR chi2(6) = 78.60
Log likelihood = -4510.888
Prob > chi2 = 0.0000

        union | Coef.    Std. Err.     z  P>|z|      [95% Conf. Interval]
-------------+--------------------------------------------------
       age   | .0710973  .0960536  0.74  0.459    -.1171643    .2593589
     grade   | .0816111  .0419074  1.95  0.051    -.0005259   .1637480
    not_smsa | .0224809  .1131786  0.20  0.843   -.1993450    .2443069
      south  | -2.856488  .6765694 -4.22  0.000   -4.182539   -1.530436
       year   | -.0636853  .0967747 -0.66  0.510   -.2533602   .1259896
south#c.year| .0264136  .0083216  3.17  0.002   .0101036    .0427235
-------------+--------------------------------------------------
```

**Technical note**

The random-effects model is calculated using quadrature, which is an approximation whose accuracy depends partially on the number of integration points used. We can use the quadchk command to see if changing the number of integration points affects the results. If the results change, the quadrature approximation is not accurate given the number of integration points. Try increasing the number of integration points using the intpoints() option and run quadchk again. Do not attempt to interpret the results of estimates when the coefficients reported by quadchk differ substantially. See [XT] quadchk for details and [XT] xtprobit for an example.

Because the xtlogit likelihood function is calculated by Gauss–Hermite quadrature, on large problems the computations can be slow. Computation time is roughly proportional to the number of points used for the quadrature.
xtlogit, re stores the following in e():

Scalars

- `e(N)`: number of observations
- `e(N_g)`: number of groups
- `e(N_cd)`: number of completely determined observations
- `e(k)`: number of parameters
- `e(k_aux)`: number of auxiliary parameters
- `e(k_eq)`: number of equations in `e(b)`
- `e(k_eq_model)`: number of equations in overall model test
- `e(df_m)`: model degrees of freedom
- `e(ll)`: log likelihood
- `e(ll_0)`: log likelihood, constant-only model
- `e(ll_c)`: log likelihood, comparison model
- `e(chi2)`: \( \chi^2 \)
- `e(chi2_c)`: \( \chi^2 \) for comparison test
- `e(N_clust)`: number of clusters
- `e(rho)`: \( \rho \)
- `e(sigma_u)`: panel-level standard deviation
- `e(n_quad)`: number of quadrature points
- `e(g_min)`: smallest group size
- `e(g_avg)`: average group size
- `e(g_max)`: largest group size
- `e(p)`: significance
- `e(rank)`: rank of `e(V)`
- `e(rank0)`: rank of `e(V)` for constant-only model
- `e(ic)`: number of iterations
- `e(rc)`: return code
- `e(converged)`: 1 if converged, 0 otherwise

Macros

- `e(cmd)`: `xtlogit`
- `e(cmdline)`: command as typed
- `e(depvar)`: name of dependent variable
- `e(varlist)`: variable denoting groups
- `e(model)`: `re`
- `e(wtype)`: weight type
- `e(wexp)`: weight expression
- `e(title)`: title in estimation output
- `e(clustvar)`: name of cluster variable
- `e(offset)`: linear offset variable
- `e(chi2type)`: Wald or LR; type of model \( \chi^2 \) test
- `e(chi2_ct)`: Wald or LR; type of model \( \chi^2 \) test corresponding to `e(chi2_c)`
- `e(vce)`: vcetype specified in `vce()`
- `e(vcetype)`: title used to label Std. Err.
- `e(intmethod)`: integration method
- `e(distrib)`: Gaussian; the distribution of the random effect
- `e(opt)`: type of optimization
- `e(which)`: max or min; whether optimizer is to perform maximization or minimization
- `e(ml_method)`: type of ml method
- `e(user)`: name of likelihood-evaluator program
- `e(technique)`: maximization technique
- `e(properties)`: `b V`
- `e(predict)`: program used to implement predict
- `e(asbalanced)`: factor variables fvset as asbalanced
- `e(asobserved)`: factor variables fvset as asobserved
Matrices
- **e(b)**: coefficient vector
- **e(Cns)**: constraints matrix
- **e(ilog)**: iteration log
- **e(gradient)**: gradient vector
- **e(V)**: variance–covariance matrix of the estimators
- **e(V_modelbased)**: model-based variance

Functions
- **e(sample)**: marks estimation sample

**xtlogit, fe** stores the following in **e()**:

Scalars
- **e(N)**: number of observations
- **e(N_g)**: number of groups
- **e(N_drop)**: number of observations dropped because of all positive or all negative outcomes
- **e(N_group_drop)**: number of groups dropped because of all positive or all negative outcomes
- **e(k)**: number of parameters
- **e(k_eq)**: number of equations in **e(b)**
- **e(k_eq_model)**: number of equations in overall model test
- **e(k dv)**: number of dependent variables
- **e(df_m)**: model degrees of freedom
- **e(r2_p)**: pseudo R-squared
- **e(ll)**: log likelihood
- **e(ll0)**: log likelihood, constant-only model
- **e(chi2)**: χ²
- **e(g_min)**: smallest group size
- **e(g_avg)**: average group size
- **e(g_max)**: largest group size
- **e(p)**: significance
- **e(rank)**: rank of **e(V)**
- **e(ic)**: number of iterations
- **e(rc)**: return code
- **e(converged)**: 1 if converged, 0 otherwise

Macros
- **e(cmd)**: clogit
- **e(cmd2)**: xtlogit
- **e(cmdline)**: command as typed
- **e(depvar)**: name of dependent variable
- **e(ivar)**: variable denoting groups
- **e(model)**: fe
- **e(wtype)**: weight type
- **e(wexp)**: weight expression
- **e(title)**: title in estimation output
- **e(offset)**: linear offset variable
- **e(chi2type)**: LR; type of model χ² test
- **e(vce)**: **vcetype** specified in **vce()**
- **e(vcetype)**: title used to label Std. Err.
- **e(group)**: name of group() variable
- **e(multiple)**: multiple if multiple positive outcomes within groups
- **e(opt)**: type of optimization
- **e(which)**: max or min; whether optimizer is to perform maximization or minimization
- **e(ml_method)**: type of ml method
- **e(user)**: name of likelihood-evaluator program
- **e(technique)**: maximization technique
- **e(properties)**: b V
- **e(predict)**: program used to implement predict
- **e(marginsok)**: predictions allowed by margins
- **e(marginsnotok)**: predictions disallowed by margins
- **e(asbalanced)**: factor variables fvset as asbalanced
- **e(asobserved)**: factor variables fvset as asobserved
Matrices
- `e(b)` coefficient vector
- `e(Cns)` constraints matrix
- `e(ilog)` iteration log
- `e(gradient)` gradient vector
- `e(V)` variance–covariance matrix of the estimators

Functions
- `e(sample)` marks estimation sample

**xtlogit**, *pa* stores the following in *e*(()):

Scalars
- `e(N)` number of observations
- `e(N_g)` number of groups
- `e(df_m)` model degrees of freedom
- `e(chi2)` $\chi^2$
- `e(p)` significance
- `e(df_pear)` degrees of freedom for Pearson $\chi^2$
- `e(chi2_dev)` $\chi^2$ test of deviance
- `e(chi2_dis)` $\chi^2$ test of deviance dispersion
- `e(deviance)` deviance
- `e(dispers)` deviance dispersion
- `e(phi)` scale parameter
- `e(g_min)` smallest group size
- `e(g_avg)` average group size
- `e(g_max)` largest group size
- `e(rank)` rank of `e(V)`
- `e(tol)` target tolerance
- `e(dif)` achieved tolerance
- `e(rc)` return code

Macros
- `e(cmd)` `xtgee`
- `e(cmd2)` `xtlogit`
- `e(cmdline)` command as typed
- `e(depvar)` name of dependent variable
- `e(ivar)` variable denoting groups
- `e(tvar)` variable denoting time within groups
- `e(model)` `pa`
- `e(family)` `binomial`
- `e(link)` `logit`; link function
- `e(corr)` correlation structure
- `e(scale)` $x^2$, dev, phi, or #; scale parameter
- `e(vtype)` weight type
- `e(wexp)` weight expression
- `e(offset)` linear offset variable
- `e(chi2type)` Wald; type of model $\chi^2$ test
- `e(vc)` vcetype specified in `vce()`
- `e(vcetype)` title used to label Std. Err.
- `e(properties)` `b V`
- `e(predict)` program used to implement predict
- `e(marginsnotok)` predictions disallowed by margins
- `e(asbalanced)` factor variables fvset as asbalanced
- `e(asobserved)` factor variables fvset as asobserved

Matrices
- `e(b)` coefficient vector
- `e(R)` estimated working correlation matrix
- `e(V)` variance–covariance matrix of the estimators

Functions
- `e(sample)` marks estimation sample
xtlogit — Fixed-effects, random-effects, and population-averaged logit models

Methods and formulas

xtlogit reports the population-averaged results obtained using xtgee, family(binomial) link(logit) to obtain estimates. The fixed-effects results are obtained using clogit. See [XT] xtgee and [R] clogit for details on the methods and formulas.

If we assume a normal distribution, \( N(0, \sigma^2_\nu) \), for the random effects \( \nu_i \),

\[
\Pr(y_{i1}, \ldots, y_{in_i} | x_{i1}, \ldots, x_{in_i}) = \int_{-\infty}^{\infty} \frac{e^{-\nu_i^2/2\sigma^2_\nu}}{\sqrt{2\pi}\sigma_\nu} \left\{ \prod_{t=1}^{n_i} F(y_{it}, x_{it}\beta + \nu_i) \right\} d\nu_i
\]

where

\[
F(y, z) = \begin{cases} 
1 & \text{if } y \neq 0 \\
\frac{1}{1 + \exp(-z)} & \text{otherwise}
\end{cases}
\]

The panel-level likelihood \( l_i \) is given by

\[
l_i = \int_{-\infty}^{\infty} \frac{e^{-\nu_i^2/2\sigma^2_\nu}}{\sqrt{2\pi}\sigma_\nu} \left\{ \prod_{t=1}^{n_i} F(y_{it}, x_{it}\beta + \nu_i) \right\} d\nu_i
\]

\[
eq \int_{-\infty}^{\infty} g(y_{it}, x_{it}, \nu_i) d\nu_i
\]

This integral can be approximated with \( M \)-point Gauss–Hermite quadrature

\[
\int_{-\infty}^{\infty} e^{-x^2} h(x) dx \approx \sum_{m=1}^{M} w^*_m h(a^*_m)
\]

This is equivalent to

\[
\int_{-\infty}^{\infty} f(x) dx \approx \sum_{m=1}^{M} w^*_m \exp \left\{ (a^*_m)^2 \right\} f(a^*_m)
\]

where the \( w^*_m \) denote the quadrature weights and the \( a^*_m \) denote the quadrature abscissas. The log likelihood, \( L \), is the sum of the logs of the panel-level likelihoods \( l_i \).

The default approximation of the log likelihood is by adaptive Gauss–Hermite quadrature, which approximates the panel-level likelihood with

\[
l_i \approx \sqrt{2\hat{\sigma}_i} \sum_{m=1}^{M} w^*_m \exp \left\{ (a^*_m)^2 \right\} g(y_{it}, x_{it}, \sqrt{2\hat{\sigma}_i}a^*_m + \hat{\mu}_i)
\]

where \( \hat{\sigma}_i \) and \( \hat{\mu}_i \) are the adaptive parameters for panel \( i \). Therefore, with the definition of \( g(y_{it}, x_{it}, \nu_i) \), the total log likelihood is approximated by
\[
L \approx \sum_{i=1}^{n} w_i \log \left[ \sqrt{2} \hat{\sigma}_i \sum_{m=1}^{M} w_m^* \exp\left\{ (a_m^*)^2 \right\} \exp\left\{ -\left( \sqrt{2} \hat{\sigma}_i a_m^* + \hat{\mu}_i \right)^2 / 2\sigma^2_\nu \right\} \prod_{t=1}^{n_i} F(y_{it}, x_{it} \beta + \sqrt{2} \hat{\sigma}_i a_m^* + \hat{\mu}_i) \right]
\]

where \( w_i \) is the user-specified weight for panel \( i \); if no weights are specified, \( w_i = 1 \).

The default method of adaptive Gauss–Hermite quadrature is to calculate the posterior mean and variance and use those parameters for \( \hat{\mu}_i \) and \( \hat{\sigma}_i \) by following the method of Naylor and Smith (1982), further discussed in Skrondal and Rabe-Hesketh (2004). We start with \( \hat{\sigma}_{i,0} = 1 \) and \( \hat{\mu}_{i,0} = 0 \), and the posterior means and variances are updated in the \( k \)th iteration. That is, at the \( k \)th iteration of the optimization for \( l_i \), we use

\[
l_{i,k} \approx \sum_{m=1}^{M} \sqrt{2} \hat{\sigma}_{i,k-1} w_m^* \exp\left\{ (a_m^*)^2 \right\} g(y_{it}, x_{it}, \sqrt{2} \hat{\sigma}_{i,k-1} a_m^* + \hat{\mu}_{i,k-1})
\]

Letting

\[
\tau_{i,m,k-1} = \sqrt{2} \hat{\sigma}_{i,k-1} a_m^* + \hat{\mu}_{i,k-1}
\]

\[
\hat{\mu}_{i,k} = \sum_{m=1}^{M} (\tau_{i,m,k-1}) \sqrt{2} \hat{\sigma}_{i,k-1} w_m^* \exp\left\{ (a_m^*)^2 \right\} g(y_{it}, x_{it}, \tau_{i,m,k-1})
\]

and

\[
\hat{\sigma}_{i,k} = \sum_{m=1}^{M} (\tau_{i,m,k-1})^2 \sqrt{2} \hat{\sigma}_{i,k-1} w_m^* \exp\left\{ (a_m^*)^2 \right\} g(y_{it}, x_{it}, \tau_{i,m,k-1}) - (\hat{\mu}_{i,k})^2
\]

and this is repeated until \( \hat{\mu}_{i,k} \) and \( \hat{\sigma}_{i,k} \) have converged for this iteration of the maximization algorithm. This adaptation is applied on every iteration until the log-likelihood change from the preceding iteration is less than a relative difference of 1e–6; after this, the quadrature parameters are fixed.

The log likelihood can also be calculated by nonadaptive Gauss–Hermite quadrature, the int-method(ghermite) option, where \( \rho = \sigma^2_\nu / (\sigma^2_\nu + 1) \):

\[
L = \sum_{i=1}^{n} w_i \log \left\{ \Pr(y_{i1}, \ldots, y_{in_i} | x_{i1}, \ldots, x_{in_i}) \right\}
\]

\[
\approx \sum_{i=1}^{n} w_i \log \left[ \frac{1}{\sqrt{n}} \sum_{m=1}^{M} w_m^* \prod_{t=1}^{n_i} F\left\{ y_{it}, x_{it} \beta + a_m^* \left( \frac{2\rho}{1 - \rho} \right)^{1/2} \right\} \right]
\]

Both quadrature formulas require that the integrated function be well approximated by a polynomial of degree equal to the number of quadrature points. The number of periods (panel size) can affect whether

\[
\prod_{t=1}^{n_i} F(y_{it}, x_{it} \beta + \nu_i)
\]
is well approximated by a polynomial. As panel size and $\rho$ increase, the quadrature approximation can become less accurate. For large $\rho$, the random-effects model can also become unidentified. Adaptive quadrature gives better results for correlated data and large panels than nonadaptive quadrature; however, we recommend that you use the `quadchk` command (see \[ XT \quad \text{quadchk} \]) to verify the quadrature approximation used in this command, whichever approximation you choose.

**xtlogit, re and the robust VCE estimator**

Specifying `vce(robust)` or `vce(cluster clustvar)` causes the Huber/White/sandwich VCE estimator to be calculated for the coefficients estimated in this regression. See \[ P \quad \_\text{robust}, \] particularly \textit{Introduction} and \textit{Methods and formulas}. Wooldridge (2013) and Arellano (2003) discuss this application of the Huber/White/sandwich VCE estimator. As discussed by Wooldridge (2013), Stock and Watson (2008), and Arellano (2003), specifying `vce(robust)` is equivalent to specifying `vce(cluster panelvar)`, where `panelvar` is the variable that identifies the panels.

Clustering on the panel variable produces a consistent VCE estimator when the disturbances are not identically distributed over the panels or there is serial correlation in $\epsilon_{it}$.

The cluster–robust VCE estimator requires that there are many clusters and the disturbances are uncorrelated across the clusters. The panel variable must be nested within the cluster variable because of the within-panel correlation that is generally induced by the random-effects transform when there is heteroskedasticity or within-panel serial correlation in the idiosyncratic errors.

**References**


Also see

[XT] xtlogit postestimation — Postestimation tools for xtlogit
[XT] quadchk — Check sensitivity of quadrature approximation
[XT] xtcloglog — Random-effects and population-averaged cloglog models
[XT] xtgee — Fit population-averaged panel-data models by using GEE
[XT] xtprobit — Random-effects and population-averaged probit models
[XT] xtset — Declare data to be panel data
[ME] melogit — Multilevel mixed-effects logistic regression
[ME] meqrlogit — Multilevel mixed-effects logistic regression (QR decomposition)
[MI] estimation — Estimation commands for use with mi estimate
[R] clogit — Conditional (fixed-effects) logistic regression
[R] logistic — Logistic regression, reporting odds ratios
[R] logit — Logistic regression, reporting coefficients
[U] 20 Estimation and postestimation commands
Description

The following postestimation commands are available after xtlogit:

<table>
<thead>
<tr>
<th>Command</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>contrast</td>
<td>contrasts and ANOVA-style joint tests of estimates</td>
</tr>
<tr>
<td>estat ic1</td>
<td>Akaike’s and Schwarz’s Bayesian information criteria (AIC and BIC)</td>
</tr>
<tr>
<td>estat summarize</td>
<td>summary statistics for the estimation sample</td>
</tr>
<tr>
<td>estat vce</td>
<td>variance–covariance matrix of the estimators (VCE)</td>
</tr>
<tr>
<td>estimates</td>
<td>cataloging estimation results</td>
</tr>
<tr>
<td>forecast2</td>
<td>dynamic forecasts and simulations</td>
</tr>
<tr>
<td>hausman</td>
<td>Hausman’s specification test</td>
</tr>
<tr>
<td>lincom</td>
<td>point estimates, standard errors, testing, and inference for linear combinations of coefficients</td>
</tr>
<tr>
<td>lrtest</td>
<td>likelihood-ratio test</td>
</tr>
<tr>
<td>margins3</td>
<td>marginal means, predictive margins, marginal effects, and average marginal effects</td>
</tr>
<tr>
<td>marginsplot</td>
<td>graph the results from margins (profile plots, interaction plots, etc.)</td>
</tr>
<tr>
<td>nlcom</td>
<td>point estimates, standard errors, testing, and inference for nonlinear combinations of coefficients</td>
</tr>
<tr>
<td>predict</td>
<td>predictions, residuals, influence statistics, and other diagnostic measures</td>
</tr>
<tr>
<td>predictnl</td>
<td>point estimates, standard errors, testing, and inference for generalized predictions</td>
</tr>
<tr>
<td>pwcompare</td>
<td>pairwise comparisons of estimates</td>
</tr>
<tr>
<td>test</td>
<td>Wald tests of simple and composite linear hypotheses</td>
</tr>
<tr>
<td>testnl</td>
<td>Wald tests of nonlinear hypotheses</td>
</tr>
</tbody>
</table>

1 estat ic is not appropriate after xtlogit, pa.
2 forecast is not appropriate with mi estimation results or after xtlogit, fe.
3 The default prediction statistic for xtlogit, fe, pu1, cannot be correctly handled by margins; however, margins can be used after xtlogit, fe with the predict(pu0) option or the predict(xb) option.
Syntax for predict

Random-effects model

```
predict [type] newvar [if] [in] [ , RE_statistic nooffset]
```

Fixed-effects model

```
predict [type] newvar [if] [in] [ , FE_statistic nooffset]
```

Population-averaged model

```
predict [type] newvar [if] [in] [ , PA_statistic nooffset]
```

<table>
<thead>
<tr>
<th>RE_statistic</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Main</td>
<td></td>
</tr>
<tr>
<td>xb</td>
<td>linear prediction; the default</td>
</tr>
<tr>
<td>pu0</td>
<td>probability of a positive outcome assuming that the random effect is zero</td>
</tr>
<tr>
<td>stdp</td>
<td>standard error of the linear prediction</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>FE_statistic</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Main</td>
<td></td>
</tr>
<tr>
<td>pc1</td>
<td>predicted probability of a positive outcome conditional on one positive outcome within group; the default</td>
</tr>
<tr>
<td>pu0</td>
<td>probability of a positive outcome assuming that the fixed effect is zero</td>
</tr>
<tr>
<td>xb</td>
<td>linear prediction</td>
</tr>
<tr>
<td>stdp</td>
<td>standard error of the linear prediction</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>PA_statistic</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Main</td>
<td></td>
</tr>
<tr>
<td>mu</td>
<td>predicted probability of depvar; considers the offset()</td>
</tr>
<tr>
<td>rate</td>
<td>predicted probability of depvar</td>
</tr>
<tr>
<td>xb</td>
<td>linear prediction</td>
</tr>
<tr>
<td>stdp</td>
<td>standard error of the linear prediction</td>
</tr>
<tr>
<td>score</td>
<td>first derivative of the log likelihood with respect to x_j*β</td>
</tr>
</tbody>
</table>

These statistics are available both in and out of sample; type predict ... if e(sample) ... if wanted only for the estimation sample.

The predicted probability for the fixed-effects model is conditional on there being only one outcome per group. See [R] clogit for details.

Menu for predict

Statistics > Postestimation > Predictions, residuals, etc.
Options for predict

xb calculates the linear prediction. This is the default for the random-effects model.

pc1 calculates the predicted probability of a positive outcome conditional on one positive outcome within group. This is the default for the fixed-effects model.

mu and rate both calculate the predicted probability of depvar. mu takes into account the offset(), and rate ignores those adjustments. mu and rate are equivalent if you did not specify offset().

mu0 calculates the probability of a positive outcome, assuming that the fixed or random effect for that observation’s panel is zero (υ = 0). This may not be similar to the proportion of observed outcomes in the group.

stdp calculates the standard error of the linear prediction.

score calculates the equation-level score, \( u_j = \frac{\partial \ln L_j(x_j\beta)}{\partial (x_j\beta)} \).

nooffset is relevant only if you specified offset(varname) for xtlogit. This option modifies the calculations made by predict so that they ignore the offset variable; the linear prediction is treated as \( x_{it}\beta \) rather than \( x_{it}\beta + \text{offset}_{it} \).

Remarks and examples

Example 1

In example 1 of [XT] xtlogit, we fit a random-effects model of union status on the person’s age and level of schooling, whether she lived in an urban area, and whether she lived in the south. In fact, we included the full interaction between south and year to capture both the overall effect of residing in the south and a separate time-trend for southerners. To test whether residing in the south affects union status, we must determine whether 1.south and south#c.year are jointly significant. First, we refit our model, store the estimation results for later use, and use test to conduct a Wald test of the joint significance of those two variables’ parameters:

```
. use http://www.stata-press.com/data/r13/union
(NLS Women 14–24 in 1968)
. xtlogit union age grade not_smsa south##c.year
(output omitted)
. estimates store fullmodel
. test 1.south 1.south#c.year
   ( 1) [union]1.south = 0
   ( 2) [union]1.south#c.year = 0
      chi2(  2) =    143.93
      Prob > chi2 =    0.0000
```

The test statistic is clearly significant, so we reject the null hypothesis that the coefficients are jointly zero and conclude that living in the south does significantly affect union status.
We can also test our hypothesis with a likelihood-ratio test. Here we fit the model without `south##c.year` and then call `lrtest` to compare this restricted model to the full model:

```
. xtlogit union age grade not_smsa
   (output omitted)
. lrtest fullmodel .
```

```
Likelihood-ratio test          LR chi2(3) = 146.55
(Assumption: . nested in fullmodel) Prob > chi2 = 0.0000
```

These results confirm our finding that living in the south affects union status.

Also see

[XT] `xtlogit` — Fixed-effects, random-effects, and population-averaged logit models

[U] 20 Estimation and postestimation commands
**xtnbreg** — Fixed-effects, random-effects, & population-averaged negative binomial models

### Syntax

**Random-effects (RE) and conditional fixed-effects (FE) overdispersion models**

```
xtnbreg depvar [ indepvars ] [ if ] [ in ] [ weight ] , [ re | fe ] RE/FE_options
```

**Population-averaged (PA) model**

```
xtnbreg depvar [ indepvars ] [ if ] [ in ] [ weight ] , pa [ PA_options ]
```

#### RE/FE_options

<table>
<thead>
<tr>
<th>Model</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>noconstant</code></td>
<td>suppress constant term; not available with <code>fe</code></td>
</tr>
<tr>
<td><code>re</code></td>
<td>use random-effects estimator; the default</td>
</tr>
<tr>
<td><code>fe</code></td>
<td>use fixed-effects estimator</td>
</tr>
<tr>
<td><code>offset(varname)</code></td>
<td>include ln(varname) in model with coefficient constrained to 1</td>
</tr>
<tr>
<td><code>constraints(varname)</code></td>
<td>include varname in model with coefficient constrained to 1</td>
</tr>
<tr>
<td><code>collinear</code></td>
<td>apply specified linear constraints</td>
</tr>
<tr>
<td></td>
<td>keep collinear variables</td>
</tr>
</tbody>
</table>

#### SE

| `vce(vcetype)` | `vcetype` may be `oim`, `bootstrap`, or `jackknife` |

#### Reporting

| `level(#)`     | set confidence level; default is `level(95)`         |
| `irr`          | report incidence-rate ratios                          |
| `noskip`       | perform overall model test as a likelihood-ratio test|
| `nocnsreport`  | do not display constraints                            |
| `display_options` | control column formats, row spacing, line width, display of omitted variables and base and empty cells, and factor-variable labeling |

#### Maximization

| `maximize_options` | control the maximization process; seldom used        |
| `coeflegend`       | display legend instead of statistics                  |
### xtnbreg — Fixed-effects, random-effects, & population-averaged negative binomial models

#### PA_options

<table>
<thead>
<tr>
<th>Model</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>noconstant</td>
<td>suppress constant term</td>
</tr>
<tr>
<td>pa</td>
<td>use population-averaged estimator</td>
</tr>
<tr>
<td>exposure(varname)</td>
<td>include ln(varname) in model with coefficient constrained to 1</td>
</tr>
<tr>
<td>offset(varname)</td>
<td>include varname in model with coefficient constrained to 1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Correlation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>corr(correlation)</td>
<td>within-panel correlation structure</td>
</tr>
<tr>
<td>force</td>
<td>estimate even if observations unequally spaced in time</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SE/Robust</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>vce(vcetype)</td>
<td>vcetype may be conventional, robust, bootstrap, or jackknife</td>
</tr>
<tr>
<td>nmp</td>
<td>use divisor $N - P$ instead of the default $N$</td>
</tr>
<tr>
<td>scale(parm)</td>
<td>overrides the default scale parameter; parm may be x2, dev, phi, or #</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Reporting</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>level(#)</td>
<td>set confidence level; default is level(95)</td>
</tr>
<tr>
<td>irr</td>
<td>report incidence-rate ratios</td>
</tr>
<tr>
<td>display_options</td>
<td>control column formats, row spacing, line width, display of omitted</td>
</tr>
<tr>
<td></td>
<td>variables and base and empty cells, and factor-variable labeling</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Optimization</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>optimize_options</td>
<td>control the optimization process; seldom used</td>
</tr>
<tr>
<td>coeflegend</td>
<td>display legend instead of statistics</td>
</tr>
</tbody>
</table>

#### correlation

<table>
<thead>
<tr>
<th>Description</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>exchangeable</td>
<td>exchangeable</td>
</tr>
<tr>
<td>independent</td>
<td>independent</td>
</tr>
<tr>
<td>unstructured</td>
<td>unstructured</td>
</tr>
<tr>
<td>fixed matname</td>
<td>user-specified</td>
</tr>
<tr>
<td>ar #</td>
<td>autoregressive of order #</td>
</tr>
<tr>
<td>stationary #</td>
<td>stationary of order #</td>
</tr>
<tr>
<td>nonstationary #</td>
<td>nonstationary of order #</td>
</tr>
</tbody>
</table>

A panel variable must be specified. For xtnbreg, pa, correlation structures other than exchangeable and independent require that a time variable also be specified. Use xtset; see [XT] xtset.

indepvars may contain factor variables; see [U] 11.4.3 Factor variables.

depvar and indepvars may contain time-series operators; see [U] 11.4.4 Time-series varlists.

by, mi estimate, and statsby are allowed; see [U] 11.1.10 Prefix commands. fp is allowed for the random-effects and fixed-effects models.

vce(bootstrap) and vce(jackknife) are not allowed with the mi estimate prefix; see [MI] mi estimate.

iweights, fweights, and pweights are allowed for the population-averaged model, and iweights are allowed in the random-effects and fixed-effects models; see [U] 11.1.6 weight. Weights must be constant within panel.

coefflegend does not appear in the dialog box.

See [U] 20 Estimation and postestimation commands for more capabilities of estimation commands.
Description

`xtnbreg` fits random-effects overdispersion models, conditional fixed-effects overdispersion models, and population-averaged negative binomial models. Here “random effects” and “fixed effects” apply to the distribution of the dispersion parameter, not to the $x\beta$ term in the model. In the random-effects and fixed-effects overdispersion models, the dispersion is the same for all elements in the same group (that is, elements with the same value of the panel variable). In the random-effects model, the dispersion varies randomly from group to group, such that the inverse of one plus the dispersion follows a Beta($r, s$) distribution. In the fixed-effects model, the dispersion parameter in a group can take on any value, because a conditional likelihood is used in which the dispersion parameter drops out of the estimation.

By default, the population-averaged model is an equal-correlation model; `xtnbreg, pa` assumes `corr(exchangeable)`. See `[XT] xtgee` for information on how to fit other population-averaged models.

Options for RE/FE models

- **Model**
  - `noconstant`; see `[R] estimation options.
  - `re` requests the random-effects estimator, which is the default.
  - `fe` requests the conditional fixed-effects estimator.
  - `exposure(varname), offset(varname), constraints(constraints), collinear`; see `[R] estimation options.

- **SE**
  - `vce(vcetype)` specifies the type of standard error reported, which includes types that are derived from asymptotic theory (`oim`) and that use bootstrap or jackknife methods (`bootstrap, jackknife`); see `[XT] vce_options`.

- **Reporting**
  - `level(#)`; see `[R] estimation options.
  - `irr` reports exponentiated coefficients $e^b$ rather than coefficients $b$. For the negative binomial model, exponentiated coefficients have the interpretation of incidence-rate ratios.
  - `noskip`; see `[R] estimation options.
  - `nocnsreport`; see `[R] estimation options.
  - `display_options: nomissed, vsquish, noemptycells, baselevels, allbaselevels, nolabel, fwrap(#), fwrapon(style), cformat(\%fmt), pformat(\%fmt), sformat(\%fmt), and nolstretch`; see `[R] estimation options.`
maximize_options: difficult, technique(algorithm_spec), iterate(#), [no]log, trace, gradient, showstep, hessian, showtolerance, tolerance(#), ltolerance(#), nrtolerance(#), nonrtolerance, and from(init_specs); see [R] maximize. These options are seldom used.

The following option is available with xtnbreg but is not shown in the dialog box: coeflegend; see [R] estimation options.

Options for PA model

Model

noconstant; see [R] estimation options.

pa requests the population-averaged estimator.

exposure(varname), offset(varname); see [R] estimation options.

Correlation

corr(correlation) specifies the within-panel correlation structure; the default corresponds to the equal-correlation model, corr(exchangeable).

When you specify a correlation structure that requires a lag, you indicate the lag after the structure’s name with or without a blank; for example, corr(ar 1) or corr(ar1).

If you specify the fixed correlation structure, you specify the name of the matrix containing the assumed correlations following the word fixed, for example, corr(fixed myr).

force specifies that estimation be forced even though the time variable is not equally spaced. This is relevant only for correlation structures that require knowledge of the time variable. These correlation structures require that observations be equally spaced so that calculations based on lags correspond to a constant time change. If you specify a time variable indicating that observations are not equally spaced, the (time dependent) model will not be fit. If you also specify force, the model will be fit, and it will be assumed that the lags based on the data ordered by the time variable are appropriate.

SE/Robust

vce(vcetype) specifies the type of standard error reported, which includes types that are derived from asymptotic theory (conventional), that are robust to some kinds of misspecification (robust), and that use bootstrap or jackknife methods (bootstrap, jackknife); see [XT] vce_options.

vce(conventional), the default, uses the conventionally derived variance estimator for generalized least-squares regression.

nmp, scale(x2|dev|phi|#); see [XT] vce_options.

Reporting

level(#); see [R] estimation options.

irr reports exponentiated coefficients $e^b$ rather than coefficients $b$. For the negative binomial model, exponentiated coefficients have the interpretation of incidence-rate ratios.
display_options: noomitted, vsquish, noemptycells, baselevels, allbaselevels, nofvlabel, fvwrap(#), fvwrapon(style), cformat(\%fmt), pformat(\%fmt), sformat(\%fmt), and nolstretch; see [R] estimation options.

Optimization

Optimize_options control the iterative optimization process. These options are seldom used.

iterate(#) specifies the maximum number of iterations. When the number of iterations equals #, the optimization stops and presents the current results, even if convergence has not been reached. The default is iterate(100).

tolerance(#) specifies the tolerance for the coefficient vector. When the relative change in the coefficient vector from one iteration to the next is less than or equal to #, the optimization process is stopped. tolerance(1e-6) is the default.

nolog suppresses display of the iteration log.

trace specifies that the current estimates be printed at each iteration.

The following option is available with xtnbreg but is not shown in the dialog box: coeflegend; see [R] estimation options.

Remarks and examples

xtnbreg is a convenience command if you want the population-averaged model. Typing

.xtnbreg ..., ... pa exposure(time)

is equivalent to typing

.xtgee ..., ... family(nbinomial) link(log) corr(exchangeable) exposure(time)

See also [XT] xtgee for information about xtnbreg.

By default, or when re is specified, xtnbreg fits a maximum-likelihood random-effects overdispersion model.

Example 1

You have (fictional) data on injury “incidents” incurred among 20 airlines in each of 4 years. (Incidents range from major injuries to exceedingly minor ones.) The government agency in charge of regulating airlines has run an experimental safety training program, and, in each of the years, some airlines have participated and some have not. You now wish to analyze whether the “incident” rate is affected by the program. You choose to estimate using random-effects negative binomial regression, as the dispersion might vary across the airlines for unidentified airline-specific reasons. Your measure of exposure is passenger miles for each airline in each year.
. use http://www.stata-press.com/data/r13/airacc
. xtnbreg i_cnt inprog, exposure(pmiles) irr

Fitting negative binomial (constant dispersion) model:
Iteration 0: log likelihood = -293.57997
Iteration 1: log likelihood = -293.57997
(output omitted)

Fitting full model:
Iteration 0: log likelihood = -295.72633
Iteration 1: log likelihood = -270.49929 (not concave)
(output omitted)

Random-effects negative binomial regression

<table>
<thead>
<tr>
<th></th>
<th>Number of obs = 80</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group variable: airline</td>
<td>Number of groups = 20</td>
</tr>
<tr>
<td>Random effects u_i ~ Beta</td>
<td>Obs per group: min = 4</td>
</tr>
<tr>
<td></td>
<td>avg = 4.0</td>
</tr>
<tr>
<td></td>
<td>max = 4</td>
</tr>
</tbody>
</table>

Wald chi2(1) = 2.04
Log likelihood = -265.38202
Prob > chi2 = 0.1532

|          | IRR     | Std. Err. | z       | P>|z|   | [95% Conf. Interval] |
|----------|---------|-----------|---------|-------|---------------------|
| i_cnt    |         |           |         |       |                     |
| inprog   | .911673 | .0590277  | -1.43   | 0.153 | .8030206 1.035027   |
| _cons    | .0367524| .0407032  | -2.98   | 0.003 | .0041936  .3220983 |
| ln(pmiles)| 1       | (exposure)|         |       |                     |
| /ln_r    | 4.794991| .951781   |         |       | 2.929535 6.660448  |
| /ln_s    | 3.268052| .4709033  |         |       | 2.345098 4.191005  |

r          | 120.9033| 115.0735  | 18.71892| 780.9007|
s          | 26.26013| 12.36598 | 10.4343 | 66.08918|

Likelihood-ratio test vs. pooled: chibar2(01) = 19.03 Prob>=chibar2 = 0.000

In the output above, the /ln_r and /ln_s lines refer to ln(r) and ln(s), where the inverse of one plus the dispersion is assumed to follow a Beta(r, s) distribution. The output also includes a likelihood-ratio test, which compares the panel estimator with the pooled estimator (that is, a negative binomial estimator with constant dispersion).

You find that the incidence rate for accidents is not significantly different for participation in the program and that the panel estimator is significantly different from the pooled estimator.
We may alternatively fit a fixed-effects overdispersion model:

```
.xtnbreg i_cnt inprog, exposure(pmiles) irr fe nolog
```

Conditional FE negative binomial regression
Number of obs = 80
Group variable: airline
Number of groups = 20
Obs per group: min = 4
avg = 4.0
max = 4
Wald chi2(1) = 2.11
Log likelihood = -174.25143 Prob > chi2 = 0.1463

| i_cnt | IRR     | Std. Err. | z     | P>|z|   | [95% Conf. Interval] |
|-------|---------|-----------|-------|-------|---------------------|
| inprog| .9062669| .0613917  | -1.45 | 0.146 | .793587 1.034946    |
| _cons | .0329025| .0331262  | -3.39 | 0.001 | .0045734 .2367111  |

ln(pmiles) 1 (exposure)

---

**Example 2**

We rerun our previous example, but this time we fit a robust equal-correlation population-averaged model:

```
.xtnbreg i_cnt inprog, exposure(pmiles) irr vce(robust) pa
```

Iteration 1: tolerance = .02499392
Iteration 2: tolerance = .0000482
Iteration 3: tolerance = 2.929e-07
GEE population-averaged model
Number of obs = 80
Group variable: airline
Number of groups = 20
Link: log
Obs per group: min = 4
avg = 4.0
max = 4
Wald chi2(1) = 1.28
Scale parameter: 1
Prob > chi2 = 0.2571

(Std. Err. adjusted for clustering on airline)

| i_cnt | Semirobust | Std. Err. | z     | P>|z|   | [95% Conf. Interval] |
|-------|------------|-----------|-------|-------|---------------------|
| inprog| .927275    | .0617857  | -1.13 | 0.257 | .8137513 1.056636   |
| _cons | .0080211   | .0004117  | -94.02| 0.000 | .0072535 .00887    |

ln(pmiles) 1 (exposure)
We compare this with a pooled estimator with clustered robust-variance estimates:

```
    . nbreg i_cnt inprog, exposure(pmiles) irr vce(cluster airline)
```

**Fitting Poisson model:**

<table>
<thead>
<tr>
<th>Iteration</th>
<th>log pseudolikelihood</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>-293.57997</td>
</tr>
<tr>
<td>1</td>
<td>-293.57997</td>
</tr>
</tbody>
</table>

**Fitting constant-only model:**

<table>
<thead>
<tr>
<th>Iteration</th>
<th>log pseudolikelihood</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>-335.13615</td>
</tr>
<tr>
<td>1</td>
<td>-279.43327</td>
</tr>
<tr>
<td>2</td>
<td>-276.09296</td>
</tr>
<tr>
<td>3</td>
<td>-274.84036</td>
</tr>
<tr>
<td>4</td>
<td>-274.81076</td>
</tr>
<tr>
<td>5</td>
<td>-274.81075</td>
</tr>
</tbody>
</table>

**Fitting full model:**

<table>
<thead>
<tr>
<th>Iteration</th>
<th>log pseudolikelihood</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>-274.56985</td>
</tr>
<tr>
<td>1</td>
<td>-274.55077</td>
</tr>
<tr>
<td>2</td>
<td>-274.55077</td>
</tr>
</tbody>
</table>

Negative binomial regression

<table>
<thead>
<tr>
<th></th>
<th>Number of obs = 80</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dispersion</td>
<td>= mean</td>
</tr>
<tr>
<td>Log pseudolikelihood</td>
<td>= -274.55077</td>
</tr>
<tr>
<td>Wald chi2(1)</td>
<td>= 0.60</td>
</tr>
<tr>
<td>Prob &gt; chi2</td>
<td>= 0.4369</td>
</tr>
</tbody>
</table>

(Std. Err. adjusted for 20 clusters in airline)

<table>
<thead>
<tr>
<th></th>
<th>Robust</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>IRR</td>
</tr>
<tr>
<td></td>
<td>Std. Err.</td>
</tr>
<tr>
<td></td>
<td>z</td>
</tr>
<tr>
<td></td>
<td>P&gt;</td>
</tr>
<tr>
<td></td>
<td>[95% Conf. Interval]</td>
</tr>
</tbody>
</table>

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>i_cnt</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>IRR</td>
<td>Std. Err.</td>
<td>z</td>
<td>P&gt;</td>
</tr>
<tr>
<td>inprog</td>
<td>.9429015</td>
<td>.0713091</td>
<td>-0.78</td>
<td>0.437</td>
</tr>
<tr>
<td>_cons</td>
<td>.007956</td>
<td>.0004237</td>
<td>-90.77</td>
<td>0.000</td>
</tr>
<tr>
<td>ln(pmiles)</td>
<td>1 (exposure)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/lnalpha</td>
<td>-2.835089</td>
<td>.3351784</td>
<td>-3.492027</td>
<td>-2.178151</td>
</tr>
<tr>
<td>alpha</td>
<td>.0587133</td>
<td>.0196794</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Stored results

**xtnbreg, re** stores the following in **e()**:  

<table>
<thead>
<tr>
<th>Scalar</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>e(N)</td>
<td>number of observations</td>
</tr>
<tr>
<td>e(N_g)</td>
<td>number of groups</td>
</tr>
<tr>
<td>e(k)</td>
<td>number of parameters</td>
</tr>
<tr>
<td>e(k_eq)</td>
<td>number of equations in e(b)</td>
</tr>
<tr>
<td>e(k_eq_model)</td>
<td>number of equations in overall model test</td>
</tr>
<tr>
<td>e(k_dv)</td>
<td>number of dependent variables</td>
</tr>
<tr>
<td>e(df_m)</td>
<td>model degrees of freedom</td>
</tr>
<tr>
<td>e(ll)</td>
<td>log likelihood</td>
</tr>
<tr>
<td>e(ll_0)</td>
<td>log likelihood, constant-only model</td>
</tr>
<tr>
<td>e(ll_c)</td>
<td>log likelihood, comparison model</td>
</tr>
<tr>
<td>e(chi2)</td>
<td>$\chi^2$ for comparison test</td>
</tr>
<tr>
<td>e(chi2_c)</td>
<td>$\chi^2$ for comparison test corresponding to e(chi2)</td>
</tr>
<tr>
<td>e(g_min)</td>
<td>smallest group size</td>
</tr>
<tr>
<td>e(g_avg)</td>
<td>average group size</td>
</tr>
<tr>
<td>e(g_max)</td>
<td>largest group size</td>
</tr>
<tr>
<td>e(r)</td>
<td>value of $r$ in $\text{Beta}(r, s)$</td>
</tr>
<tr>
<td>e(s)</td>
<td>value of $s$ in $\text{Beta}(r, s)$</td>
</tr>
<tr>
<td>e(rank)</td>
<td>rank of e(V)</td>
</tr>
<tr>
<td>e(rank0)</td>
<td>rank of e(V) for constant-only model</td>
</tr>
<tr>
<td>e(ic)</td>
<td>number of iterations</td>
</tr>
<tr>
<td>e(rc)</td>
<td>return code</td>
</tr>
<tr>
<td>e(converged)</td>
<td>1 if converged, 0 otherwise</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Macro</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>e(cmd)</td>
<td>xtnbreg</td>
</tr>
<tr>
<td>e(cmdline)</td>
<td>command as typed</td>
</tr>
<tr>
<td>e(depvar)</td>
<td>name of dependent variable</td>
</tr>
<tr>
<td>e(ivar)</td>
<td>variable denoting groups</td>
</tr>
<tr>
<td>e(model)</td>
<td>re</td>
</tr>
<tr>
<td>e(wtype)</td>
<td>weight type</td>
</tr>
<tr>
<td>e(wexp)</td>
<td>weight expression</td>
</tr>
<tr>
<td>e(title)</td>
<td>title in estimation output</td>
</tr>
<tr>
<td>e(offset)</td>
<td>linear offset variable</td>
</tr>
<tr>
<td>e(chi2type)</td>
<td>Wald or LR: type of model $\chi^2$ test</td>
</tr>
<tr>
<td>e(chi2_ct)</td>
<td>Wald or LR: type of model $\chi^2$ test corresponding to e(chi2_c)</td>
</tr>
<tr>
<td>e(vcetype)</td>
<td>vcetype specified in vce()</td>
</tr>
<tr>
<td>e(technique)</td>
<td>title used to label Std. Err.</td>
</tr>
<tr>
<td>e(method)</td>
<td>estimation method</td>
</tr>
<tr>
<td>e(distrib)</td>
<td>Beta; the distribution of the random effect</td>
</tr>
<tr>
<td>e(opt)</td>
<td>type of optimization</td>
</tr>
<tr>
<td>e(which)</td>
<td>max or min: whether optimizer is to perform maximization or minimization</td>
</tr>
<tr>
<td>e(ml_method)</td>
<td>type of ml method</td>
</tr>
<tr>
<td>e(user)</td>
<td>name of likelihood-evaluator program</td>
</tr>
<tr>
<td>e(properties)</td>
<td>b V</td>
</tr>
<tr>
<td>e(predict)</td>
<td>program used to implement predict</td>
</tr>
<tr>
<td>e(asbalanced)</td>
<td>factor variables fvset as asbalanced</td>
</tr>
<tr>
<td>e(asobserved)</td>
<td>factor variables fvset as asobserved</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Matrix</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>e(b)</td>
<td>coefficient vector</td>
</tr>
<tr>
<td>e(Cns)</td>
<td>constraints matrix</td>
</tr>
<tr>
<td>e(ilog)</td>
<td>iteration log</td>
</tr>
<tr>
<td>e(gradient)</td>
<td>gradient vector</td>
</tr>
<tr>
<td>e(V)</td>
<td>variance–covariance matrix of the estimators</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Function</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>e(sample)</td>
<td>marks estimation sample</td>
</tr>
</tbody>
</table>
xtnbreg, fe stores the following in e():

Scalars
- e(N) number of observations
- e(N_g) number of groups
- e(k) number of parameters
- e(k_eq) number of equations in e(b)
- e(k_eq_model) number of equations in overall model test
- e(k_dv) number of dependent variables
- e(df_m) model degrees of freedom
- e(r2_p) pseudo $R^2$
- e(ll) log likelihood
- e(ll_0) log likelihood, constant-only model
- e(chi2) $\chi^2$
- e(g_min) smallest group size
- e(g_avg) average group size
- e(g_max) largest group size
- e(p) significance
- e(rank) rank of e(V)
- e(ic) number of iterations
- e(rc) return code
- e(converged) 1 if converged, 0 otherwise

Macros
- e(cmd) xtnbreg
- e(cmdline) command as typed
- e(depvar) name of dependent variable
- e(ivar) variable denoting groups
- e(model) fe
- e(wtype) weight type
- e(wexp) weight expression
- e(title) title in estimation output
- e(offset) linear offset variable
- e(chi2type) LR; type of model $\chi^2$ test
- e(vcetype) vcetype specified in vce()
- e(method) requested estimation method
- e(opt) type of optimization
- e(which) max or min; whether optimizer is to perform maximization or minimization
- e(ml_method) type of ml method
- e(user) name of likelihood-evaluator program
- e(technique) maximization technique
- e(properties) b V
- e(predict) program used to implement predict
- e(asbalanced) factor variables fvset as asbalanced
- e(asobserved) factor variables fvset as asobserved

Matrices
- e(b) coefficient vector
- e(Cns) constraints matrix
- e(ilog) iteration log
- e(gradients) gradient vector
- e(V) variance–covariance matrix of the estimators

Functions
- e(sample) marks estimation sample
xtnbreg, pa stores the following in e():

Scalars
- e(N) number of observations
- e(N_g) number of groups
- e(df_m) model degrees of freedom
- e(chi2) $\chi^2$
- e(p) significance
- e(df_pear) degrees of freedom for Pearson $\chi^2$
- e(chi2_dev) $\chi^2$ test of deviance
- e(chi2_dis) $\chi^2$ test of deviance dispersion
- e(deviance) deviance
- e(dispers) deviance dispersion
- e(phi) scale parameter
- e(g_min) smallest group size
- e(g_avg) average group size
- e(g_max) largest group size
- e(rank) rank of e(V)
- e(tol) target tolerance
- e(dif) achieved tolerance
- e(rc) return code

Macros
- e(cmd) xtgee
- e(cmd2) xtnbreg
- e(cmdline) command as typed
- e(depvar) name of dependent variable
- e(ivar) variable denoting groups
- e(tvar) variable denoting time within groups
- e(model) pa
- e(family) negative binomial($k=1$)
- e(link) log; link function
- e(corr) correlation structure
- e(scale) $\chi^2$, dev, phi, or #; scale parameter
- e(type) weight type
- e(wexp) weight expression
- e(offset) linear offset variable
- e(chi2type) Wald; type of model $\chi^2$ test
- e(vce) vcetype specified in vce()
- e(vcetype) title used to label Std. Err.
- e(nmp) nmp, if specified
- e(nbalpha) $\alpha$
- e(properties) b V
- e(predict) program used to implement predict
- e(marginsnotok) predictions disallowed by margins
- e(asbalanced) factor variables fvset as asbalanced
- e(asobserved) factor variables fvset as asobserved

Matrices
- e(b) coefficient vector
- e(R) estimated working correlation matrix
- e(V) variance–covariance matrix of the estimators
- e(V_modelbased) model-based variance

Functions
- e(sample) marks estimation sample
Methods and formulas

*xtnbreg*, pa reports the population-averaged results obtained by using *xtgee*, family(nbinomial) link(log) to obtain estimates. See [XT] *xtgee* for details on the methods and formulas.

For the random-effects and fixed-effects overdispersion models, let $y_{it}$ be the count for the $i$th observation in the $t$th group. We begin with the model $y_{it} | \gamma_{it} \sim \text{Poisson}(\gamma_{it})$, where $\gamma_{it} \sim \text{gamma}(\lambda_{it}, \delta_i)$ with $\lambda_{it} = \exp(x_{it}\beta + \text{offset}_{it})$ and $\delta_i$ is the dispersion parameter. This yields the model

$$
\Pr(Y_{it} = y_{it} | x_{it}, \delta_i) = \frac{\Gamma(\lambda_{it} + y_{it})}{\Gamma(\lambda_{it})\Gamma(y_{it} + 1)} \left( \frac{1}{1 + \delta_i} \right) \gamma_{it}^{y_{it}} \delta_i^{\lambda_{it} - 1} \exp(-\gamma_{it})
$$

(See Hausman, Hall, and Griliches [1984, eq. 3.1, 922]; our $\delta$ is the inverse of their $\delta$.) Looking at within-panel effects only, we find that this specification yields a negative binomial model for the $i$th group with dispersion (variance divided by the mean) equal to $1 + \delta_i$, that is, constant dispersion within group. This parameterization of the negative binomial model differs from the default parameterization of *nbreg*, which has dispersion equal to $1 + \alpha \exp(x\beta + \text{offset})$; see [R] *nbreg*.

For a random-effects overdispersion model, we allow $\delta_i$ to vary randomly across groups; namely, we assume that $1/(1 + \delta_i) \sim \text{Beta}(r, s)$. The joint probability of the counts for the $i$th group is

$$
\Pr(Y_{i1} = y_{i1}, \ldots, Y_{in_i} = y_{in_i} | X_i) = \int_0^\infty \prod_{t=1}^{n_i} \Pr(Y_{it} = y_{it} | x_{it}, \delta_i) f(\delta_i) \, d\delta_i
$$

$$
= \frac{\Gamma(r + s)\Gamma(r + \sum_{t=1}^{n_i} \lambda_{it})\Gamma(s + \sum_{t=1}^{n_i} y_{it})}{\Gamma(r)\Gamma(s)\Gamma(r + s + \sum_{t=1}^{n_i} \lambda_{it} + \sum_{t=1}^{n_i} y_{it})} \prod_{t=1}^{n_i} \frac{\Gamma(\lambda_{it} + y_{it})}{\Gamma(\lambda_{it})\Gamma(y_{it} + 1)}
$$

for $X_i = (x_{i1}, \ldots, x_{in_i})$ and where $f$ is the probability density function for $\delta_i$. The resulting log likelihood is

$$
\ln L = \sum_{i=1}^n w_i \left[ \ln \Gamma(r + s) + \ln \Gamma(r + \sum_{k=1}^{n_i} \lambda_{ik}) + \ln \Gamma(s + \sum_{k=1}^{n_i} y_{ik}) - \ln \Gamma(r) - \ln \Gamma(s) \right.
$$

$$
- \ln \Gamma(r + s + \sum_{k=1}^{n_i} \lambda_{ik} + \sum_{k=1}^{n_i} y_{ik}) + \sum_{t=1}^{n_i} \left\{ \ln \Gamma(\lambda_{it} + y_{it}) - \ln \Gamma(\lambda_{it}) - \ln \Gamma(y_{it} + 1) \right\}
$$

where $\lambda_{it} = \exp(x_{it}\beta + \text{offset}_{it})$ and $w_i$ is the weight for the $i$th group (Hausman, Hall, and Griliches 1984, eq. 3.5, 927).

For the fixed-effects overdispersion model, we condition the joint probability of the counts for each group on the sum of the counts for the group (that is, the observed $\sum_{t=1}^{n_i} y_{it}$). This yields

$$
\Pr(Y_{i1} = y_{i1}, \ldots, Y_{in_i} = y_{in_i} | X_i, \sum_{t=1}^{n_i} Y_{it} = \sum_{t=1}^{n_i} y_{it})
$$

$$
= \frac{\Gamma(\sum_{t=1}^{n_i} \lambda_{it})\Gamma(\sum_{t=1}^{n_i} y_{it} + 1)}{\Gamma(\sum_{t=1}^{n_i} \lambda_{it} + \sum_{t=1}^{n_i} y_{it})} \prod_{t=1}^{n_i} \frac{\Gamma(\lambda_{it} + y_{it})}{\Gamma(\lambda_{it})\Gamma(y_{it} + 1)}
$$
The conditional log likelihood is

\[
\ln L = \sum_{i=1}^{n} w_i \left[ \ln \Gamma \left( \sum_{t=1}^{n_i} \lambda_{it} \right) + \ln \Gamma \left( \sum_{t=1}^{n_i} y_{it} + 1 \right) - \ln \Gamma \left( \sum_{t=1}^{n_i} \lambda_{it} + \sum_{t=1}^{n_i} y_{it} \right) \right]
\]

\[
+ \sum_{t=1}^{n_i} \left\{ \ln \Gamma (\lambda_{it} + y_{it}) - \ln \Gamma (\lambda_{it}) - \ln \Gamma (y_{it} + 1) \right\}
\]

See Hausman, Hall, and Griliches (1984) for a more thorough development of the random-effects and fixed-effects models. Also see Cameron and Trivedi (2013) for a good textbook treatment of this model.

References


Also see

[XT] xtnbreg postestimation — Postestimation tools for xtnbreg

[XT] xtgee — Fit population-averaged panel-data models by using GEE

[XT] xtpoisson — Fixed-effects, random-effects, and population-averaged Poisson models

[XT] xtset — Declare data to be panel data

[ME] membreg — Multilevel mixed-effects negative binomial regression

[MI] estimation — Estimation commands for use with mi estimate

[R] nbreg — Negative binomial regression

[U] 20 Estimation and postestimation commands
Description

The following postestimation commands are available after `xtnbreg`:

<table>
<thead>
<tr>
<th>Command</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>contrast</td>
<td>contrasts and ANOVA-style joint tests of estimates</td>
</tr>
<tr>
<td>estat ic^1</td>
<td>Akaike’s and Schwarz’s Bayesian information criteria (AIC and BIC)</td>
</tr>
<tr>
<td>estat summarize</td>
<td>summary statistics for the estimation sample</td>
</tr>
<tr>
<td>estat vce</td>
<td>variance–covariance matrix of the estimators (VCE)</td>
</tr>
<tr>
<td>estimates</td>
<td>cataloging estimation results</td>
</tr>
<tr>
<td>forecast^2</td>
<td>dynamic forecasts and simulations</td>
</tr>
<tr>
<td>hausman</td>
<td>Hausman’s specification test</td>
</tr>
<tr>
<td>lincom</td>
<td>point estimates, standard errors, testing, and inference for linear</td>
</tr>
<tr>
<td></td>
<td>combinations of coefficients</td>
</tr>
<tr>
<td>lrtest</td>
<td>likelihood-ratio test</td>
</tr>
<tr>
<td>margins</td>
<td>marginal means, predictive margins, marginal effects, and average</td>
</tr>
<tr>
<td></td>
<td>marginal effects</td>
</tr>
<tr>
<td>marginsplot</td>
<td>graph the results from margins (profile plots, interaction plots, etc.)</td>
</tr>
<tr>
<td>nlcom</td>
<td>point estimates, standard errors, testing, and inference for nonlinear</td>
</tr>
<tr>
<td></td>
<td>combinations of coefficients</td>
</tr>
<tr>
<td>predict</td>
<td>predictions, residuals, influence statistics, and other diagnostic measures</td>
</tr>
<tr>
<td>predictnl</td>
<td>point estimates, standard errors, testing, and inference for generalized</td>
</tr>
<tr>
<td></td>
<td>predictions</td>
</tr>
<tr>
<td>pwcompare</td>
<td>pairwise comparisons of estimates</td>
</tr>
<tr>
<td>test</td>
<td>Wald tests of simple and composite linear hypotheses</td>
</tr>
<tr>
<td>testnl</td>
<td>Wald tests of nonlinear hypotheses</td>
</tr>
</tbody>
</table>

^1 `estat ic` is not appropriate after `xtnbreg, pa`.

^2 `forecast` is not appropriate with `mi` estimation results.

Syntax for predict

**Random-effects (RE) and conditional fixed-effects (FE) overdispersion models**

```
predict [type] newvar [if] [in] [, RE/FE_statistic nooffset]
```

**Population-averaged (PA) model**

```
predict [type] newvar [if] [in] [, PA_statistic nooffset]
```
### RE/FE_statistic

<table>
<thead>
<tr>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Main</strong></td>
</tr>
<tr>
<td>xb</td>
</tr>
<tr>
<td>stdp</td>
</tr>
<tr>
<td>nu0</td>
</tr>
<tr>
<td>iru0</td>
</tr>
<tr>
<td>pr0(n)</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>pr0(a,b)</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

### PA_statistic

<table>
<thead>
<tr>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Main</strong></td>
</tr>
<tr>
<td>mu</td>
</tr>
<tr>
<td>rate</td>
</tr>
<tr>
<td>xb</td>
</tr>
<tr>
<td>stdp</td>
</tr>
<tr>
<td>score</td>
</tr>
</tbody>
</table>

These statistics are available both in and out of sample; type `predict ... if e(sample) ...` if wanted only for the estimation sample.

### Menu for predict

Statistics > Postestimation > Predictions, residuals, etc.

### Options for predict

- **xb** calculates the linear prediction. This is the default for the random-effects and fixed-effects models.
- **mu** and **rate** both calculate the predicted number of events. **mu** takes into account the `offset()`, and **rate** ignores those adjustments. **mu** and **rate** are equivalent if you did not specify `offset()`. **mu** is the default for the population-averaged model.
- **stdp** calculates the standard error of the linear prediction.
- **nu0** calculates the predicted number of events, assuming a zero random or fixed effect.
- **iru0** calculates the predicted incidence rate, assuming a zero random or fixed effect.
- **pr0(n)** calculates the probability $Pr(y_j = n)$ assuming the random effect is zero, where $n$ is a nonnegative integer that may be specified as a number or a variable (only allowed after `xtnbreg, re`).
- **pr0(a,b)** calculates the probability $Pr(a \leq y_j \leq b)$ assuming the random effect is zero, where $a$ and $b$ are nonnegative integers that may be specified as numbers or variables (only allowed after `xtnbreg, re`);
$b$ missing ($b \geq .$) means $+\infty$; 
\texttt{pr0(20,.)} calculates $Pr(y_j \geq 20)$; 
\texttt{pr0(20,b)} calculates $Pr(y_j \geq 20)$ in observations for which $b \geq .$ and calculates $Pr(20 \leq y_j \leq b)$ elsewhere.

\texttt{pr0(.b)} produces a syntax error. A missing value in an observation on the variable $a$ causes a missing value in that observation for \texttt{pr0(a,b)}.

\texttt{score} calculates the equation-level score, $u_j = \partial \ln L_j(x_j\beta) / \partial (x_j\beta)$.

\texttt{nooffset} is relevant only if you specified \texttt{offset(varname)} for \texttt{xtnbreg}. It modifies the calculations made by \texttt{predict} so that they ignore the offset variable; the linear prediction is treated as $x_{it}/\beta$ rather than $x_{it}/\beta + \text{offset}_{it}$.

**Methods and formulas**

The probabilities calculated using the \texttt{pr0(n)} option are the probability $Pr(y_{it} = n)$ for a RE model assuming the random effect is zero. A negative binomial model is an overdispersed Poisson model, and the nominal overdispersion can be calculated as $\delta = s/(r - 1)$, where $r$ and $s$ are as given in the estimation results. Define $\mu_{it} = \exp(x_{it}/\beta + \text{offset}_{it})$. Then the probabilities in \texttt{pr0(n)} are calculated as the probability that $y_{it} = n$, where $y_{it}$ has a negative binomial distribution with mean $\delta \mu_{it}$ and variance $\delta(1 + \delta)\mu_{it}$.

**Also see**

[XT] \texttt{xtnbreg} — Fixed-effects, random-effects, & population-averaged negative binomial models

[U] 20 Estimation and postestimation commands
Syntax

xtologit  depvar  [  indepvars  ]  [  if  ]  [  in  ]  [  ,  options  ]

options  Description

Model

offset(varname)  include  varname  in  model  with  coefficient  constrained  to  1
constraints(constraints)  apply  specified  linear  constraints
collinear  keep  collinear  variables

SE/Robust

vce(vcetype)  vcetype  may  be  oim,  robust,  cluster  clustvar,  bootstrap,  or  jackknife

Reporting

level(#)  set  confidence  level;  default  is  level(95)
or
nokip  report  odds  ratios
noxnsreport  perform  overall  model  test  as  a  likelihood-ratio  test
display_options  do  not  display  constraints

column  formats,  row  spacing,  line  width,  display  of  omitted
variables  and  base  and  empty  cells,  and  factor-variable  labeling

Integration

intmethod(intmethod)  integration  method;  intmethod  may  be  mvaghermite  (the  default)  or  ghermite
intpoints(#)  use  #  quadrature  points;  default  is  intpoints(12)

Maximization

maximize_options  control  the  maximization  process;  seldom  used
startgrid(numlist)  improve  starting  value  of  the  random-intercept  parameter  by

performing  a  grid  search

nodisplay  suppress  display  of  header  and  coefficients
coeflegend  display  legend  instead  of  statistics

A  panel  variable  must  be  specified;  see  [XT]  xtset.

indepvars  may  contain  factor  variables;  see  [U]  11.4.3  Factor  variables.
depvar  and  indepvars  may  contain  time-series  operators;  see  [U]  11.4.4  Time-series  varlists.

by,  fp,  and  statsby  are  allowed;  see  [U]  11.1.10  Prefix  commands.

startgrid(),  nodisplay,  and  coeflegend  do  not  appear  in  the  dialog  box.

See  [U]  20  Estimation  and  postestimation  commands  for  more  capabilities  of  estimation  commands.
Menu

Statistics > Longitudinal/panel data > Ordinal outcomes > Logistic regression (RE)

Description

*xtologit* fits random-effects ordered logistic models. The actual values taken on by the dependent variable are irrelevant, although larger values are assumed to correspond to “higher” outcomes. The conditional distribution of the dependent variable given the random effects is assumed to be multinomial with success probability determined by the logistic cumulative distribution function.

Options

**Model**

offset(*varname*), constraints(*constraints*), collinear; see [*R*] estimation options.

**SE/Robust**

*vce(vcetype)* specifies the type of standard error reported, which includes types that are derived from asymptotic theory (*oim*), that are robust to some kinds of misspecification (*robust*), that allow for intragroup correlation (*cluster clustvar*), and that use bootstrap or jackknife methods (*bootstrap, jackknife*); see [*XT*] *vce* options.

Specifying *vce(robust)* is equivalent to specifying *vce(cluster panelvar)*; see *xtologit* and the robust VCE estimator in Methods and formulas.

**Reporting**

level(*#*); see [*R*] estimation options.

or reports the estimated coefficients transformed to odds ratios, that is, $e^b$ rather than $b$. Standard errors and confidence intervals are similarly transformed. This option affects how results are displayed, not how they are estimated. or may be specified at estimation or when replaying previously estimated results.

noskip, nocnsreport; see [*R*] estimation options.

*display_options*: noomitted, vsquish, noemptycells, baselevels, allbaselevels,nofvlabel, fwrap(*#*), fwraption(*style*), cformat(*fmt*), pformat(*fmt*), sformat(*fmt*), and nolstretch; see [*R*] estimation options.

**Integration**

*intmethod(*intmethod*), intpoints(*#*); see [*R*] estimation options.

**Maximization**

*maximize_options*: difficult, technique(*algorithm_spec*), iterate(*#*), [no] log, trace, gradient, showstep, hessian, showtolerance, tolerance(*#*), itolerance(*#*), nrtolerance(*#*), nonrtolerance, and from(*initspecs*); see [*R*] maximize. These options are seldom used.
The following options are available with `xtologit` but are not shown in the dialog box:

```
startgrid(numlist) performs a grid search to improve the starting value of the random-intercept parameter. No grid search is performed by default unless the starting value is found to not be feasible; in this case, `xtologit` runs `startgrid(0.1 1 10)` and chooses the value that works best. You may already be using a default form of `startgrid()` without knowing it. If you see `xtologit` displaying Grid node 1, Grid node 2, ... following Grid node 0 in the iteration log, that is `xtologit` doing a default search because the original starting value was not feasible.
```

`nodisplay` is for programmers. It suppresses the display of the header and the coefficients.

`coeflegend`; see [R] estimation options.

Remarks and examples

`xtologit` fits random-effects ordered logistic models. Ordered logistic models are used to estimate relationships between an ordinal dependent variable and a set of independent variables. An ordinal variable is a variable that is categorical and ordered, for instance, “poor”, “good”, and “excellent”, which might indicate a person’s current health status or the repair record of a car. If there are only two outcomes, see [XT] `xtlogit`, [XT] `xtprobit`, and [XT] `xtcloglog`. This entry is concerned only with more than two outcomes.

Example 1

We use the data from the “Television, School, and Family Smoking Prevention and Cessation Project” (Flay et al. 1988; Rabe-Hesketh and Skrondal 2012, chap. 11), where schools were randomly assigned into one of four groups defined by two treatment variables. Students within each school are nested in classes, and classes are nested in schools. In this example, we ignore the variability of classes within schools; see example 2 of [ME] `meologit` for a model that incorporates the additional class-level variance component. The dependent variable is the tobacco and health knowledge score (`thk`) collapsed into four ordered categories. We regress the outcome on the treatment variables and their interaction and control for the pretreatment score.
. use http://www.stata-press.com/data/r13/tvsfpors
. xtset school
    panel variable: school (unbalanced)
. xtologit thk prethk cc##tv

Fitting comparison model:
Iteration 0:  log likelihood = -2212.775
Iteration 1:  log likelihood = -2125.509
Iteration 2:  log likelihood = -2125.103
Iteration 3:  log likelihood = -2125.103

Refining starting values:
Grid node 0:  log likelihood = -2136.243

Fitting full model:
Iteration 0:  log likelihood = -2136.243  (not concave)
Iteration 1:  log likelihood = -2120.258
Iteration 2:  log likelihood = -2119.757
Iteration 3:  log likelihood = -2119.743
Iteration 4:  log likelihood = -2119.743

Random-effects ordered logistic regression
Number of obs = 1600
Group variable: school  Number of groups = 28
Random effects u_i ~ Gaussian
Obs per group:
    min = 18
    avg = 57.1
    max = 137
Integration method: mvaghermite  Integration points = 12

Wald chi2(4) = 128.1
Log likelihood = -2119.743  Prob > chi2 = 0.0000

| Coef.  | Std. Err. | z    | P>|z| [95% Conf. Interval] |
|--------|-----------|------|--------|------------------------|
| thk    |           |      |        |                        |
| prethk | 0.403291  | 0.03886 | 10.38 | 0.000 0.327125 0.4794534 |
| 1.cc   | 0.923790  | 0.20407 | 4.53  | 0.000 0.5238127 1.323768 |
| 1.tv   | 0.274993  | 0.19774 | 1.39  | 0.164 -0.1125744 0.6625618 |
| cc#tv  | -0.465926 | 0.2845963 | -1.64 | 0.102 -1.023724 0.918728 |
| /cut1  | -0.088449 | 0.164106 | -0.54 | 0.590 -0.4100916 0.233193 |
| /cut2  | 1.153364  | 0.165616 | 6.96  | 0.000 0.8287625 1.477965 |
| /cut3  | 2.33195   | 0.1734199 | 13.45 | 0.000 1.992053 2.671846 |
| /sigma2_u | 0.0735112 | 0.0383106 | 0.0264695 0.2041551 |

LR test vs. ologit regression:  chibar2(01) = 10.72  Prob>chibar2 = 0.0005

The estimation table reports the parameter estimates, the estimated cutpoints \((\kappa_1, \kappa_2, \kappa_3)\), and the estimated panel-level variance component labeled \(\sigma^2_u\). The parameter estimates can be interpreted just as the output from a standard ordered logistic regression would be interpreted; see \[R\] ologit. For example, we find that students with higher preintervention scores tend to have higher postintervention scores.

Underneath the parameter estimates and the cutpoints, the table shows the estimated variance component. The estimate of \(\sigma^2_u\) is 0.074 with standard error 0.038. The reported likelihood-ratio test shows that there is enough variability between schools to favor a random-effects ordered logistic regression over a standard ordered logistic regression.
## Technical note

The random-effects model is calculated using quadrature, which is an approximation whose accuracy depends partially on the number of integration points used. We can use the `quadchk` command to see if changing the number of integration points affects the results. If the results change, the quadrature approximation is not accurate given the number of integration points. Try increasing the number of integration points using the `intpoints()` option and run `quadchk` again. Do not attempt to interpret the results of estimates when the coefficients reported by `quadchk` differ substantially. See [XT] `quadchk` for details and [XT] `xtprobit` for an example.

Because the `xtologit` likelihood function is calculated by Gauss–Hermite quadrature, on large problems the computations can be slow. Computation time is roughly proportional to the number of points used for the quadrature.

## Stored results

`xtologit` stores the following in `e()`:

### Scalars
- `e(N)`: number of observations
- `e(N_g)`: number of groups
- `e(k)`: number of parameters
- `e(k_aux)`: number of auxiliary parameters
- `e(k_eq)`: number of equations in `e(b)`
- `e(k_eq_model)`: number of equations in overall model test
- `e(k_dv)`: number of dependent variables
- `e(k_cat)`: number of categories
- `e(df_m)`: model degrees of freedom
- `e(ll)`: log likelihood
- `e(ll_0)`: log likelihood, constant-only model
- `e(ll_c)`: log likelihood, comparison model
- `e(chi2)`: \( \chi^2 \)
- `e(chi2_c)`: \( \chi^2 \) for comparison test
- `e(N_clust)`: number of clusters
- `e(sigma_u)`: panel-level standard deviation
- `e(n_quad)`: number of quadrature points
- `e(g_min)`: smallest group size
- `e(g_avg)`: average group size
- `e(g_max)`: largest group size
- `e(p)`: significance
- `e(rank)`: rank of `e(V)`
- `e(rank0)`: rank of `e(V)` for constant-only model
- `e(ic)`: number of iterations
- `e(rc)`: return code
- `e(converged)`: 1 if converged, 0 otherwise

### Macros
- `e(cmd)`: `xtologit`
- `e(cmdline)`: command as typed
- `e(depvar)`: name of dependent variable
- `e(covariates)`: list of covariates
- `e(ivar)`: variable denoting groups
- `e(title)`: title in estimation output
- `e(clustvar)`: name of cluster variable
- `e(offset)`: linear offset variable
- `e(chi2type)`: Wald or LR; type of model \( \chi^2 \) test
- `e(vce)`: vcetype specified in `vce()`
- `e(vcetype)`: title used to label Std. Err.
xtologit — Random-effects ordered logistic models

**Integration Method**
e(intmethod) integration method

**Type of Optimization**
e(opt) type of optimization

**Distribution of the Random Effect**
e(distrib) Gaussian; the distribution of the random effect

**Name of Likelihood-Evaluator Program**
e(user) name of likelihood-evaluator program

**Maximization Technique**
e(technique) maximization technique

**Type of ml Method**
e(ml_method) type of ml method

**Predictions Allowed by margins**
e(marginsok) predictions allowed by margins

**Factor Variables**
e(asbalanced) factor variables fvset as asbalanced

**Factor Variables**
e(asobserved) factor variables fvset as asobserved

**Matrices**
e(b) coefficient vector

**Constraints Matrix**
e(Cns) constraints matrix

**Iteration Log**
e(ilog) iteration log

**Gradient Vector**
e(gradient) gradient vector

**Category Values**
e(cat) category values

**Variance–Covariance Matrix of the Estimators**
e(V) variance–covariance matrix of the estimators

**Model-Based Variance**
e(V_modelbased) model-based variance

**Functions**
e(sample) marks estimation sample

**Methods and Formulas**

**xtologit** fits via maximum likelihood the random-effects model

\[
Pr(y_{it} > k | \kappa, x_{it}, \nu_i) = H(x_{it}\beta + \nu_i - \kappa_k)
\]

for \(i = 1, \ldots, n\) panels, where \(t = 1, \ldots, n_i\), \(\nu_i\) are independent and identically distributed \(N(0, \sigma^2_\nu)\), and \(\kappa\) is a set of cutpoints \(\kappa_1, \kappa_2, \ldots, \kappa_{K-1}\), where \(K\) is the number of possible outcomes; and \(H(\cdot)\) is the logistic cumulative distribution function.

From the above, we can derive the probability of observing outcome \(k\) for response \(y_{it}\) as

\[
p_{itk} \equiv Pr(y_{it} = k | \kappa, x_{it}, \nu_i) = \Pr(\kappa_{k-1} < x_{it}\beta + \nu_i + \epsilon_{it} \leq \kappa_k)
\]

\[
= \Pr(\kappa_{k-1} - x_{it}\beta - \nu_i < \epsilon_{it} \leq \kappa_k - x_{it}\beta - \nu_i)
\]

\[
= H(\kappa_k - x_{it}\beta - \nu_i) - H(\kappa_{k-1} - x_{it}\beta - \nu_i)
\]

\[
= \frac{1}{1 + \exp(-\kappa_k + x_{it}\beta + \nu_i)} - \frac{1}{1 + \exp(-\kappa_{k-1} + x_{it}\beta + \nu_i)}
\]

where \(\kappa_0\) is taken as \(-\infty\) and \(\kappa_K\) is taken as \(+\infty\). Here \(x_{it}\) does not contain a constant term, because its effect is absorbed into the cutpoints.

We may also express this model in terms of a latent linear response, where observed ordinal responses \(y_{it}\) are generated from the latent continuous responses, such that

\[
y_{it}^* = x_{it}\beta + \nu_i + \epsilon_{it}
\]

and

\[
y_{it} = \begin{cases} 
1 & \text{if } y_{it}^* \leq \kappa_1 \\
2 & \text{if } \kappa_1 < y_{it}^* \leq \kappa_2 \\
\vdots \\
K & \text{if } \kappa_{K-1} < y_{it}^*
\end{cases}
\]

The errors \(\epsilon_{it}\) are distributed as logistic with mean zero and variance \(\pi^2/3\) and are independent of \(\nu_i\).
Given a set of panel-level random effects $\nu_i$, we can define the conditional distribution for response $y_{it}$ as

$$f(y_{it}, \kappa, \mathbf{x}_i \beta + \nu_i) = \prod_{k=1}^{K} p_{itk}^{I_k(y_{it})} = \exp \sum_{k=1}^{K} \left\{ I_k(y_{it}) \log(p_{itk}) \right\}$$

where

$$I_k(y_{it}) = \begin{cases} 1 & \text{if } y_{it} = k \\ 0 & \text{otherwise} \end{cases}$$

For panel $i$, $i = 1, \ldots, M$, the conditional distribution of $\mathbf{y}_i = (y_{i1}, \ldots, y_{in_i})'$ is

$$\prod_{t=1}^{n_i} f(y_{it}, \kappa, \mathbf{x}_i \beta + \nu_i)$$

and the panel-level likelihood $l_i$ is given by

$$l_i(\beta, \kappa, \sigma^2_{\nu}) = \int_{-\infty}^{\infty} \frac{e^{-\nu_i^2/2\sigma^2_{\nu}}}{\sqrt{2\pi\sigma_{\nu}}} \left\{ \prod_{t=1}^{n_i} f(y_{it}, \kappa, \mathbf{x}_i \beta + \nu_i) \right\} d\nu_i$$

$$\equiv \int_{-\infty}^{\infty} g(y_{it}, \kappa, \mathbf{x}_i, \nu_i) d\nu_i$$

This integral can be approximated with $M$-point Gauss–Hermite quadrature

$$\int_{-\infty}^{\infty} f(x) dx \approx \sum_{m=1}^{M} w^*_m h(a^*_m)$$

This is equivalent to

$$\int_{-\infty}^{\infty} f(x) dx \approx \sum_{m=1}^{M} w^*_m \exp \left\{ (a^*_m)^2 \right\} f(a^*_m)$$

where the $w^*_m$ denote the quadrature weights and the $a^*_m$ denote the quadrature abscissas. The log likelihood, $L$, is the sum of the logs of the panel-level likelihoods $l_i$.

The default approximation of the log likelihood is by mean–variance adaptive Gauss–Hermite quadrature, which approximates the panel-level likelihood with

$$l_i \approx \sqrt{2\hat{\sigma}_i^2} \sum_{m=1}^{M} w^*_m \exp \left\{ (a^*_m)^2 \right\} g(y_{it}, \kappa, \mathbf{x}_i, \sqrt{2\hat{\sigma}_i}a^*_m + \hat{\mu}_i)$$

where $\hat{\sigma}_i$ and $\hat{\mu}_i$ are the adaptive parameters for panel $i$. The method of calculating the posterior mean and variance and using those parameters for $\hat{\mu}_i$ and $\hat{\sigma}_i$ is described in detail in Naylor and Smith (1982) and Skrondal and Rabe-Hesketh (2004). We start with $\hat{\sigma}_{i,0} = 1$ and $\hat{\mu}_{i,0} = 0$, and the posterior means and variances are updated in the $j$th iteration. That is, at the $j$th iteration of the optimization for $l_i$, we use

$$l_{i,j} \approx \sum_{m=1}^{M} \sqrt{2\hat{\sigma}_{i,j-1}^2} w^*_m \exp \left\{ (a^*_m)^2 \right\} g(y_{it}, \kappa, \mathbf{x}_i, \sqrt{2\hat{\sigma}_{i,j-1}}a^*_m + \hat{\mu}_{i,j-1})$$
Letting
\[ \tau_{i,m,j-1} = \sqrt{2} \hat{\sigma}_{i,j-1} a_m^* + \hat{\mu}_{i,j-1} \]
\[ \hat{\mu}_{i,j} = \sum_{m=1}^{M} (\tau_{i,m,j-1}) \frac{\sqrt{2} \hat{\sigma}_{i,j-1} w_m^* \exp\{ (a_m^*)^2 \} g(y_{it}, \kappa, x_{it}, \tau_{i,m,j-1})}{l_{i,j}} \]
and
\[ \hat{\sigma}_{i,j} = \sum_{m=1}^{M} (\tau_{i,m,j-1})^2 \frac{\sqrt{2} \hat{\sigma}_{i,j-1} w_m^* \exp\{ (a_m^*)^2 \} g(y_{it}, \kappa, x_{it}, \tau_{i,m,j-1})}{l_{i,j}} - (\hat{\mu}_{i,j})^2 \]
This is repeated until \( \hat{\mu}_{i,j} \) and \( \hat{\sigma}_{i,j} \) have converged for this iteration of the maximization algorithm. This adaptation is applied on every iteration.

The log likelihood can also be calculated by nonadaptive Gauss–Hermite quadrature with the option intmethod(ghermite), where \( \rho = \frac{\sigma_v^2}{\sigma_v^2 + 1} \):

\[ L = \sum_{i=1}^{n} w_i \log \left\{ \Pr(y_{i1}, \ldots, y_{in_i} | \kappa, x_{i1}, \ldots, x_{in_i}) \right\} \]
\[ \approx \sum_{i=1}^{n} w_i \log \left[ \frac{1}{\sqrt{\pi}} \sum_{m=1}^{M} w_m^* \prod_{t=1}^{n_i} f \left\{ y_{it}, \kappa, x_{it} \beta + a_m^* \left( \frac{2 \rho}{1 - \rho} \right)^{1/2} \right\} \right] \]

Both quadrature formulas require that the integrated function be well approximated by a polynomial of degree equal to the number of quadrature points. The number of periods (panel size) can affect whether
\[ \prod_{t=1}^{n_i} f(y_{it}, \kappa, x_{it} \beta + \nu_i) \]
is well approximated by a polynomial. As panel size and \( \rho \) increase, the quadrature approximation can become less accurate. For large \( \rho \), the random-effects model can also become unidentified. Adaptive quadrature gives better results for correlated data and large panels than nonadaptive quadrature; however, we recommend that you use the quadchk command (see [XT] quadchk) to verify the quadrature approximation used in this command, whichever approximation you choose.

**xtologit and the robust VCE estimator**

Specifying vce(robust) or vce(cluster clustvar) causes the Huber/White/sandwich VCE estimator to be calculated for the coefficients estimated in this regression. See [P] _robust, particularly Introduction and Methods and formulas. Wooldridge (2013) and Arellano (2003) discuss this application of the Huber/White/sandwich VCE estimator. As discussed by Wooldridge (2013), Stock and Watson (2008), and Arellano (2003), specifying vce(robust) is equivalent to specifying vce(cluster panelvar), where panelvar is the variable that identifies the panels.

Clustering on the panel variable produces a consistent VCE estimator when the disturbances are not identically distributed over the panels or there is serial correlation in \( \epsilon_{it} \).

The cluster–robust VCE estimator requires that there are many clusters and the disturbances are uncorrelated across the clusters. The panel variable must be nested within the cluster variable because of the within-panel correlation that is generally induced by the random-effects transform when there is heteroskedasticity or within-panel serial correlation in the idiosyncratic errors.
References


Also see

[XT] xtologit postestimation — Postestimation tools for xtologit
[XT] quadchk — Check sensitivity of quadrature approximation
[XT] xtprobit — Random-effects ordered probit models
[XT] xtset — Declare data to be panel data
[ME] meologit — Multilevel mixed-effects ordered logistic regression
[R] logistic — Logistic regression, reporting odds ratios
[R] logit — Logistic regression, reporting coefficients
[U] 20 Estimation and postestimation commands
Description

The following postestimation commands are available after `xtologit`:

<table>
<thead>
<tr>
<th>Command</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>contrast</td>
<td>contrasts and ANOVA-style joint tests of estimates</td>
</tr>
<tr>
<td>estat ic</td>
<td>Akaike’s and Schwarz’s Bayesian information criteria (AIC and BIC)</td>
</tr>
<tr>
<td>estat summarize</td>
<td>summary statistics for the estimation sample</td>
</tr>
<tr>
<td>estat vce</td>
<td>variance–covariance matrix of the estimators (VCE)</td>
</tr>
<tr>
<td>estimates</td>
<td>cataloging estimation results</td>
</tr>
<tr>
<td>hausman</td>
<td>Hausman’s specification test</td>
</tr>
<tr>
<td>lincom</td>
<td>point estimates, standard errors, testing, and inference for linear combinations of coefficients</td>
</tr>
<tr>
<td>lrtest</td>
<td>likelihood-ratio test</td>
</tr>
<tr>
<td>margins</td>
<td>marginal means, predictive margins, marginal effects, and average marginal effects</td>
</tr>
<tr>
<td>marginsplot</td>
<td>graph the results from margins (profile plots, interaction plots, etc.)</td>
</tr>
<tr>
<td>nlcom</td>
<td>point estimates, standard errors, testing, and inference for nonlinear combinations of coefficients</td>
</tr>
<tr>
<td>predict</td>
<td>predictions, residuals, influence statistics, and other diagnostic measures</td>
</tr>
<tr>
<td>predictnl</td>
<td>point estimates, standard errors, testing, and inference for generalized predictions</td>
</tr>
<tr>
<td>pwcompare</td>
<td>pairwise comparisons of estimates</td>
</tr>
<tr>
<td>test</td>
<td>Wald tests of simple and composite linear hypotheses</td>
</tr>
<tr>
<td>testnl</td>
<td>Wald tests of nonlinear hypotheses</td>
</tr>
</tbody>
</table>
Syntax for predict

```
predict [type] { stub* | newvar | newvarlist } [ if ] [ in ] [, statistic
    outcome(outcome) nooffset]
```

<table>
<thead>
<tr>
<th>statistic</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Main</td>
<td></td>
</tr>
<tr>
<td>xb</td>
<td>linear prediction; the default</td>
</tr>
<tr>
<td>pu0</td>
<td>probability of the specified outcome (outcome()) assuming that the random effect is zero</td>
</tr>
<tr>
<td>stdp</td>
<td>standard error of the linear prediction</td>
</tr>
</tbody>
</table>

If you do not specify outcome(), pu0 (with one new variable specified) assumes outcome(#1).
You specify one or k new variables with pu0, where k is the number of outcomes.
You specify one new variable with xb and stdp.
These statistics are available both in and out of sample; type predict ... if e(sample) ... if wanted only for the estimation sample.

Menu for predict

Statistics > Postestimation > Predictions, residuals, etc.

Options for predict

```
xb, the default, calculates the linear prediction.

pu0 calculates predicted probabilities, assuming that the random effect for that observation’s panel is zero (ν = 0).
You specify one or k new variables, where k is the number of categories of the dependent variable.
If you specify the outcome() option, the probabilities will be predicted for the requested outcome only, in which case, you specify only one new variable. If you specify only one new variable and do not specify outcome(), outcome(1) is assumed.

stdp calculates the standard error of the linear prediction.
outcome(outcome) specifies the outcome for which the predicted probabilities are to be calculated.
outcome() should contain either one value of the dependent variable or one of #1, #2, ..., with #1 meaning the first category of the dependent variable, #2 meaning the second category, etc.
nooffset is relevant only if you specified offset(varname) for xtologit. This option modifies the calculations made by predict so that they ignore the offset variable; the linear prediction is treated as x_{it}\beta rather than x_{it}\beta + offset_{it}.
Remarks and examples

Example 1

In example 1 of [XT] `xtologit`, we modeled the tobacco and health knowledge score (`thk`)—coded 1, 2, 3, 4—among students as a function of two treatments (`cc` and `tv`) using a random-effects ordered logistic model. Here we refit the model, obtain the predicted probabilities for all 4 outcomes, and list the first 10 observations.

```
. use http://www.stata-press.com/data/r13/tvsfpors
. xtset school
    panel variable:  school (unbalanced)
. xtologit thk prethk cc##tv
    (output omitted)
. predict pr*, pu0
    (1 missing values generated)
. list thk pr1-pr4 in 1/10
   thk  pr1      pr2      pr3      pr4
    1.  3  .1395758  .2200463  .2863958  .3539821
    2.  4  .0675217  .1329124  .2484952  .5510707
    3.  3  .0675217  .1329124  .2484952  .5510707
    4.  4  .0977827  .1750507  .2765777  .4505889
    5.  4  .0977827  .1750507  .2765777  .4505889
    6.  3  .0675217  .1329124  .2484952  .5510707
    7.  2  .1395758  .2200463  .2863958  .3539821
    8.  4  .0675217  .1329124  .2484952  .5510707
    9.  4  .0461466  .09731  .2089935  .6475499
   10.  4  .0977827  .1750507  .2765777  .4505889
```

For each observation, our best guess for the predicted outcome is the one with the highest predicted probability. For example, for the very first observation in the table above, we would choose outcome 4 as the most likely to occur.

These predicted probabilities assume the random effects are zero for all panels. If you are interested in predicted probabilities that incorporate the random effects, see [ME] `meologit` and [ME] `meologit postestimation`.

Also see

[XT] `xtologit` — Random-effects ordered logistic models

[U] 20 Estimation and postestimation commands
xtoprobit — Random-effects ordered probit models

## Syntax

```
xtoprobit depvar [ indepvars ] [ if ] [ in ] [ , options ]
```

### options

**Model**

- `offset(varname)`: include `varname` in model with coefficient constrained to 1
- `constraints(constraints)`: apply specified linear constraints
- `collinear`: keep collinear variables

**SE/Robust**

- `vce(vcetype)`: `vcetype` may be `oim`, `robust`, `cluster clustvar`, `bootstrap`, or `jackknife`

**Reporting**

- `level(#)`: set confidence level; default is `level(95)`
- `noskip`: perform overall model test as a likelihood-ratio test
- `nocnsreport`: do not display constraints
- `display_options`: control column formats, row spacing, line width, display of omitted variables and base and empty cells, and factor-variable labeling

**Integration**

- `intmethod(intmethod)`: integration method; `intmethod` may be `mvaghermite` (the default) or `ghermite`
- `intpoints(#)`: use # quadrature points; default is `intpoints(12)`

**Maximization**

- `maximize_options`: control the maximization process; seldom used
- `startgrid(numlist)`: improve starting value of the random-intercept parameter by performing a grid search
- `nodisplay`: suppress display of header and coefficients
- `coeflegend`: display legend instead of statistics

### Notes

A panel variable must be specified; see [XT] xtset.

`indepvars` may contain factor variables; see [U] 11.4.3 Factor variables.

`depvar` and `indepvars` may contain time-series operators; see [U] 11.4.4 Time-series varlists.

by, fp, and statsby are allowed; see [U] 11.1.10 Prefix commands.

`startgrid()`, `nodisplay`, and `coeflegend` do not appear in the dialog box.

See [U] 20 Estimation and postestimation commands for more capabilities of estimation commands.

275
xtoprobit fits random-effects ordered probit models. The actual values taken on by the dependent variable are irrelevant, although larger values are assumed to correspond to “higher” outcomes. The conditional distribution of the dependent variable given the random effects is assumed to be multinomial, with success probability determined by the standard normal cumulative distribution function.

Options

Model

offset(varname), constraints(constraints), collinear; see [R] estimation options.

SE/Robust

cve(vcetype) specifies the type of standard error reported, which includes types that are derived from asymptotic theory (oim), that are robust to some kinds of misspecification (robust), that allow for intragroup correlation (cluster clustvar), and that use bootstrap or jackknife methods (bootstrap, jackknife); see [XT] vce_options.

Specifying vce(robust) is equivalent to specifying vce(cluster panelvar); see xtoprobit and the robust VCE estimator in Methods and formulas.

Reporting

level(#) ; see [R] estimation options.
noskip, nocnsreport; see [R] estimation options.

Display_options: noomitted, vsquish, noemptycells, baselevels, allbaselevels, nolabel, fvwrap(#), fvwrapon(style), cformat(%fmt), pformat(%fmt), sformat(%fmt), and nolstretch; see [R] estimation options.

Integration

intmethod(intmethod), intpoints(#); see [R] estimation options.

Maximization

maximize_options: difficult, technique(algorithm_spec), iterate(#), [no]log, trace, gradient, showstep, hessian, showtolerance, tolerance(#), ltolerance(#), nrtolerance(#), nonrtolerance, and from(init_specs); see [R] maximize. These options are seldom used.
The following options are available with *xtoprobit* but are not shown in the dialog box: startgrid( numlist ) performs a grid search to improve the starting value of the random-intercept parameter. No grid search is performed by default unless the starting value is found to not be feasible; in this case, *xtoprobit* runs startgrid(0.1 1 10) and chooses the value that works best. You may already be using a default form of startgrid() without knowing it. If you see *xtoprobit* displaying Grid node 1, Grid node 2, ... following Grid node 0 in the iteration log, that is *xtoprobit* doing a default search because the original starting value was not feasible.

nodisplay is for programmers. It suppresses the display of the header and the coefficients.

coeffield; see [R] estimation options.

Remarks and examples

*xtoprobit* fits random-effects ordered probit models. Ordered probit models are used to estimate relationships between an ordinal dependent variable and a set of independent variables. An ordinal variable is a variable that is categorical and ordered, for instance, “poor”, “good”, and “excellent”, which might indicate a person’s current health status or the repair record of a car. If there are only two outcomes, see [XT] *xtprobit*, [XT] *xtlogit*, and [XT] *xtcloglog*. This entry is concerned only with more than two outcomes.

Example 1

We use the data from the “Television, School, and Family Smoking Prevention and Cessation Project” (Flay et al. 1988; Rabe-Hesketh and Skrondal 2012, chap. 11), where schools were randomly assigned into one of four groups defined by two treatment variables. Students within each school are nested in classes, and classes are nested in schools. In this example, we ignore the variability of classes within schools; see example 2 of [ME] *meprobit* for a model that incorporates the additional class-level variance component. The dependent variable is the tobacco and health knowledge score (thk) collapsed into four ordered categories. We regress the outcome on the treatment variables and their interaction and control for the pretreatment score.
. use http://www.stata-press.com/data/r13/tvsfpors
. xtset school
   panel variable: school (unbalanced)
. xtoprobit thk prethk cc##tv

Fitting comparison model:
Iteration 0:  log likelihood = -2212.775
Iteration 1:  log likelihood = -2127.8111
Iteration 2:  log likelihood = -2127.7612
Iteration 3:  log likelihood = -2127.7612

Refining starting values:
Grid node 0:  log likelihood = -2149.7302

Fitting full model:
Iteration 0:  log likelihood = -2149.7302  (not concave)
Iteration 1:  log likelihood = -2129.6838  (not concave)
Iteration 2:  log likelihood = -2123.5143
Iteration 3:  log likelihood = -2122.2896
Iteration 4:  log likelihood = -2121.7949
Iteration 5:  log likelihood = -2121.7716
Iteration 6:  log likelihood = -2121.7715

Random-effects ordered probit regression
Number of obs = 1600
Group variable: school
Number of groups = 28
Random effects u_i ~ Gaussian
Obs per group:  min = 18
               avg = 57.1
               max = 137
Integration method: mvaghermite
Integration points = 12
Log likelihood = -2121.7715
Wald chi2(4)   = 128.05
Prob > chi2    = 0.0000

| thk     | Coef.   | Std. Err. | z     | P>|z|  | [95% Conf. Interval] |
|---------|---------|-----------|-------|------|----------------------|
| prethk  | 0.2369804 | 0.0227739 | 10.41 | 0.000 | 0.1923444          | 0.2816164 |
| 1.cc    | 0.5490957 | 0.1255108 | 4.37  | 0.000 | 0.303099           | 0.7950923 |
| 1.tv    | 0.1695405 | 0.1215889 | 1.39  | 0.163 | -0.0687693         | 0.4078504 |
| cc#tv   |         |           |       |      |                      |            |
| 1 1     | -0.2951837 | 0.1751969 | -1.68 | 0.092 | -0.6385634         | 0.0481959 |
| /cut1   | -0.0682011 | 0.1003374 | -0.68 | 0.497 | -0.2648587         | 0.1284565 |
| /cut2   | 0.67681    | 0.1008836 | 6.71  | 0.000 | 0.4790817          | 0.8745382 |
| /cut3   | 1.390649   | 0.1037494 | 13.40 | 0.000 | 1.187304           | 1.593995  |
| /sigma2_u| 0.0288527 | 0.0146201 |       |      | 0.0106874          | 0.0778937 |

LR test vs. oprobit regression:  chibar2(01) = 11.98  Prob>chibar2 = 0.0003

The estimation table reports the parameter estimates, the estimated cutpoints \((\kappa_1, \kappa_2, \kappa_3)\), and the estimated panel-level variance component labeled \(\text{sigma2\_u}\). The parameter estimates can be interpreted just as the output from a standard ordered probit regression would be interpreted; see \[R\] oprobit. For example, we find that students with higher preintervention scores tend to have higher postintervention scores.

Underneath the parameter estimates and the cutpoints, the table shows the estimated variance component. The estimate of \(\sigma^2_u\) is 0.029 with standard error 0.015. The reported likelihood-ratio test shows that there is enough variability between schools to favor a random-effects ordered probit regression over a standard ordered probit regression.
Technical note

The random-effects model is calculated using quadrature, which is an approximation whose accuracy depends partially on the number of integration points used. We can use the `quadchk` command to see if changing the number of integration points affects the results. If the results change, the quadrature approximation is not accurate given the number of integration points. Try increasing the number of integration points using the `intpoints()` option and run `quadchk` again. Do not attempt to interpret the results of estimates when the coefficients reported by `quadchk` differ substantially. See `[XT] quadchk` for details and `[XT] xtoprobit` for an example.

Because the `xtoprobit` likelihood function is calculated by Gauss–Hermite quadrature, on large problems the computations can be slow. Computation time is roughly proportional to the number of points used for the quadrature.

Stored results

`xtoprobit` stores the following in `e()`:

Scalars
- `e(N)` number of observations
- `e(N_g)` number of groups
- `e(k)` number of parameters
- `e(k_aux)` number of auxiliary parameters
- `e(k_eq)` number of equations in `e(b)`
- `e(k_eq_model)` number of equations in overall model test
- `e(k_dv)` number of dependent variables
- `e(k_cat)` number of categories
- `e(df_m)` model degrees of freedom
- `e(ll)` log likelihood
- `e(ll_0)` log likelihood, constant-only model
- `e(ll_c)` log likelihood, comparison model
- `e(chi2)` $\chi^2$
- `e(chi2_c)` $\chi^2$ for comparison test
- `e(N_clust)` number of clusters
- `e(sigma_u)` panel-level standard deviation
- `e(n_quad)` number of quadrature points
- `e(g_min)` smallest group size
- `e(g_avg)` average group size
- `e(g_max)` largest group size
- `e(p)` significance
- `e(rank)` rank of `e(V)`
- `e(rank0)` rank of `e(V)` for constant-only model
- `e(ic)` number of iterations
- `e(rc)` return code
- `e(converged)` 1 if converged, 0 otherwise

Macros
- `e(cmd)` `xtoprobit`
- `e(cmdline)` command as typed
- `e(depvar)` name of dependent variable
- `e(covariates)` list of covariates
- `e(ivar)` variable denoting groups
- `e(title)` title in estimation output
- `e(clustvar)` name of cluster variable
- `e(offset)` linear offset variable
- `e(chi2type)` Wald or LR; type of model $\chi^2$ test
- `e(vce)` `vcetype` specified in `vce()`
- `e(vcetype)` title used to label Std. Err.
xtoprobit — Random-effects ordered probit models

- **e(intmethod)**: Integration method
- **e(distrib)**: Gaussian; the distribution of the random effect
- **e(opt)**: Type of optimization
- **e(which)**: Max or min; whether optimizer is to perform maximization or minimization
- **e(ml_method)**: Type of ml method
- **e(user)**: Name of likelihood-evaluator program
- **e(technique)**: Maximization technique
- **e(properties)**: \( b \ V \)
- **e(predict)**: Program used to implement predict
- **e(marginsok)**: Predictions allowed by margins
- **e(asbalanced)**: Factor variables fvset as asbalanced
- **e(asobserved)**: Factor variables fvset as asobserved

Matrices
- **e(b)**: Coefficient vector
- **e(Cns)**: Constraints matrix
- **e(ilog)**: Iteration log
- **e(gradient)**: Gradient vector
- **e(cat)**: Category values
- **e(V)**: Variance–covariance matrix of the estimators
- **e(V_modelbased)**: Model-based variance

Functions
- **e(sample)**: Marks estimation sample

**Methods and formulas**

**xtoprobit** fits via maximum likelihood the random-effects model

\[
\Pr(y_{it} > k | \kappa, x_{it}, \nu_i) = \Phi(x_{it}\beta + \nu_i - \kappa_k)
\]

for \( i = 1, \ldots, n \) panels, where \( t = 1, \ldots, n_i, \nu_i \) are independent and identically distributed \( N(0, \sigma^2_\nu) \), and \( \kappa \) is a set of cutpoints \( \kappa_1, \kappa_2, \ldots, \kappa_{K-1} \), where \( K \) is the number of possible outcomes; and \( \Phi(\cdot) \) is the standard normal cumulative distribution function.

From the above, we can derive the probability of observing outcome \( k \) for response \( y_{it} \) as

\[
p_{itk} \equiv \Pr(y_{it} = k | \kappa, x_{it}, \nu_i) = \Pr(\kappa_{k-1} < x_{it}\beta + \nu_i + \epsilon_{it} \leq \kappa_k)
\]

\[
= \Pr(\kappa_{k-1} - x_{it}\beta - \nu_i < \epsilon_{it} \leq \kappa_k - x_{it}\beta - \nu_i)
\]

\[
= \Phi(\kappa_k - x_{it}\beta - \nu_i) - \Phi(\kappa_{k-1} - x_{it}\beta - \nu_i)
\]

where \( \kappa_0 \) is taken as \( -\infty \), and \( \kappa_K \) is taken as \( +\infty \). Here \( x_{it} \) does not contain a constant term, because its effect is absorbed into the cutpoints.

We may also express this model in terms of a latent linear response, where observed ordinal responses \( y_{it} \) are generated from the latent continuous responses, such that

\[
y_{it}^* = x_{it}\beta + \nu_i + \epsilon_{it}
\]

and

\[
y_{it} = \begin{cases} 
1 & \text{if } y_{it}^* \leq \kappa_1 \\
2 & \text{if } \kappa_1 < y_{it}^* \leq \kappa_2 \\
& \vdots \\
K & \text{if } \kappa_{K-1} < y_{it}^*
\end{cases}
\]

The errors \( \epsilon_{it} \) are distributed as standard normal with mean zero and variance one and are independent of \( \nu_i \).
Given a set of panel-level random effects \( \nu_i \), we can define the conditional distribution for response \( y_{it} \) as

\[
f(y_{it}, \kappa, x_{it}\beta + \nu_i) = \prod_{k=1}^{K} p_{itk}^{I_k(y_{it})} = \exp \sum_{k=1}^{K} \left\{ I_k(y_{it}) \log(p_{itk}) \right\}
\]

where

\[
I_k(y_{it}) = \begin{cases} 1 & \text{if } y_{it} = k \\ 0 & \text{otherwise} \end{cases}
\]

For panel \( i, i = 1, \ldots, M \), the conditional distribution of \( y_i = (y_{i1}, \ldots, y_{in_i})' \) is

\[
\prod_{t=1}^{n_i} f(y_{it}, \kappa, x_{it}\beta + \nu_i)
\]

and the panel-level likelihood \( l_i \) is given by

\[
l_i(\beta, \kappa, \sigma^2_\nu) = \int_{-\infty}^{\infty} \frac{e^{-\nu_i^2/2\sigma^2\nu}}{\sqrt{2\pi}\sigma_\nu} \left\{ \prod_{t=1}^{n_i} f(y_{it}, \kappa, x_{it}\beta + \nu_i) \right\} d\nu_i
\]

\[
\equiv \int_{-\infty}^{\infty} g(y_{it}, \kappa, x_{it}, \nu_i) d\nu_i
\]

This integral can be approximated with \( M \)-point Gauss–Hermite quadrature

\[
\int_{-\infty}^{\infty} f(x)dx \approx \sum_{m=1}^{M} w^*_m h(a^*_m)
\]

This is equivalent to

\[
\int_{-\infty}^{\infty} f(x)dx \approx \sum_{m=1}^{M} w^*_m \exp \left\{ (a^*_m)^2 \right\} f(a^*_m)
\]

where the \( w^*_m \) denote the quadrature weights and the \( a^*_m \) denote the quadrature abscissas. The log likelihood, \( L \), is the sum of the logs of the panel-level likelihoods \( l_i \).

The default approximation of the log likelihood is by mean–variance adaptive Gauss–Hermite quadrature, which approximates the panel-level likelihood with

\[
l_i \approx \sqrt{2\hat{\sigma}_i} \sum_{m=1}^{M} w^*_m \exp \left\{ (a^*_m)^2 \right\} g(y_{it}, \kappa, x_{it}, \sqrt{2\hat{\sigma}_i}a^*_m + \hat{\mu}_i)
\]

where \( \hat{\sigma}_i \) and \( \hat{\mu}_i \) are the adaptive parameters for panel \( i \). The method of calculating the posterior mean and variance and using those parameters for \( \hat{\mu}_i \) and \( \hat{\sigma}_i \) is described in detail in Naylor and Smith (1982) and Skrondal and Rabe-Hesketh (2004). We start with \( \hat{\sigma}_{i,0} = 1 \) and \( \hat{\mu}_{i,0} = 0 \), and the posterior means and variances are updated in the \( j \)th iteration. That is, at the \( j \)th iteration of the optimization for \( l_i \), we use

\[
l_{i,j} \approx \sum_{m=1}^{M} \sqrt{2\hat{\sigma}_{i,j-1}} w^*_m \exp \left\{ (a^*_m)^2 \right\} g(y_{it}, \kappa, x_{it}, \sqrt{2\hat{\sigma}_{i,j-1}a^*_m + \hat{\mu}_{i,j-1}})
\]
Letting

\[ \tau_{i,m,j-1} = \sqrt{2\hat{\sigma}_{i,j-1}} a^*_m + \hat{\mu}_{i,j-1} \]

\[ \hat{\mu}_{i,j} = \sum_{m=1}^{M} (\tau_{i,m,j-1}) \sqrt{2\hat{\sigma}_{i,j-1}} w^*_m \exp\{ (a^*_m)^2 \} g(y_{it}, \kappa, x_{it}, \tau_{i,m,j-1}) \]

and

\[ \hat{\sigma}_{i,j} = \sum_{m=1}^{M} (\tau_{i,m,j-1})^2 \sqrt{2\hat{\sigma}_{i,j-1}} w^*_m \exp\{ (a^*_m)^2 \} g(y_{it}, \kappa, x_{it}, \tau_{i,m,j-1}) - (\hat{\mu}_{i,j})^2 \]

This is repeated until \( \hat{\mu}_{i,j} \) and \( \hat{\sigma}_{i,j} \) have converged for this iteration of the maximization algorithm. This adaptation is applied on every iteration.

The log likelihood can also be calculated by nonadaptive Gauss–Hermite quadrature with the option \text{intmethod}(ghermite), where \( \rho = \sigma^2_\nu / (\sigma^2_\nu + 1) \):

\[ L = \sum_{i=1}^{n} w_i \log \left\{ \operatorname{Pr}(y_{i1}, \ldots, y_{in_i} \mid \kappa, x_{i1}, \ldots, x_{in_i}) \right\} \]

\[ \approx \sum_{i=1}^{n} w_i \log \left[ \frac{1}{\sqrt{\pi}} \sum_{m=1}^{M} w^*_m \prod_{t=1}^{n_i} f \left\{ y_{it}, \kappa, x_{it} \beta + a^*_m \left( \frac{2\rho}{1-\rho} \right)^{1/2} \right\} \right] \]

Both quadrature formulas require that the integrated function be well approximated by a polynomial of degree equal to the number of quadrature points. The number of periods (panel size) can affect whether

\[ \prod_{t=1}^{n_i} f(y_{it}, \kappa, x_{it} \beta + \nu_i) \]

is well approximated by a polynomial. As panel size and \( \rho \) increase, the quadrature approximation can become less accurate. For large \( \rho \), the random-effects model can also become unidentified. Adaptive quadrature gives better results for correlated data and large panels than nonadaptive quadrature; however, we recommend that you use the \text{quadchk} command (see \text{[XT] quadchk}) to verify the quadrature approximation used in this command, whichever approximation you choose.

### xtoprobit and the robust VCE estimator

Specifying \text{vce(robust)} or \text{vce(cluster clustvar)} causes the Huber/White/sandwich VCE estimator to be calculated for the coefficients estimated in this regression. See \text{[P] \_robust}, particularly \text{Introduction} and \text{Methods and formulas}. Wooldridge (2013) and Arellano (2003) discuss this application of the Huber/White/sandwich VCE estimator. As discussed by Wooldridge (2013), Stock and Watson (2008), and Arellano (2003), specifying \text{vce(robust)} is equivalent to specifying \text{vce(cluster panelvar)}, where \text{panelvar} is the variable that identifies the panels.

Clustering on the panel variable produces a consistent VCE estimator when the disturbances are not identically distributed over the panels or there is serial correlation in \( \epsilon_{it} \).

The cluster–robust VCE estimator requires that there are many clusters and the disturbances are uncorrelated across the clusters. The panel variable must be nested within the cluster variable because of the within-panel correlation that is generally induced by the random-effects transform when there is heteroskedasticity or within-panel serial correlation in the idiosyncratic errors.
References


Also see

[XT] `xtoprobit postestimation` — Postestimation tools for xtoprobit

[XT] `quadchk` — Check sensitivity of quadrature approximation

[XT] `xtologit` — Random-effects ordered logistic models

[XT] `xtset` — Declare data to be panel data

[ME] `meoprobit` — Multilevel mixed-effects ordered probit regression

[R] `probit` — Probit regression

[U] `20 Estimation and postestimation commands`
Description

The following postestimation commands are available after `xtprobit`:

<table>
<thead>
<tr>
<th>Command</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>contrast</td>
<td>contrasts and ANOVA-style joint tests of estimates</td>
</tr>
<tr>
<td>estat ic</td>
<td>Akaike’s and Schwarz’s Bayesian information criteria (AIC and BIC)</td>
</tr>
<tr>
<td>estat summarize</td>
<td>summary statistics for the estimation sample</td>
</tr>
<tr>
<td>estat vce</td>
<td>variance–covariance matrix of the estimators (VCE)</td>
</tr>
<tr>
<td>estimates</td>
<td>cataloging estimation results</td>
</tr>
<tr>
<td>hausman</td>
<td>Hausman’s specification test</td>
</tr>
<tr>
<td>lincom</td>
<td>point estimates, standard errors, testing, and inference for linear combinations of coefficients</td>
</tr>
<tr>
<td>lrtest</td>
<td>likelihood-ratio test</td>
</tr>
<tr>
<td>margins</td>
<td>marginal means, predictive margins, marginal effects, and average marginal effects</td>
</tr>
<tr>
<td>marginsplot</td>
<td>graph the results from margins (profile plots, interaction plots, etc.)</td>
</tr>
<tr>
<td>nlcom</td>
<td>point estimates, standard errors, testing, and inference for nonlinear combinations of coefficients</td>
</tr>
<tr>
<td>predict</td>
<td>predictions, residuals, influence statistics, and other diagnostic measures</td>
</tr>
<tr>
<td>predictnl</td>
<td>point estimates, standard errors, testing, and inference for generalized predictions</td>
</tr>
<tr>
<td>pwcompare</td>
<td>pairwise comparisons of estimates</td>
</tr>
<tr>
<td>test</td>
<td>Wald tests of simple and composite linear hypotheses</td>
</tr>
<tr>
<td>testnl</td>
<td>Wald tests of nonlinear hypotheses</td>
</tr>
</tbody>
</table>
Syntax for predict

\[ \text{predict} \ [\text{type}] \ \{\ stub* | new|newvarlist \} \ [\text{if}] \ [\text{in}] \ [, \ statistic \ outcome(outcome) \ nooffset] \]

<table>
<thead>
<tr>
<th>statistic</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Main</td>
<td></td>
</tr>
<tr>
<td>xb</td>
<td>linear prediction; the default</td>
</tr>
<tr>
<td>pu0</td>
<td>probability of the specified outcome (outcome()) assuming that the random effect is zero</td>
</tr>
<tr>
<td>stdp</td>
<td>standard error of the linear prediction</td>
</tr>
</tbody>
</table>

If you do not specify outcome(), pu0 (with one new variable specified) assumes outcome(#1).
You specify one or \( k \) new variables with pu0, where \( k \) is the number of outcomes.
You specify one new variable with xb and stdp.
These statistics are available both in and out of sample; type \text{predict} ... if e(sample) ... if wanted only for the estimation sample.

Menu for predict

Statistics > Postestimation > Predictions, residuals, etc.

Options for predict

Main

\[ \text{xb}, \text{the default, calculates the linear prediction.} \]
\[ \text{pu0} \text{ calculates predicted probabilities, assuming that the random effect for that observation’s panel is zero (} \nu = 0) \].

You specify one or \( k \) new variables, where \( k \) is the number of categories of the dependent variable.
If you specify the outcome() option, the probabilities will be predicted for the requested outcome only, in which case, you specify only one new variable. If you specify only one new variable and do not specify outcome(), outcome(1) is assumed.

\[ \text{stdp} \text{ calculates the standard error of the linear prediction.} \]
\[ \text{outcome(outcome)} \text{ specifies the outcome for which the predicted probabilities are to be calculated.} \]
\[ \text{outcome()} \text{ should contain either one value of the dependent variable or one of } \#1, \#2, \ldots, \text{ with } \#1 \text{ meaning the first category of the dependent variable, } \#2 \text{ meaning the second category, etc.} \]

\[ \text{nooffset} \text{ is relevant only if you specified offset(varname) for xtoprobit. This option modifies the calculations made by predict so that they ignore the offset variable; the linear prediction is treated as } x_{it} \beta \text{ rather than } x_{it} \beta + \text{offset}_{it}. \]
Remarks and examples

Example 1

In example 1 of [XT] xtoprobit, we modeled the tobacco and health knowledge score (thk)—coded 1, 2, 3, 4—among students as a function of two treatments (cc and tv) using a random-effects ordered probit model. Here we refit the model, obtain the predicted probabilities for all 4 outcomes, and list the first 10 observations.

```
. use http://www.stata-press.com/data/r13/tvsfpors
. xtset school
   panel variable: school (unbalanced)
. xtoprobit thk prethk cc##tv
   (output omitted)
. predict pr*, pu0
   (1 missing values generated)
. list thk pr1-pr4 in 1/10
  thk  pr1      pr2      pr3      pr4
 1.   3  .1375798  .2269989  .2788329  .3565884
 2.   4  .0587658  .1472831  .2515963  .5423548
 3.   3  .0587658  .1472831  .2515963  .5423548
 4.   4  .0920497  .1878205  .2720888  .4480409
 5.   4  .0920497  .1878205  .2720888  .4480409
 6.   3  .0587658  .1472831  .2515963  .5423548
 7.   2  .1375798  .2269989  .2788329  .3565884
 8.   4  .0587658  .1472831  .2515963  .5423548
 9.   4  .0357571  .1094559  .2204553  .6343318
10.   4  .0920497  .1878205  .2720888  .4480409
```

For each observation, our best guess for the predicted outcome is the one with the highest predicted probability. For example, for the very first observation in the table above, we would choose outcome 4 as the most likely to occur.

These predicted probabilities assume the random effects are zero for all panels. If you are interested in predicted probabilities that incorporate the random effects, see [ME] meoprobit and [ME] meoprobit postestimation.

Also see

[XT] xtoprobit — Random-effects ordered probit models

[U] 20 Estimation and postestimation commands
Title

**xtcse** — Linear regression with panel-corrected standard errors

## Syntax

```
xtcse depvar [indepvars] [if] [in] [weight] [ , options ]
```

### options

<table>
<thead>
<tr>
<th>Model</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>noconstant</td>
<td>suppress constant term</td>
</tr>
<tr>
<td>correlation(independent)</td>
<td>use independent autocorrelation structure</td>
</tr>
<tr>
<td>correlation(ar1)</td>
<td>use AR1 autocorrelation structure</td>
</tr>
<tr>
<td>correlation(par1)</td>
<td>use panel-specific AR1 autocorrelation structure</td>
</tr>
<tr>
<td>rho(type(calc))</td>
<td>specify method to compute autocorrelation parameter; seldom used</td>
</tr>
<tr>
<td>npl</td>
<td>weight panel-specific autocorrelations by panel sizes</td>
</tr>
<tr>
<td>hetonly</td>
<td>assume panel-level heteroskedastic errors</td>
</tr>
<tr>
<td>independent</td>
<td>assume independent errors across panels</td>
</tr>
</tbody>
</table>

### by/in

**casewise**

include only observations with complete cases

**pairwise**

include all available observations with nonmissing pairs

### SE

**nmk**

normalize standard errors by \( N - k \) instead of \( N \)

### Reporting

**level(#)**

set confidence level; default is `level(95)`

**detail**

report list of gaps in time series

**display_options**

control column formats, row spacing, line width, display of omitted variables and base and empty cells, and factor-variable labeling

**coeflegend**

display legend instead of statistics

A panel variable and a time variable must be specified; use `xtset`; see [XT] `xtset`.

`indepvars` may contain factor variables; see [U] 11.4.3 Factor variables.

`depvar` and `indepvars` may contain time-series operators; see [U] 11.4.4 Time-series varlists.

`by` and `statsby` are allowed; see [U] 11.1.10 Prefix commands.

`iweights` and `aweights` are allowed; see [U] 11.1.6 weight.

`coeflegend` does not appear in the dialog box.

See [U] 20 Estimation and postestimation commands for more capabilities of estimation commands.
Menu

Statistics > Longitudinal/panel data > Contemporaneous correlation > Regression with panel-corrected standard errors (PCSE)

Description

_xtpcse_ calculates panel-corrected standard error (PCSE) estimates for linear cross-sectional time-series models where the parameters are estimated by either OLS or Prais–Winsten regression. When computing the standard errors and the variance–covariance estimates, _xtpcse_ assumes that the disturbances are, by default, heteroskedastic and contemporaneously correlated across panels.

See [XT] _xtgls_ for the generalized least-squares estimator for these models.

Options

**Model**

noconstant; see [R] estimation options.

correlation(_corr_) specifies the form of assumed autocorrelation within panels.

  correlation(independent), the default, specifies that there is no autocorrelation.

  correlation(ar1) specifies that, within panels, there is first-order autocorrelation AR(1) and that the coefficient of the AR(1) process is common to all the panels.

  correlation(psar1) specifies that, within panels, there is first-order autocorrelation and that the coefficient of the AR(1) process is specific to each panel. _psar1_ stands for panel-specific AR(1).

rhotype(_calc_) specifies the method to be used to calculate the autocorrelation parameter. Allowed strings for _calc_ are

  regress regression using lags; the default

  freg regression using leads

  tscorr time-series autocorrelation calculation

  dw Durbin–Watson calculation

All above methods are consistent and asymptotically equivalent; this is a rarely used option.

np1 specifies that the panel-specific autocorrelations be weighted by $T_i$ rather than by the default $T_i - 1$ when estimating a common $\rho$ for all panels, where $T_i$ is the number of observations in panel $i$. This option has an effect only when panels are unbalanced and the correlation(ar1) option is specified.

hetonly and independent specify alternative forms for the assumed covariance of the disturbances across the panels. If neither is specified, the disturbances are assumed to be heteroskedastic (each panel has its own variance) and contemporaneously correlated across the panels (each pair of panels has its own covariance). This is the standard PCSE model.

hetonly specifies that the disturbances are assumed to be panel-level heteroskedastic only with no contemporaneous correlation across panels.

independent specifies that the disturbances are assumed to be independent across panels; that is, there is one disturbance variance common to all observations.
casewise and pairwise specify how missing observations in unbalanced panels are to be treated when estimating the interpanel covariance matrix of the disturbances. The default is casewise selection.

casewise specifies that the entire covariance matrix be computed only on the observations (periods) that are available for all panels. If an observation has missing data, all observations of that period are excluded when estimating the covariance matrix of disturbances. Specifying casewise ensures that the estimated covariance matrix will be of full rank and will be positive definite.

pairwise specifies that, for each element in the covariance matrix, all available observations (periods) that are common to the two panels contributing to the covariance be used to compute the covariance.

The casewise and pairwise options have an effect only when the panels are unbalanced and neither hetonly nor independent is specified.

SE

nmk specifies that standard errors be normalized by \( N - k \), where \( k \) is the number of parameters estimated, rather than \( N \), the number of observations. Different authors have used one or the other normalization. Greene (2012, 280) remarks that whether a degree-of-freedom correction improves the small-sample properties is an open question.

level(#); see [R] estimation options.

detail specifies that a detailed list of any gaps in the series be reported.

display_options: noomitted, vsquish, noemptycells, baselevels, allbaselevels, nofvlabel, fvwrap(#), fvwrapon(style), cformat(%fmt), pformat(%fmt), sformat(%fmt), and nolstretch; see [R] estimation options.

The following option is available with xtpcse but is not shown in the dialog box:

coefflegend; see [R] estimation options.

Remarks and examples

\texttt{xtpcse} is an alternative to feasible generalized least squares (FGLS)—see [XT] \texttt{xtgls}—for fitting linear cross-sectional time-series models when the disturbances are not assumed to be independent and identically distributed (i.i.d.). Instead, the disturbances are assumed to be either heteroskedastic across panels or heteroskedastic and contemporaneously correlated across panels. The disturbances may also be assumed to be autocorrelated within panel, and the autocorrelation parameter may be constant across panels or different for each panel.

We can write such models as

\[ y_{it} = x_{it}\beta + \epsilon_{it} \]

where \( i = 1, \ldots, m \) is the number of units (or panels); \( t = 1, \ldots, T_i \); \( T_i \) is the number of periods in panel \( i \); and \( \epsilon_{it} \) is a disturbance that may be autocorrelated along \( t \) or contemporaneously correlated across \( i \).
This model can also be written panel by panel as

\[
\begin{bmatrix}
y_1 \\
y_2 \\
\vdots \\
y_m
\end{bmatrix} =
\begin{bmatrix}
X_1 \\
X_2 \\
\vdots \\
X_m
\end{bmatrix} \beta +
\begin{bmatrix}
\epsilon_1 \\
\epsilon_2 \\
\vdots \\
\epsilon_m
\end{bmatrix}
\]

For a model with heteroskedastic disturbances and contemporaneous correlation but with no autocorrelation, the disturbance covariance matrix is assumed to be

\[
E[\epsilon\epsilon'] = \Omega =
\begin{bmatrix}
\sigma_{11}I_{T_i} & \sigma_{12}I_{T_i} & \cdots & \sigma_{1m}I_{T_i} \\
\sigma_{21}I_{T_i} & \sigma_{22}I_{T_i} & \cdots & \sigma_{2m}I_{T_i} \\
\vdots & \vdots & \ddots & \vdots \\
\sigma_{m1}I_{T_i} & \sigma_{m2}I_{T_i} & \cdots & \sigma_{mm}I_{T_i}
\end{bmatrix}
\]

where \( \sigma_{ii} \) is the variance of the disturbances for panel \( i \), \( \sigma_{ij} \) is the covariance of the disturbances between panel \( i \) and panel \( j \) when the panels’ periods are matched, and \( I \) is a \( T_i \) by \( T_i \) identity matrix with balanced panels. The panels need not be balanced for xtpcse, but the expression for the covariance of the disturbances will be more general if they are unbalanced.

This could also be written as

\[
E[\epsilon\epsilon'] = \Sigma \otimes I_{T_i \times T_i}
\]

where \( \Sigma \) is the panel-by-panel covariance matrix and \( I \) is an identity matrix.

See [XT] xtgls for a full taxonomy and description of possible disturbance covariance structures.

xtpcse and xtgls follow two different estimation schemes for this family of models. xtpcse produces OLS estimates of the parameters when no autocorrelation is specified, or Prais–Winsten (see [TS] prais) estimates when autocorrelation is specified. If autocorrelation is specified, the estimates of the parameters are conditional on the estimates of the autocorrelation parameter(s). The estimate of the variance–covariance matrix of the parameters is asymptotically efficient under the assumed covariance structure of the disturbances and uses the FGLS estimate of the disturbance covariance matrix; see Kmenta (1997, 121).

xtgls produces full FGLS parameter and variance–covariance estimates. These estimates are conditional on the estimates of the disturbance covariance matrix and are conditional on any autocorrelation parameters that are estimated; see Kmenta (1997), Greene (2012), Davidson and MacKinnon (1993), or Judge et al. (1985).

Both estimators are consistent, as long as the conditional mean \( (x_{it}\beta) \) is correctly specified. If the assumed covariance structure is correct, FGLS estimates produced by xtgls are more efficient. Beck and Katz (1995) have shown, however, that the full FGLS variance–covariance estimates are typically unacceptably optimistic (anticonservative) when used with the type of data analyzed by most social scientists—10–20 panels with 10–40 periods per panel. They show that the OLS or Prais–Winsten estimates with PCSEs have coverage probabilities that are closer to nominal.

Because the covariance matrix elements, \( \sigma_{ij} \), are estimated from panels \( i \) and \( j \), using those observations that have common time periods, estimators for this model achieve their asymptotic behavior as the \( T_i \)'s approach infinity. In contrast, the random- and fixed-effects estimators assume a different model and are asymptotic in the number of panels \( m \); see [XT] xtreg for details of the random- and fixed-effects estimators.
Although \texttt{xtpcse} allows other disturbance covariance structures, the term PCSE, as used in the literature, refers specifically to models that are both heteroskedastic and contemporaneously correlated across panels, with or without autocorrelation.

\section*{Example 1: Controlling for heteroskedasticity and cross-panel correlation}

Grunfeld and Griliches (1960) analyzed a company’s current-year gross investment (\texttt{invest}) as determined by the company’s prior year market value (\texttt{mvalue}) and the prior year’s value of the company’s plant and equipment (\texttt{kstock}). The dataset includes 10 companies over 20 years, from 1935 through 1954, and is a classic dataset for demonstrating cross-sectional time-series analysis. Greene (2012, 1112) reproduces the dataset.

To use \texttt{xtpcse}, the data must be organized in “long form”; that is, each observation must represent a record for a specific company at a specific time; see \texttt{[D] reshape}. In the Grunfeld data, \texttt{company} is a categorical variable identifying the company, and \texttt{year} is a variable recording the year. Here are the first few records:

\begin{verbatim}
. use http://www.stata-press.com/data/r13/grunfeld
. list in 1/5
     +-----------------+------------------+-+-----------------+
     | company | year | invest | mvalue | kstock | time |
     +-----------------+------------------+-+-----------------+
     1.         1 | 1935 | 317.6  | 3078.5 | 2.8    | 1    |
     2.         1 | 1936 | 391.8  | 4661.7 | 52.6   | 2    |
     3.         1 | 1937 | 410.6  | 5387.1 | 156.9  | 3    |
     4.         1 | 1938 | 257.7  | 2792.2 | 209.2  | 4    |
     5.         1 | 1939 | 330.8  | 4313.2 | 203.4  | 5    |
     +-----------------+------------------+-+-----------------+
\end{verbatim}

To compute PCSEs, Stata must be able to identify the panel to which each observation belongs and be able to match the periods across the panels. We tell Stata how to do this matching by specifying the panel and time variables with \texttt{xtset}; see \texttt{[XT] xtset}. Because the data are annual, we specify the yearly option.

\begin{verbatim}
. xtset company year, yearly
    panel variable: company (strongly balanced)
    time variable: year, 1935 to 1954
    delta: 1 year
\end{verbatim}

We can obtain OLS parameter estimates for a linear model of \texttt{invest} on \texttt{mvalue} and \texttt{kstock} while allowing the standard errors (and variance–covariance matrix of the estimates) to be consistent when the disturbances from each observation are not independent. Specifically, we want the standard errors to be robust to each company having a different variance of the disturbances and to each company’s observations being correlated with those of the other companies through time.
This model is fit in Stata by typing

```
.xtpcse invest mvalue kstock
```

Linear regression, correlated panels corrected standard errors (PCSEs)

<table>
<thead>
<tr>
<th>Group variable:</th>
<th>company</th>
<th>Number of obs = 200</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time variable:</td>
<td>year</td>
<td>Number of groups = 10</td>
</tr>
<tr>
<td>Panels:</td>
<td>correlated (balanced)</td>
<td>Obs per group: min = 20</td>
</tr>
<tr>
<td>Autocorrelation:</td>
<td>no autocorrelation</td>
<td>avg = 20</td>
</tr>
<tr>
<td></td>
<td>max = 20</td>
<td></td>
</tr>
<tr>
<td>Estimated covariances</td>
<td>= 55</td>
<td>R-squared = 0.8124</td>
</tr>
<tr>
<td>Estimated autocorrelations</td>
<td>= 0</td>
<td>Wald chi2(2) = 637.41</td>
</tr>
<tr>
<td>Estimated coefficients</td>
<td>= 3</td>
<td>Prob &gt; chi2 = 0.0000</td>
</tr>
</tbody>
</table>

| Panel-corrected     | Coef. Std. Err. z  P>|z|  [95% Conf. Interval] |
|---------------------|---------------------|------------------|------------------|------------------|
| invest              | mvalue              | .1155622 .0072124 16.02 0.000  .101426 .1296983 |
|                     | kstock              | .2306785 .0278862 8.27 0.000   .1760225 .2853345 |
|                     | _cons               | -42.71437 6.780965 -6.30 0.000 -56.00482 -29.42392 |

Example 2: Comparing the FGLS and PCSE approaches

```
.xtgls invest mvalue kstock, panels(correlated)
```

Cross-sectional time-series FGLS regression

Coefficients: generalized least squares

| Panels: heteroskedastic with cross-sectional correlation |
| Correlation: no autocorrelation |
| Estimated covariances | = 55      | Number of obs = 200 |
| Estimated autocorrelations | = 0   | Number of groups = 10 |
| Estimated coefficients | = 3   | Time periods = 20 |
| Wald chi2(2) = 3738.07 |
| Prob > chi2 = 0.0000 |

| invest | Coef. Std. Err. z  P>|z|  [95% Conf. Interval] |
|--------|---------------------|------------------|------------------|------------------|
| mvalue | .1127515 .0022364 50.42 0.000   .1083683 .1171347 |
| kstock | .2231176 .0057363 38.90 0.000   .2118746 .2343605 |
| _cons  | -39.84382 1.717563 -23.20 0.000 -43.21018 -36.47746 |

The coefficients between the two models are close; the constants differ substantially, but we are generally not interested in the constant. As Beck and Katz observed, the standard errors for the FGLS model are 50%–100% smaller than those for the OLS model with PCSE.

If we were also concerned about autocorrelation of the disturbances, we could obtain a model with a common AR(1) parameter by specifying correlation(ar1).
. xtpcse invest mvalue kstock, correlation(ar1)
(note: estimates of rho outside [-1,1] bounded to be in the range [-1,1])

Prais-Winsten regression, correlated panels corrected standard errors (PCSEs)

Group variable: company Number of obs = 200
Time variable: year Number of groups = 10
Panels: correlated (balanced) Obs per group: min = 20
Autocorrelation: common AR(1) avg = 20

max = 20
Estimated covariances = 55 R-squared = 0.5468
Estimated autocorrelations = 1 Wald chi2(2) = 93.71
Estimated coefficients = 3 Prob > chi2 = 0.0000

| Panel-corrected | invest | Coef.  | Std. Err. | z     | P>|z|  | [95% Conf. Interval] |
|-----------------|--------|--------|-----------|-------|------|----------------------|
| mvalue          | .0950157 | .0129934 | 7.31     | 0.000 | .0695492 | .1204822 |
| kstock          | .306005 | .0603718 | 5.07     | 0.000 | .1876784 | .4243317 |
| _cons           | -39.12569 | 30.50355 | -1.28    | 0.200 | -98.91154 | 20.66016 |

rho = .9059774

The estimate of the autocorrelation parameter is high (0.906), and the standard errors are larger than for the model without autocorrelation, which is to be expected if there is autocorrelation.

Example 3: Controlling for cross-panel correlation and autocorrelation

Let’s estimate panel-specific autocorrelation parameters and change the method of estimating the autocorrelation parameter to the one typically used to estimate autocorrelation in time-series analysis.

. xtpcse invest mvalue kstock, correlation(psar1) rhotype(tscorr)

Prais-Winsten regression, correlated panels corrected standard errors (PCSEs)

Group variable: company Number of obs = 200
Time variable: year Number of groups = 10
Panels: correlated (balanced) Obs per group: min = 20
Autocorrelation: panel-specific AR(1) avg = 20

max = 20
Estimated covariances = 55 R-squared = 0.8670
Estimated autocorrelations = 10 Wald chi2(2) = 444.53
Estimated coefficients = 3 Prob > chi2 = 0.0000

| Panel-corrected | invest | Coef.  | Std. Err. | z     | P>|z|  | [95% Conf. Interval] |
|-----------------|--------|--------|-----------|-------|------|----------------------|
| mvalue          | .1052613 | .0086018 | 12.24    | 0.000 | .0884021 | .1221205 |
| kstock          | .3386743 | .0367568 | 9.21     | 0.000 | .2666322 | .4107163 |
| _cons           | -58.18714 | 12.83687 | -4.60    | 0.000 | -82.95496 | -33.41933 |

rhos = .5135627 .87017 .9023497 .63368 .8571502 ... .8752707

Beck and Katz (1995, 121) make a case against estimating panel-specific AR parameters, as opposed to one AR parameter for all panels.
Example 4: Controlling for heteroskedasticity only; not quite PCSEs

We can also diverge from PCSEs to estimate standard errors that are panel corrected, but only for panel-level heteroskedasticity; that is, each company has a different variance of the disturbances. Allowing also for autocorrelation, we would type

```
.xtpcse invest mvalue kstock, correlation(ar1) hetonly
(note: estimates of rho outside [-1,1] bounded to be in the range [-1,1])
```

Prais-Winsten regression, heteroskedastic panels corrected standard errors

<p>| | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Group variable:</td>
<td>company</td>
<td>Number of obs</td>
<td>=</td>
<td>200</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time variable:</td>
<td>year</td>
<td>Number of groups</td>
<td>=</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Panels:</td>
<td>heteroskedastic (balanced)</td>
<td>Obs per group: min</td>
<td>=</td>
<td>20</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Autocorrelation:</td>
<td>common AR(1)</td>
<td>avg</td>
<td>=</td>
<td>20</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>max</td>
<td>=</td>
<td>20</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Estimated covariances</td>
<td>=</td>
<td>10</td>
<td></td>
<td>R-squared</td>
<td>=</td>
<td>0.5468</td>
<td></td>
</tr>
<tr>
<td>Estimated autocorrelations</td>
<td>=</td>
<td>1</td>
<td></td>
<td>Wald chi2(2)</td>
<td>=</td>
<td>91.72</td>
<td></td>
</tr>
<tr>
<td>Estimated coefficients</td>
<td>=</td>
<td>3</td>
<td></td>
<td>Prob &gt; chi2</td>
<td>=</td>
<td>0.0000</td>
<td></td>
</tr>
</tbody>
</table>

| invest | Het-corrected Coef. | Std. Err. | z    | P>|z| | [95% Conf. Interval] |
|--------|---------------------|-----------|------|-------|----------------------|
| mvalue | .0950157            | .0130872  | 7.26 | 0.000 | .0693653              | .1206661     |
| kstock | .306005             | .061432   | 4.98 | 0.000 | .1856006              | .4264095     |
| _cons  | -39.12569           | 26.16935  | -1.50| 0.135 | -90.41666             | 12.16529     |
| rho    | .9059774            |           |      |       |                      |              |

With this specification, we do not obtain what are referred to in the literature as PCSEs. These standard errors are in the same spirit as PCSEs but are from the asymptotic covariance estimates of OLS without allowing for contemporaneous correlation.
Stored results

`xtpcse` stores the following in `e()`:

Scalars
- `e(N)`: number of observations
- `e(N_g)`: number of groups
- `e(N_gaps)`: number of gaps
- `e(n Cf)`: number of estimated coefficients
- `e(n_cv)`: number of estimated covariances
- `e(n_cr)`: number of estimated correlations
- `e(n_Sigma)`: observations used to estimate elements of `Sigma`
- `e(mss)`: model sum of squares
- `e(df)`: degrees of freedom
- `e(df_m)`: model degrees of freedom
- `e(rss)`: residual sum of squares
- `e(g_min)`: smallest group size
- `e(g_avg)`: average group size
- `e(g_max)`: largest group size
- `e(r2)`: $R^2$
- `e(chi2)`: $\chi^2$
- `e(p)`: significance
- `e(rmse)`: root mean squared error
- `e(rank)`: rank of `e(V)`
- `e(rc)`: return code

Macros
- `e(cmd)`:
- `e(cmdline)`:
- `e(depvar)`:
- `e(ivar)`:
- `e(tvar)`:
- `e(wtype)`:
- `e(wexp)`:
- `e(title)`:
- `e(panels)`:
- `e(corr)`:
- `e(rhotype)`:
- `e(rho)`:
- `e(cons)`:
- `e(missmeth)`:
- `e(balance)`:
- `e(chi2type)`:
- `e(vcetype)`:
- `e(properties)`:
- `e(predict)`:
- `e(marginsok)`:
- `e(asbalanced)`:
- `e(asobserved)`:

Matrices
- `e(b)`:
- `e(Sigma)`: $\Sigma$ matrix
- `e(rhomat)`:
- `e(V)`:

Functions
- `e(sample)`:

marks estimation sample
Methods and formulas

If no autocorrelation is specified, the parameters \( \beta \) are estimated by OLS; see [R] `regress`. If autocorrelation is specified, the parameters \( \beta \) are estimated by Prais–Winsten; see [TS] `prais`.

When autocorrelation with panel-specific coefficients of correlation is specified (by using option `correlation(psar1)`), each panel-level \( \rho_i \) is computed from the residuals of an OLS regression across all panels; see [TS] `prais`. When autocorrelation with a common coefficient of correlation is specified (by using option `correlation(ar1)`), the common correlation coefficient is computed as

\[
\rho = \frac{\rho_1 + \rho_2 + \cdots + \rho_m}{m}
\]

where \( \rho_i \) is the estimated autocorrelation coefficient for panel \( i \) and \( m \) is the number of panels.

The covariance of the OLS or Prais–Winsten coefficients is

\[
\text{Var}(\beta) = (X'X)^{-1}X'\Omega(X'X)^{-1}
\]

where \( \Omega \) is the full covariance matrix of the disturbances.

When the panels are balanced, we can write \( \Omega \) as

\[
\Omega = \Sigma_{m \times m} \otimes I_{T_i \times T_i}
\]

where \( \Sigma \) is the \( m \) by \( m \) panel-by-panel covariance matrix of the disturbances; see Remarks and examples.

`xtpcse` estimates the elements of \( \Sigma \) as

\[
\hat{\Sigma}_{ij} = \frac{\epsilon_i' \epsilon_j}{T_{ij}}
\]

where \( \epsilon_i \) and \( \epsilon_j \) are the residuals for panels \( i \) and \( j \), respectively, that can be matched by period, and where \( T_{ij} \) is the number of residuals between the panels \( i \) and \( j \) that can be matched by time period.

When the panels are balanced (each panel has the same number of observations and all periods are common to all panels), \( T_{ij} = T \), where \( T \) is the number of observations per panel.

When panels are unbalanced, `xtpcse` by default uses casewise selection, in which only those residuals from periods that are common to all panels are used to compute \( \hat{S}_{ij} \). Here \( T_{ij} = T^* \), where \( T^* \) is the number of periods common to all panels. When pairwise is specified, each \( \hat{S}_{ij} \) is computed using all observations that can be matched by period between the panels \( i \) and \( j \).

Acknowledgments

We thank the following people for helpful comments: Nathaniel Beck of the Department of Politics at New York University, Jonathan Katz of the Division of the Humanities and Social Science at California Institute of Technology, and Robert John Franzese Jr. of the Center for Political Studies at the Institute for Social Research at the University of Michigan.
References


Also see

[XT] xtpcse postestimation — Postestimation tools for xtpcse

[XT] xtset — Declare data to be panel data

[XT] xtreg — Fixed-, between-, and random-effects and population-averaged linear models

[XT] xtregar — Fixed- and random-effects linear models with an AR(1) disturbance

[R] regress — Linear regression

[TS] newey — Regression with Newey–West standard errors

[TS] prais — Prais–Winsten and Cochrane–Orcutt regression

[U] 20 Estimation and postestimation commands
## Description

The following postestimation commands are available after `xtpcse`:

<table>
<thead>
<tr>
<th>Command</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>contrast</code></td>
<td>contrasts and ANOVA-style joint tests of estimates</td>
</tr>
<tr>
<td><code>estat summarize</code></td>
<td>summary statistics for the estimation sample</td>
</tr>
<tr>
<td><code>estat vce</code></td>
<td>variance–covariance matrix of the estimators (VCE)</td>
</tr>
<tr>
<td><code>estimates</code></td>
<td>cataloging estimation results</td>
</tr>
<tr>
<td><code>forecast</code></td>
<td>dynamic forecasts and simulations</td>
</tr>
<tr>
<td><code>lincom</code></td>
<td>point estimates, standard errors, testing, and inference for linear combinations of coefficients</td>
</tr>
<tr>
<td><code>margins</code></td>
<td>marginal means, predictive margins, marginal effects, and average marginal effects</td>
</tr>
<tr>
<td><code>marginsplot</code></td>
<td>graph the results from margins (profile plots, interaction plots, etc.)</td>
</tr>
<tr>
<td><code>nlcom</code></td>
<td>point estimates, standard errors, testing, and inference for nonlinear combinations of coefficients</td>
</tr>
<tr>
<td><code>predict</code></td>
<td>predictions, residuals, influence statistics, and other diagnostic measures</td>
</tr>
<tr>
<td><code>predictnl</code></td>
<td>point estimates, standard errors, testing, and inference for generalized predictions</td>
</tr>
<tr>
<td><code>pwcompare</code></td>
<td>pairwise comparisons of estimates</td>
</tr>
<tr>
<td><code>test</code></td>
<td>Wald tests of simple and composite linear hypotheses</td>
</tr>
<tr>
<td><code>testnl</code></td>
<td>Wald tests of nonlinear hypotheses</td>
</tr>
</tbody>
</table>

1 `forecast` is not appropriate with mi estimation results.

## Syntax for predict

```
predict [type] newvar [if] [in] [, xb stdp]
```

These statistics are available both in and out of sample; type `predict ... if e(sample) ...` if wanted only for the estimation sample.

## Menu for predict

Statistics  >  Postestimation  >  Predictions, residuals, etc.

## Options for predict

- `xb`, the default, calculates the linear prediction.
- `stdp` calculates the standard error of the linear prediction.
Also see

[XT] xtpcse — Linear regression with panel-corrected standard errors

[U] 20 Estimation and postestimation commands
**xtpoisson — Fixed-effects, random-effects, and population-averaged Poisson models**

### Syntax

**Random-effects (RE) model**

```stata
xtpoisson depvar [indepvars] [if] [in] [weight] [ , re RE_options ]
```

**Conditional fixed-effects (FE) model**

```stata
xtpoisson depvar [indepvars] [if] [in] [weight] , fe [FE_options ]
```

**Population-averaged (PA) model**

```stata
xtpoisson depvar [indepvars] [if] [in] [weight] , pa [PA_options ]
```
### Model

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>noconstant</code></td>
<td>suppress constant term</td>
</tr>
<tr>
<td><code>re</code></td>
<td>use random-effects estimator; the default</td>
</tr>
<tr>
<td><code>exposure(varname)</code></td>
<td>include ln(varname) in model with coefficient constrained to 1</td>
</tr>
<tr>
<td><code>offset(varname)</code></td>
<td>include varname in model with coefficient constrained to 1</td>
</tr>
<tr>
<td><code>normal</code></td>
<td>use a normal distribution for random effects instead of gamma</td>
</tr>
<tr>
<td><code>constraints(constraints)</code></td>
<td>apply specified linear constraints</td>
</tr>
<tr>
<td><code>collinear</code></td>
<td>keep collinear variables</td>
</tr>
</tbody>
</table>

### SE/Robust

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>vce(vcetype)</code></td>
<td>vcetype may be oim, robust, cluster clustvar, bootstrap, or jackknife</td>
</tr>
</tbody>
</table>

### Reporting

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>level(#)</code></td>
<td>set confidence level; default is level(95)</td>
</tr>
<tr>
<td><code>irr</code></td>
<td>report incidence-rate ratios</td>
</tr>
<tr>
<td><code>noskip</code></td>
<td>perform overall model test as a likelihood-ratio test</td>
</tr>
<tr>
<td><code>nocnsreport</code></td>
<td>do not display constraints</td>
</tr>
<tr>
<td><code>display_options</code></td>
<td>control column formats, row spacing, line width, display of omitted variables and base and empty cells, and factor-variable labeling</td>
</tr>
</tbody>
</table>

### Integration

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>intmethod(intmethod)</code></td>
<td>integration method; intmethod may be mvaghermite (the default) or ghermite</td>
</tr>
<tr>
<td><code>intpoints(#)</code></td>
<td>use # quadrature points; default is intpoints(12)</td>
</tr>
</tbody>
</table>

### Maximization

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>maximize_options</code></td>
<td>control the maximization process; seldom used</td>
</tr>
<tr>
<td><code>coeflegend</code></td>
<td>display legend instead of statistics</td>
</tr>
</tbody>
</table>
### Model

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>fe</code></td>
<td>use fixed-effects estimator</td>
</tr>
<tr>
<td><code>exposure(varname)</code></td>
<td>include ln(varname) in model with coefficient constrained to 1</td>
</tr>
<tr>
<td><code>offset(varname)</code></td>
<td>include varname in model with coefficient constrained to 1</td>
</tr>
<tr>
<td><code>constraints(constraints)</code></td>
<td>apply specified linear constraints</td>
</tr>
<tr>
<td><code>collinear</code></td>
<td>keep collinear variables</td>
</tr>
</tbody>
</table>

### SE/Robust

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>vce(vcetype)</code></td>
<td><code>vcetype</code> may be oim, robust, bootstrap, or jackknife</td>
</tr>
</tbody>
</table>

### Reporting

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>level(#)</code></td>
<td>set confidence level; default is level(95)</td>
</tr>
<tr>
<td><code>irr</code></td>
<td>report incidence-rate ratios</td>
</tr>
<tr>
<td><code>nocnsreport</code></td>
<td>do not display constraints</td>
</tr>
</tbody>
</table>

### Display Options

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>display_options</code></td>
<td>control column formats, row spacing, line width, display of omitted variables and base and empty cells, and factor-variable labeling</td>
</tr>
</tbody>
</table>

### Maximization

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>maximize_options</code></td>
<td>control the maximization process; seldom used</td>
</tr>
<tr>
<td><code>coeflegend</code></td>
<td>display legend instead of statistics</td>
</tr>
</tbody>
</table>
**Model**

- **noconstant**: suppress constant term
- **pa**: use population-averaged estimator
- **exposure(varname)**: include ln(varname) in model with coefficient constrained to 1
- **offset(varname)**: include varname in model with coefficient constrained to 1

**Correlation**

- **corr(correlation)**: within-panel correlation structure
- **force**: estimate if observations unequally spaced in time

**SE/Robust**

- **vce(vcetype)**: vcetype may be conventional, robust, bootstrap, or jackknife
- **nmp**: use divisor \(N - P\) instead of the default \(N\)
- **scale(parm)**: overrides the default scale parameter; \(parm\) may be \(x^2\), dev, phi, or \#

**Reporting**

- **level(#)***: set confidence level; default is level(95)
- **irr**: report incidence-rate ratios
- **display_options**: control column formats, row spacing, line width, display of omitted variables and base and empty cells, and factor-variable labeling

**Optimization**

- **optimize_options**: control the optimization process; seldom used
- **coeflegend**: display legend instead of statistics

**correlation**

- **exchangeable**: exchangeable
- **independent**: independent
- **unstructured**: unstructured
- **fixed matname**: user-specified
- **ar #**: autoregressive of order #
- **stationary #**: stationary of order #
- **nonstationary #**: nonstationary of order #

A panel variable must be specified. For xtpoisson, pa, correlation structures other than exchangeable and independent require that a time variable also be specified. Use xtset; see [XT] xtset.

**deplvars** may contain factor variables; see [U] 11.4.3 Factor variables. depvar and indeplvars may contain time-series operators; see [U] 11.4.4 Time-series varlists.

by, mi estimate, and statsby are allowed; see [U] 11.1.10 Prefix commands. fp is allowed for the random-effects and fixed-effects models.

vce(bootstrap) and vce(jackknife) are not allowed with the mi estimate prefix; see [MI] mi estimate.
iweights, fweights, and pweights are allowed for the population-averaged model and iweights are allowed for the random-effects and fixed-effects models; see [U] 11.1.6 weight. Weights must be constant within panel.

coeflegend does not appear in the dialog box.

See [U] 20 Estimation and postestimation commands for more capabilities of estimation commands.
xtpoisson — Fixed-effects, random-effects, and population-averaged Poisson models

Menu
Statistics > Longitudinal/panel data > Count outcomes > Poisson regression (FE, RE, PA)

Description

xtpoisson fits random-effects, conditional fixed-effects, and population-averaged Poisson models. Whenever we refer to a fixed-effects model, we mean the conditional fixed-effects model.

By default, the population-averaged model is an equal-correlation model; xtpoisson, pa assumes corr(exchangeable). See [XT] xtgee for information on how to fit other population-averaged models.

Options for RE model

<table>
<thead>
<tr>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>noconstant; see [R] estimation options.</td>
</tr>
<tr>
<td>re, the default, requests the random-effects estimator.</td>
</tr>
<tr>
<td>exposure(varname), offset(varname); see [R] estimation options.</td>
</tr>
<tr>
<td>normal specifies that the random effects follow a normal distribution instead of a gamma distribution.</td>
</tr>
<tr>
<td>constraints(constraints), collinear; see [R] estimation options.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SE/Robust</th>
</tr>
</thead>
<tbody>
<tr>
<td>vce(vcetype) specifies the type of standard error reported, which includes types that are derived from asymptotic theory (oim), that are robust to some kinds of misspecification (robust), that allow for intragroup correlation (cluster clustvar), and that use bootstrap or jackknife methods (bootstrap, jackknife); see [XT] vce_options.</td>
</tr>
<tr>
<td>Specifying vce(robust) is equivalent to specifying vce(cluster panelvar); see xtpoisson, re normal and the robust VCE estimator in Methods and formulas.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Reporting</th>
</tr>
</thead>
<tbody>
<tr>
<td>level(#); see [R] estimation options.</td>
</tr>
<tr>
<td>irr reports exponentiated coefficients $e^b$ rather than coefficients $b$. For the Poisson model, exponentiated coefficients are interpreted as incidence-rate ratios.</td>
</tr>
<tr>
<td>noskip; see [R] estimation options.</td>
</tr>
<tr>
<td>nocnsreport; see [R] estimation options.</td>
</tr>
<tr>
<td>display_options: nomitted, vsquish, noemptycells, baselevels, allbaselevels, nolabel, fvwrap(#), fwrapon(style), cformat(%,fmt), pformat(%,fmt), sformat(%,fmt), and nolstretch; see [R] estimation options.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Integration</th>
</tr>
</thead>
<tbody>
<tr>
<td>intmethod(intmethod), intpoints(#); see [R] estimation options. normal must also be specified.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Maximization</th>
</tr>
</thead>
<tbody>
<tr>
<td>maximize_options: difficult, technique(algorithm_spec), iterate(#), [no]log, trace, gradient, showstep, hessian, showtolerance, tolerance(#), ltolerance(#), nrtolerance(#), nonrtolerance, and from(init_specs); see [R] maximize. These options are seldom used.</td>
</tr>
</tbody>
</table>
The following option is available with *xtpoisson* but is not shown in the dialog box:

*coeflegend*; see [R] estimation options.

### Options for FE model

**Model**

*fe* requests the fixed-effects estimator.

*exposure(varname)*, *offset(varname)*, *constraints(constraints)*, *collinear*; see [R] estimation options.

**SE/Robust**

*vce(vcetype)* specifies the type of standard error reported, which includes types that are derived from asymptotic theory (*oim*), that are robust to some kinds of misspecification (*robust*), and that use bootstrap or jackknife methods (*bootstrap*, *jackknife*); see [XT] *vce_options*.

*vce(robust)* invokes a cluster–robust estimate of the VCE in which the ID variable specifies the clusters.

**Reporting**

*level(#)*) see [R] estimation options.

*ierr* reports exponentiated coefficients $e^b$ rather than coefficients $b$. For the Poisson model, exponentiated coefficients are interpreted as incidence-rate ratios.

*nocnsreport*; see [R] estimation options.

*display_options*: *noomitted*, *vsquish*, *noemptycells*, *baselevels*, *allbaselevels*, *nofvlabel*, *fvwrap(#)*, *fvwrapon(style)*, *cformat(%fmt)*, *pformat(%fmt)*, *sformat(%fmt)*, and *nolstretch*; see [R] estimation options.

**Maximization**

*maximize_options*: *dificult*, *technique(algorithm_spec)*, *iterate(#)*, [*no*] *log*, *trace*, *gradient*, *showstep*, *hessian*, *showtolerance*, *tolerance(#)*, *ltolerance(#)*, *nrtolerance(#)*, *nonrtolerance*, and *from(init_specs)*; see [R] maximize. These options are seldom used.

The following option is available with *xtpoisson* but is not shown in the dialog box:

*coeflegend*; see [R] estimation options.

### Options for PA model

**Model**

*noconstant*; see [R] estimation options.

*pa* requests the population-averaged estimator.

*exposure(varname)*, *offset(varname)*; see [R] estimation options.
correlation) specifies the within-panel correlation structure; the default corresponds to the equal-correlation model, corr(exchangeable).

When you specify a correlation structure that requires a lag, you indicate the lag after the structure's name with or without a blank; for example, corr(ar 1) or corr(ar1).

If you specify the fixed correlation structure, you specify the name of the matrix containing the assumed correlations following the word fixed, for example, corr(fixed myr).

force specifies that estimation be forced even though the time variable is not equally spaced. This is relevant only for correlation structures that require knowledge of the time variable. These correlation structures require that observations be equally spaced so that calculations based on lags correspond to a constant time change. If you specify a time variable indicating that observations are not equally spaced, the (time dependent) model will not be fit. If you also specify force, the model will be fit, and it will be assumed that the lags based on the data ordered by the time variable are appropriate.

evce(vcetype) specifies the type of standard error reported, which includes types that are derived from asymptotic theory (conventional), that are robust to some kinds of misspecification (robust), and that use bootstrap or jackknife methods (bootstrap, jackknife); see [XT] vce_options. vce(conventional), the default, uses the conventionally derived variance estimator for generalized least-squares regression.

level(#); see [R] estimation options.

irr reports exponentiated coefficients $e^b$ rather than coefficients $b$. For the Poisson model, exponentiated coefficients are interpreted as incidence-rate ratios.

display_options: noomitted, vsquish, noemptycells, baselevels, allbaselevels, nofvlabel, fvwrap(#), fvwrapon(style), cformat(%fmt), pformat(%fmt), sformat(%fmt), and nolstretch; see [R] estimation options.

optimize_options control the iterative optimization process. These options are seldom used.

iterate(#) specifies the maximum number of iterations. When the number of iterations equals #, the optimization stops and presents the current results, even if convergence has not been reached. The default is iterate(100).

tolerance(#) specifies the tolerance for the coefficient vector. When the relative change in the coefficient vector from one iteration to the next is less than or equal to #, the optimization process is stopped. tolerance(1e-6) is the default.

nolog suppresses display of the iteration log.

trace specifies that the current estimates be printed at each iteration.

The following option is available with xtpoisson but is not shown in the dialog box: coeflegend; see [R] estimation options.
**Remarks and examples**

`xtpoisson` is a convenience command if you want the population-averaged model. Typing

```
.xtpoisson ..., ..., pa exposure(time)
```

is equivalent to typing

```
.xtgee ..., ..., family(poisson) link(log) corr(exchangeable) exposure(time)
```

Also see [XT `xtgee`] for information about `xtpoisson`.

By default or when `re` is specified, `xtpoisson` fits via maximum likelihood the random-effects model

$$Pr(Y_{it} = y_{it} | x_{it}) = F(y_{it}, x_{it} \beta + \nu_i)$$

for $i = 1, \ldots, n$ panels, where $t = 1, \ldots, n_i$, and $F(x, z) = Pr(X = x)$, where $X$ is Poisson distributed with mean $\exp(z)$. In the standard random-effects model, $\nu_i$ is assumed to be i.i.d. such that $\exp(\nu_i)$ is gamma with mean one and variance $\alpha$, which is estimated from the data. If `normal` is specified, $\nu_i$ is assumed to be i.i.d. $N(0, \sigma^2_\nu)$.

**Example 1**

We have data on the number of ship accidents for five different types of ships ([McCullagh and Nelder 1989, 205](#)). We wish to analyze whether the “incident” rate is affected by the period in which the ship was constructed and operated. Our measure of exposure is months of service for the ship, and in this model, we assume that the exponentiated random effects are distributed as gamma with mean one and variance $\alpha$. 

---

---
. use http://www.stata-press.com/data/r13/ships
. xtpoisson accident op_75_79 co_65_69 co_70_74 co_75_79, exp(service) irr

Fitting Poisson model:
Iteration 0:  log likelihood = -147.37993
Iteration 1:  log likelihood = -80.372714
Iteration 2:  log likelihood = -80.116093
Iteration 3:  log likelihood = -80.115916
Iteration 4:  log likelihood = -80.115916

Fitting full model:
Iteration 0:  log likelihood = -79.653186
Iteration 1:  log likelihood = -76.990836  (not concave)
Iteration 2:  log likelihood = -74.824942
Iteration 3:  log likelihood = -74.811243
Iteration 4:  log likelihood = -74.811217
Iteration 5:  log likelihood = -74.811217

Random-effects Poisson regression  Number of obs = 34
Group variable: ship   Number of groups = 5
Random effects u_i ~ Gamma
Obs per group:  min = 6
               avg = 6.8
               max = 7
Wald chi2(4) = 50.90
Log likelihood = -74.811217  Prob > chi2 = 0.0000

|                | IRR   | Std. Err. | z     | P>|z|  | [95% Conf. Interval] |
|----------------|-------|-----------|-------|------|----------------------|
| accident       |       |           |       |      |                      |
| op_75_79       | 1.466 | .173      | 3.24  | 0.001| 1.162957    1.848777 |
| co_65_69       | 2.032 | .304      | 4.74  | 0.000| 1.515982    2.72512 |
| co_70_74       | 2.357 | .399      | 5.05  | 0.000| 1.690033    3.286774 |
| co_75_79       | 1.642 | .381      | 2.14  | 0.033| 1.04174     2.58786 |
| _cons          | .001  | .000      | -30.24| 0.000| .0008952    .0021044 |
| ln(service)     | -2.368| .847      | -4.023| 0.000| -1.98710    -4.36990 |

Likelihood-ratio test of alpha=0:  chibar2(01) = 10.61  Prob>chibar2 = 0.001

The output also includes a likelihood-ratio test of \( \alpha = 0 \), which compares the panel estimator with the pooled (Poisson) estimator.

We find that the incidence rate for accidents is significantly different for the periods of construction and operation of the ships and that the random-effects model is significantly different from the pooled model.
We may alternatively fit a fixed-effects specification instead of a random-effects specification:

```
xtpoisson accident op_75_79 co_65_69 co_70_74 co_75_79, exp(service) irr fe
```

Iteration 0:  log likelihood = -80.738973
Iteration 1:  log likelihood = -54.857546
Iteration 2:  log likelihood = -54.641897
Iteration 3:  log likelihood = -54.641859
Iteration 4:  log likelihood = -54.641859

Conditional fixed-effects Poisson regression

<table>
<thead>
<tr>
<th></th>
<th>Number of obs</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Group variable:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ship</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Obs per group:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>min</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>avg</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>max</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wald chi2(4)</td>
<td></td>
<td>48.44</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log likelihood</td>
<td></td>
<td>-54.641859</td>
<td>0.0000</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| accident     | IRR    | Std. Err. | z    | P>|z|  | [95% Conf. Interval] |
|--------------|--------|-----------|------|------|----------------------|
| op_75_79     | 1.468831 | .1737218  | 3.25 | 0.001| 1.164926 - 1.852019 |
| co_65_69     | 2.008003 | .3004803  | 4.66 | 0.000| 1.497577 - 2.692398 |
| co_70_74     | 2.26693  | .384865   | 4.82 | 0.000| 1.625274 - 3.161912 |
| co_75_79     | 1.573695 | .3669393  | 1.94 | 0.052| .9964273 - 2.485397 |
| ln(service)   | 1 (exposure) |                  |      |      |                      |

Both of these models fit the same thing but will differ in efficiency, depending on whether the assumptions of the random-effects model are true.

We could have assumed that the random effects followed a normal distribution, \( N(0, \sigma^2_\nu) \), instead of a “log-gamma” distribution, and obtained

```
xtpoisson accident op_75_79 co_65_69 co_70_74 co_75_79, exp(service) irr > normal nolog
```

Random-effects Poisson regression

<table>
<thead>
<tr>
<th></th>
<th>Number of obs</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Group variable:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ship</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Obs per group:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>min</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>avg</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>max</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wald chi2(4)</td>
<td></td>
<td>50.95</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log likelihood</td>
<td></td>
<td>-74.780982</td>
<td>0.0000</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| accident     | IRR    | Std. Err. | z    | P>|z|  | [95% Conf. Interval] |
|--------------|--------|-----------|------|------|----------------------|
| op_75_79     | 1.466677 | .1734403  | 3.24 | 0.001| 1.163259 - 1.849236 |
| co_65_69     | 2.032604 | .3040933  | 4.74 | 0.000| 1.516025 - 2.725205 |
| co_70_74     | 2.357045 | .3998397  | 5.05 | 0.000| 1.690338 - 3.286717 |
| co_75_79     | 1.646935 | .3820235  | 2.15 | 0.031| 1.045278 - 2.594905 |
| _cons        | .0013075 | .0002775  | -31.28| 0.000| .0008625 - .001982 |
| ln(service)   | 1 (exposure) |                  |      |      |                      |
| /lnsig2u     | -2.351868 | .8586262  | -2.74| 0.006| -4.034745 - .6689918|
| sigma_u      | .3085306 | .1324562  |     |      | .1330045 - .7156988 |

Likelihood-ratio test of sigma_u=0: chibar2(01) = 10.67 Pr>=chibar2 = 0.001

The output includes the additional panel-level variance component. This is parameterized as the log of the variance \( \ln(\sigma^2_\nu) \) (labeled lnsig2u in the output). The standard deviation \( \sigma_\nu \) is also included in the output labeled sigma_u.
When $\sigma_u$ is zero, the panel-level variance component is unimportant and the panel estimator is no different from the pooled estimator. A likelihood-ratio test of this is included at the bottom of the output. This test formally compares the pooled estimator (poisson) with the panel estimator. Here $\sigma_\nu$ is significantly greater than zero, so a panel estimator is indicated.

Example 2

This time we fit a robust equal-correlation population-averaged model:

```
xtpoisson accident op_75_79 co_65_69 co_70_74 co_75_79, exp(service) pa
correlation(exchangeable)
```

```
Iteration 1: tolerance = .04083192
Iteration 2: tolerance = .00270188
Iteration 3: tolerance = .00030663
Iteration 4: tolerance = .00003466
Iteration 5: tolerance = 3.891e-06
Iteration 6: tolerance = 4.359e-07
```

GEE population-averaged model

```
GEE population-averaged model
Number of obs = 34
Group variable: ship Number of groups = 5
Link: log Obs per group: min = 6
Family: Poisson avg = 6.8
Correlation: exchangeable max = 7
Wald chi2(4) = 252.94
Scale parameter: (Std. Err. adjusted for clustering on ship)
```

```
<table>
<thead>
<tr>
<th>accident</th>
<th>Robust</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>IRR</td>
<td>Std. Err.</td>
<td>z</td>
<td>P&gt;</td>
<td>z</td>
</tr>
<tr>
<td>op_75_79</td>
<td>1.483299</td>
<td>.1197901</td>
<td>4.88</td>
<td>0.000</td>
<td>1.266153</td>
</tr>
<tr>
<td>co_65_69</td>
<td>2.038477</td>
<td>.1809524</td>
<td>8.02</td>
<td>0.000</td>
<td>1.712955</td>
</tr>
<tr>
<td>co_70_74</td>
<td>2.643467</td>
<td>.4093947</td>
<td>6.28</td>
<td>0.000</td>
<td>1.951407</td>
</tr>
<tr>
<td>co_75_79</td>
<td>1.876656</td>
<td>.33075</td>
<td>3.57</td>
<td>0.000</td>
<td>1.328511</td>
</tr>
<tr>
<td>_cons</td>
<td>.0010255</td>
<td>.0000721</td>
<td>-97.90</td>
<td>0.000</td>
<td>.0008935</td>
</tr>
<tr>
<td>ln(service)</td>
<td>1 (exposure)</td>
<td>IRR</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
We may compare this with a pooled estimator with clustered robust-variance estimates:

```
. poisson accident op_75_79 co_65_69 co_70_74 co_75_79, exp(service)
> vce(cluster ship) irr
```

```
Iteration 0: log pseudolikelihood = -147.37993
Iteration 1: log pseudolikelihood = -80.372714
Iteration 2: log pseudolikelihood = -80.116093
Iteration 3: log pseudolikelihood = -80.115916
Iteration 4: log pseudolikelihood = -80.115916
```

Poisson regression

```
Number of obs = 34
Wald chi2(3) = .
Prob > chi2 = .
Log pseudolikelihood = -80.115916
```

(Std. Err. adjusted for 5 clusters in ship)

```
Robust
accident IRR Std. Err. z P>|z| [95% Conf. Interval]
```

```
op_75_79 1.47324 .1287036 4.44 0.000 1.2414 1.748377
co_65_69 2.125914 .2850531 5.62 0.000 1.634603 2.764897
co_70_74 2.860138 .6213563 4.84 0.000 1.868384 4.378325
co_75_79 2.021926 .4265285 3.34 0.001 1.337221 3.057227
_cons .0009609 .0000277 -240.66 0.000 .000908 .0010168
ln(service) 1 (exposure)
```

Technical note

The random-effects model is calculated using quadrature, which is an approximation whose accuracy depends partially on the number of integration points used. We can use the `quadchk` command to see if changing the number of integration points affects the results. If the results change, the quadrature approximation is not accurate given the number of integration points. Try increasing the number of integration points using the `intpoints()` option and run `quadchk` again. Do not attempt to interpret the results of estimates when the coefficients reported by `quadchk` differ substantially. See `[XT] quadchk` for details and `[XT] xtpoisson` for an example.

Because the `xtpoisson, re normal` likelihood function is calculated by Gauss–Hermite quadrature, on large problems the computations can be slow. Computation time is roughly proportional to the number of points used for the quadrature.
Stored results

xtpoisson, re stores the following in e():

Scalars
- e(N) number of observations
- e(N_g) number of groups
- e(N_cd) number of completely determined observations
- e(k) number of parameters
- e(k_aux) number of auxiliary parameters
- e(k_eq) number of equations in e(b)
- e(k_eq_model) number of equations in overall model test
- e(k_dv) number of dependent variables
- e(ll) log likelihood
- e(ll_0) log likelihood, constant-only model
- e(ll_c) log likelihood, comparison model
- e(chi2) \( \chi^2 \)
- e(chi2_c) \( \chi^2 \) for comparison test
- e(sigma_u) panel-level standard deviation
- e(alpha) value of alpha
- e(g_min) smallest group size
- e(g_avg) average group size
- e(g_max) largest group size
- e(p) significance
- e(rank) rank of e(V)
- e(rank0) rank of e(V) for constant-only model
- e(ic) number of iterations
- e(rc) return code
- e(converged) 1 if converged, 0 otherwise

Macros
- e(cmd) xtpoisson
- e(cmdline) command as typed
- e(depvar) name of dependent variable
- e(ivar) variable denoting groups
- e(model) re
- e(wtype) weight type
- e(wexp) weight expression
- e(ttitle) title in estimation output
- e(offset) linear offset variable
- e(chi2type) Wald or LR; type of model \( \chi^2 \) test
- e(chi2_c) Wald or LR; type of model \( \chi^2 \) test corresponding to e(chi2_c)
- e(vc0) vc0 specified in vce()
- e(vcetype) title used to label Std. Err.
- e(method) requested estimation method
- e(distrib) Gamma; the distribution of the random effect
- e(opt) type of optimization
- e(which) max or min; whether optimizer is to perform maximization or minimization
- e(ml_method) type of ml method
- e(user) name of likelihood-evaluator program
- e(technique) maximization technique
- e(properties) b V
- e(predict) program used to implement predict
- e(asbalanced) factor variables fvset as asbalanced
- e(asobserved) factor variables fvset as asobserved
Matrices
\[ e(b) \]  
coefficient vector
\[ e(Cns) \]  
constraints matrix
\[ e(ilog) \]  
iteration log
\[ e(gradient) \]  
gradient vector
\[ e(V) \]  
variance–covariance matrix of the estimators

Functions
\[ e(sample) \]  
marks estimation sample

\[ \text{xtpoisson, re normal} \]  
stores the following in \( e() \):

Scalars
\[ e(N) \]  
number of observations
\[ e(N_g) \]  
number of groups
\[ e(N_cd) \]  
number of completely determined observations
\[ e(k) \]  
number of parameters
\[ e(k_aux) \]  
number of auxiliary parameters
\[ e(k_eq) \]  
number of equations in \( e(b) \)
\[ e(k_eq_model) \]  
number of equations in overall model test
\[ e(k_dv) \]  
number of dependent variables
\[ e(df_m) \]  
model degrees of freedom
\[ e(ll) \]  
log likelihood
\[ e(ll_0) \]  
log likelihood, constant-only model
\[ e(ll_c) \]  
log likelihood, comparison model
\[ e(chi2) \]  
\( \chi^2 \)
\[ e(chi2_c) \]  
\( \chi^2 \) for comparison test
\[ e(N_clust) \]  
number of clusters
\[ e(sigma_u) \]  
panel-level standard deviation
\[ e(n_quad) \]  
number of quadrature points
\[ e(g_min) \]  
smallest group size
\[ e(g_avg) \]  
average group size
\[ e(g_max) \]  
largest group size
\[ e(p) \]  
significance
\[ e(rank) \]  
rank of \( e(V) \)
\[ e(rank0) \]  
rank of \( e(V) \) for constant-only model
\[ e(ic) \]  
number of iterations
\[ e(rc) \]  
return code
\[ e(converged) \]  
1 if converged, 0 otherwise
xtpoisson, fe stores the following in e():

Scalars
\[
\begin{align*}
\text{e(N)} & \quad \text{number of observations} \\
\text{e(N_g)} & \quad \text{number of groups} \\
\text{e(k)} & \quad \text{number of parameters} \\
\text{e(k_eq)} & \quad \text{number of equations in e(b)} \\
\text{e(k_eq_model)} & \quad \text{number of equations in overall model test} \\
\text{e(k_dv)} & \quad \text{number of dependent variables} \\
\text{e(df_m)} & \quad \text{model degrees of freedom} \\
\text{e(ll)} & \quad \log \text{ likelihood} \\
\text{e(ll_0)} & \quad \log \text{ likelihood, constant-only model} \\
\text{e(ll_c)} & \quad \log \text{ likelihood, comparison model} \\
\text{e(chi2)} & \quad \chi^2 \\
\text{e(g_min)} & \quad \text{smallest group size} \\
\text{e(g_avg)} & \quad \text{average group size} \\
\text{e(g_max)} & \quad \text{largest group size} \\
\text{e(p)} & \quad \text{significance} \\
\text{e(rank)} & \quad \text{rank of e(V)} \\
\text{e(ic)} & \quad \text{number of iterations} \\
\text{e(rc)} & \quad \text{return code} \\
\text{e(converged)} & \quad 1 \text{ if converged, 0 otherwise}
\end{align*}
\]
Macros

- `e(cmd)` - xtpoisson
- `e(cmdline)` - command as typed
- `e(depvar)` - name of dependent variable
- `e(ivar)` - variable denoting groups
- `e(model)` - `fe`
- `e(wtype)` - weight type
- `e(wexp)` - weight expression
- `e(title)` - title in estimation output
- `e(offset)` - linear offset variable
- `e(chi2type)` - Wald; type of model $\chi^2$ test
- `e(vce)` - vcetype specified in vce()
- `e(vcetype)` - title used to label Std. Err.
- `e(method)` - requested estimation method
- `e(opt)` - type of optimization
- `e(which)` - `max` or `min`; whether optimizer is to perform maximization or minimization
- `e(ml_method)` - type of ml method
- `e(user)` - name of likelihood-evaluator program
- `e(technique)` - maximization technique
- `e(properties)` - b V
- `e(predict)` - program used to implement predict
- `e(asbalanced)` - factor variables fvset as asbalanced
- `e(asobserved)` - factor variables fvset as asobserved

Matrices

- `e(b)` - coefficient vector
- `e(Cns)` - constraints matrix
- `e(ilog)` - iteration log
- `e(gradient)` - gradient vector
- `e(V)` - variance–covariance matrix of the estimators

Functions

- `e(sample)` - marks estimation sample

xtpoisson, pa stores the following in `e()`:

Scalars

- `e(N)` - number of observations
- `e(N_g)` - number of groups
- `e(df_m)` - model degrees of freedom
- `e(chi2)` - $\chi^2$
- `e(p)` - significance
- `e(df_pear)` - degrees of freedom for Pearson $\chi^2$
- `e(chi2_dev)` - $\chi^2$ test of deviance
- `e(chi2_dis)` - $\chi^2$ test of deviance dispersion
- `e(deviance)` - deviance
- `e(dispers)` - deviance dispersion
- `e(phi)` - scale parameter
- `e(g_min)` - smallest group size
- `e(g_avg)` - average group size
- `e(g_max)` - largest group size
- `e(rank)` - rank of `e(V)`
- `e(tol)` - target tolerance
- `e(dif)` - achieved tolerance
- `e(rc)` - return code
xtpoisson, pa reports the population-averaged results obtained by using xtgee, family(poisson) link(log) to obtain estimates. See [XT] xtgee for details about the methods and formulas.

xtpoisson, fe with robust standard errors implements the formula presented in Wooldridge (1999). The formula is a cluster–robust estimate of the VCE in which the ID variable specifies the clusters.

Although Hausman, Hall, and Griliches (1984) wrote the seminal article on the random-effects and fixed-effects models, Cameron and Trivedi (2013) provide a good textbook treatment. Allison (2009, chap. 4) succinctly discusses these models and illustrates the differences between them using Stata.

For a random-effects specification, we know that

\[
\Pr(y_{it1}, \ldots, y_{itn_i} | \alpha_i, x_{i1}, \ldots, x_{in_i}) = \left( \prod_{t=1}^{n_i} \frac{\lambda_{it}^{y_{it}}}{y_{it}!} \right) \exp \left\{ - \exp(\alpha_i) \sum_{t=1}^{n_i} \lambda_{it} \right\} \exp \left( \alpha_i \sum_{t=1}^{n_i} y_{it} \right)
\]
where \( \lambda_{it} = \exp(x_{it}\beta) \). We may rewrite the above as [defining \( \epsilon_i = \exp(\alpha_i) \)]

\[
\Pr(y_{i1}, \ldots, y_{ini} | \epsilon_i, x_{i1}, \ldots, x_{ini}) = \left\{ \prod_{t=1}^{n_i} \frac{\lambda_{it} \epsilon_i y_{it}}{y_{it}!} \right\} \exp \left( -\sum_{t=1}^{n_i} \lambda_{it} \epsilon_i \right) = \left( \prod_{t=1}^{n_i} \frac{\lambda_{it} y_{it}}{y_{it}!} \right) \exp \left( -\epsilon_i \sum_{t=1}^{n_i} \lambda_{it} \right) \epsilon_i \sum_{t=1}^{n_i} y_{it}
\]

We now assume that \( \epsilon_i \) follows a gamma distribution with mean one and variance \( 1/\theta \) so that unconditional on \( \epsilon_i \)

\[
\Pr(y_{i1}, \ldots, y_{ini} | X_i) = \frac{\theta^\theta}{\Gamma(\theta)} \left( \prod_{t=1}^{n_i} \frac{\lambda_{it} y_{it}}{y_{it}!} \right) \int_0^\infty \exp \left( -\epsilon_i \sum_{t=1}^{n_i} \lambda_{it} \right) \epsilon_i^{\theta-1} \exp(-\theta \epsilon_i) d\epsilon_i
\]

\[
= \frac{\theta^\theta}{\Gamma(\theta)} \left( \prod_{t=1}^{n_i} \frac{\lambda_{it} y_{it}}{y_{it}!} \right) \int_0^\infty \exp \left\{ -\epsilon_i \left( \theta + \sum_{t=1}^{n_i} \lambda_{it} \right) \right\} \epsilon_i^{\theta + \sum_{t=1}^{n_i} y_{it} - 1} d\epsilon_i
\]

\[
= \left( \prod_{t=1}^{n_i} \frac{\lambda_{it} y_{it}}{y_{it}!} \right) \frac{\Gamma \left( \theta + \sum_{t=1}^{n_i} y_{it} \right)}{\Gamma(\theta)} \left( \frac{\theta}{\theta + \sum_{t=1}^{n_i} \lambda_{it}} \right)^{\theta} \left( \frac{1}{\theta + \sum_{t=1}^{n_i} \lambda_{it}} \right)^{n_i} \sum_{t=1}^{n_i} y_{it}
\]

for \( X_i = (x_{i1}, \ldots, x_{ini}) \).

The log likelihood (assuming gamma heterogeneity) is then derived using

\[
u_i = \frac{\theta}{\theta + \sum_{t=1}^{n_i} \lambda_{it}} \quad \lambda_{it} = \exp(x_{it}\beta)
\]

\[
\Pr(Y_{i1} = y_{i1}, \ldots, Y_{ini} = y_{ini} | X_i) = \frac{\prod_{t=1}^{n_i} \lambda_{it} y_{it} \Gamma(\theta + \sum_{t=1}^{n_i} y_{it})}{\prod_{t=1}^{n_i} y_{it}! \Gamma(\theta) \sum_{t=1}^{n_i} \lambda_{it}} u_i^\theta (1 - u_i) \sum_{t=1}^{n_i} y_{it}
\]

such that the log likelihood may be written as

\[
L = \sum_{i=1}^{n} w_i \left\{ \log \Gamma \left( \theta + \sum_{t=1}^{n_i} y_{it} \right) - \sum_{t=1}^{n_i} \log(1 + y_{it}) - \log(\Gamma(\theta)) + \theta \log u_i 
+ \log(1 - u_i) \sum_{t=1}^{n_i} y_{it} + \sum_{t=1}^{n_i} y_{it} (x_{it}\beta) - \left( \sum_{t=1}^{n_i} y_{it} \right) \log \left( \sum_{t=1}^{n_i} \lambda_{it} \right) \right\}
\]

where \( w_i \) is the user-specified weight for panel \( i \); if no weights are specified, \( w_i = 1 \).
Alternatively, if we assume a normal distribution, \( N(0, \sigma^2_\nu) \), for the random effects \( \nu_i \)

\[
\Pr(y_{i1}, \ldots, y_{in_i}|X_i) = \int_{-\infty}^{\infty} \frac{e^{-\nu_i^2/2\sigma^2_\nu}}{\sqrt{2\pi\sigma^2_\nu}} \left\{ \prod_{t=1}^{n_i} F(y_{it}, x_{it}\beta + \nu_i) \right\} d\nu_i
\]

where

\[
F(y, z) = \exp\left\{ -\exp(z) + yz - \log(y!) \right\}.
\]

The panel-level likelihood \( l_i \) is given by

\[
l_i = \int_{-\infty}^{\infty} \frac{e^{-\nu_i^2/2\sigma^2_\nu}}{\sqrt{2\pi\sigma^2_\nu}} \left\{ \prod_{t=1}^{n_i} F(y_{it}, x_{it}\beta + \nu_i) \right\} d\nu_i
\]

\[
\equiv \int_{-\infty}^{\infty} g(y_{it}, x_{it}, \nu_i) d\nu_i
\]

This integral can be approximated with \( M \)-point Gauss–Hermite quadrature

\[
\int_{-\infty}^{\infty} e^{-x^2} h(x) dx \approx \sum_{m=1}^{M} w_m^* h(a_m^*)
\]

This is equivalent to

\[
\int_{-\infty}^{\infty} f(x) dx \approx \sum_{m=1}^{M} w_m^* \exp\left\{ (a_m^*)^2 \right\} f(a_m^*)
\]

where the \( w_m^* \) denote the quadrature weights and the \( a_m^* \) denote the quadrature abscissas. The log likelihood, \( L \), is the sum of the logs of the panel-level likelihoods \( l_i \).

The default approximation of the log likelihood is by adaptive Gauss–Hermite quadrature, which approximates the panel-level likelihood with

\[
l_i \approx \sqrt{2\sigma_i} \sum_{m=1}^{M} w_m^* \exp\left\{ (a_m^*)^2 \right\} g(y_{it}, x_{it}, \sqrt{2\sigma_i}a_m^* + \hat{\mu}_i)
\]

where \( \hat{\sigma}_i \) and \( \hat{\mu}_i \) are the adaptive parameters for panel \( i \). Therefore, with the definition of \( g(y_{it}, x_{it}, \nu_i) \), the total log likelihood is approximated by

\[
L \approx \sum_{i=1}^{n} w_i \log\left[ \sqrt{2\sigma_i} \sum_{m=1}^{M} w_m^* \exp\left\{ (a_m^*)^2 \right\} \exp\left\{ -(\sqrt{2\sigma_i}a_m^* + \hat{\mu}_i)^2 / 2\sigma^2_\nu \right\} / \sqrt{2\pi\sigma^2_\nu} \right] \\
\prod_{t=1}^{n_i} F(y_{it}, x_{it}\beta + \sqrt{2\sigma_i}a_m^* + \hat{\mu}_i)
\]

where \( w_i \) is the user-specified weight for panel \( i \); if no weights are specified, \( w_i = 1 \).
The default method of adaptive Gauss–Hermite quadrature is to calculate the posterior mean and variance and use those parameters for $\hat{\mu}_i$ and $\hat{\sigma}_i$ by following the method of Naylor and Smith (1982), further discussed in Skrondal and Rabe-Hesketh (2004). We start with $\hat{\sigma}_{i,0} = 1$ and $\hat{\mu}_{i,0} = 0$, and the posterior means and variances are updated in the $k$th iteration. That is, at the $k$th iteration of the optimization for $l_i$, we use

$$l_{i,k} \approx \sum_{m=1}^{M} \sqrt{2} \hat{\sigma}_{i,k-1} w_m^* \exp \left\{ \left( a_m^* \right)^2 \right\} g(y_{it}, x_{it}, \sqrt{2} \hat{\sigma}_{i,k-1} a_m^* + \hat{\mu}_{i,k-1})$$

Letting

$$\tau_{i,m,k-1} = \sqrt{2} \hat{\sigma}_{i,k-1} a_m^* + \hat{\mu}_{i,k-1}$$

$$\hat{\mu}_{i,k} = \sum_{m=1}^{M} \left( \tau_{i,m,k-1} \right) \frac{\sqrt{2} \hat{\sigma}_{i,k-1} w_m^* \exp \left\{ \left( a_m^* \right)^2 \right\} g(y_{it}, x_{it}, \tau_{i,m,k-1})}{l_{i,k}}$$

and

$$\hat{\sigma}_{i,k} = \sum_{m=1}^{M} \left( \tau_{i,m,k-1} \right)^2 \frac{\sqrt{2} \hat{\sigma}_{i,k-1} w_m^* \exp \left\{ \left( a_m^* \right)^2 \right\} g(y_{it}, x_{it}, \tau_{i,m,k-1})}{l_{i,k}} - \left( \hat{\mu}_{i,k} \right)^2$$

and this is repeated until $\hat{\mu}_{i,k}$ and $\hat{\sigma}_{i,k}$ have converged for this iteration of the maximization algorithm. This adaptation is applied on every iteration until the log-likelihood change from the preceding iteration is less than a relative difference of 1e–6; after this, the quadrature parameters are fixed.

The log likelihood can also be calculated by nonadaptive Gauss–Hermite quadrature, the int-method(ghermite) option, where $\rho = \sigma^2_\nu / (\sigma^2_\nu + 1)$:

$$L = \sum_{i=1}^{n} w_i \log \left\{ \text{Pr}(y_{i1}, \ldots, y_{in_i} \mid x_{i1}, \ldots, x_{in_i}) \right\}$$

$$\approx \sum_{i=1}^{n} w_i \log \left[ \frac{1}{\sqrt{\pi}} \sum_{m=1}^{M} w_m^* \prod_{t=1}^{n_i} F \left( y_{it}, x_{it}\beta + a_m^* \left( \frac{2\rho}{1-\rho} \right)^{1/2} \right) \right]$$

Both quadrature formulas require that the integrated function be well approximated by a polynomial of degree equal to the number of quadrature points. The number of periods (panel size) can affect whether

$$\prod_{t=1}^{n_i} F(y_{it}, x_{it}\beta + \nu_i)$$

is well approximated by a polynomial. As panel size and $\rho$ increase, the quadrature approximation can become less accurate. For large $\rho$, the random-effects model can also become unidentified. Adaptive quadrature gives better results for correlated data and large panels than nonadaptive quadrature; however, we recommend that you use the quadchk command (see [XT] quadchk) to verify the quadrature approximation used in this command, whichever approximation you choose.
For a fixed-effects specification, we know that

\[
\Pr(Y_{it} = y_{it} | x_{it}) = \exp\{-\exp(\alpha_i + x_{it}\beta)\} \exp(\alpha_i + x_{it}\beta)^{y_{it}} / y_{it}!
\]

\[
= \frac{1}{y_{it}!} \exp\{-\exp(\alpha_i) \exp(x_{it}\beta) + \alpha_i y_{it}\} \exp(x_{it}\beta)^{y_{it}}
\]

\[
\equiv F_{it}
\]

Because we know that the observations are independent, we may write the joint probability for the observations within a panel as

\[
\Pr(Y_{i1} = y_{i1}, \ldots, Y_{in_i} = y_{in_i} | X_i)
\]

\[
= \prod_{t=1}^{n_i} \frac{1}{y_{it}!} \exp\{-\exp(\alpha_i) \exp(x_{it}\beta) + \alpha_i y_{it}\} \exp(x_{it}\beta)^{y_{it}}
\]

\[
= \left( \prod_{t=1}^{n_i} \frac{\exp(x_{it}\beta)^{y_{it}}}{y_{it}!} \right) \exp\left\{- \exp(\alpha_i) \sum_t \exp(x_{it}\beta) + \alpha_i \sum_t y_{it} \right\}
\]

and we also know that the sum of \( n_i \) Poisson independent random variables, each with parameter \( \lambda_{it} \) for \( t = 1, \ldots, n_i \), is distributed as Poisson with parameter \( \sum_t \lambda_{it} \). Thus

\[
\Pr\left(\sum_t Y_{it} = \sum_t y_{it} | X_i, \sum_t Y_{it} = \sum_t y_{it}\right) =
\]

\[
= \frac{1}{(\sum_t y_{it})!} \exp\left\{- \exp(\alpha_i) \sum_t \exp(x_{it}\beta) + \alpha_i \sum_t y_{it} \right\} \left\{ \sum_t \exp(x_{it}\beta) \right\} \sum_t y_{it}
\]

So, the conditional likelihood is conditioned on the sum of the outcomes in the set (panel). The appropriate function is given by

\[
\Pr(Y_{i1} = y_{i1}, \ldots, Y_{in_i} = y_{in_i} | X_i, \sum_t Y_{it} = \sum_t y_{it} ) =
\]

\[
\frac{1}{(\sum_t y_{it})!} \exp\left\{- \exp(\alpha_i) \sum_t \exp(x_{it}\beta) + \alpha_i \sum_t y_{it} \right\} \left\{ \sum_t \exp(x_{it}\beta) \right\} \sum_t y_{it}
\]

\[
= \left( \sum_t y_{it} \right)! \prod_{t=1}^{n_i} \frac{\exp(x_{it}\beta)^{y_{it}}}{y_{it}! \{\sum_k \exp(x_{ik}\beta)\}^{y_{it}}}
\]

which is free of \( \alpha_i \).
The conditional log likelihood is given by

\[ L = \log \prod_{i=1}^{n} \left( \sum_{t=1}^{n_i} y_{it} \right)! \prod_{t=1}^{n_i} y_{it}! \left\{ \sum_{\ell=1}^{n_i} \exp(x_{it}\beta) \right\}^{w_i} \]

\[ = \log \prod_{i=1}^{n} \left\{ \left( \sum_{t=1}^{n_i} y_{it} \right)! \prod_{t=1}^{n_i} y_{it}! \right\}^{w_i} \prod_{i=1}^{n} \left( \sum_{t=1}^{n_i} y_{it} + 1 \right)!^{-1} \prod_{t=1}^{n_i} y_{it}! \left\{ \sum_{\ell=1}^{n_i} \exp(x_{i\ell}\beta) \right\}^{y_{i\ell}} \]

\[ = \sum_{i=1}^{n} w_i \left\{ \log \Gamma \left( \sum_{t=1}^{n_i} y_{it} + 1 \right) - \sum_{t=1}^{n_i} \log \Gamma(y_{it} + 1) + \sum_{t=1}^{n_i} y_{it} \log p_{it} \right\} \]

where

\[ p_{it} = \frac{e^{x_{it}\beta}}{\sum_{\ell} e^{x_{i\ell}\beta}} \]

xtpoisson, re normal and the robust VCE estimator

Specifying vce(robust) or vce(cluster clustvar) causes the Huber/White/sandwich VCE estimator to be calculated for the coefficients estimated in this regression. See [P] _robust, particularly Introduction and Methods and formulas. Wooldridge (2013) and Arellano (2003) discuss this application of the Huber/White/sandwich VCE estimator. As discussed by Wooldridge (2013), Stock and Watson (2008), and Arellano (2003), specifying vce(robust) is equivalent to specifying vce(cluster panelvar), where panelvar is the variable that identifies the panels.

Clustering on the panel variable produces a consistent VCE estimator when the disturbances are not identically distributed over the panels or there is serial correlation in \( \epsilon_{it} \).

The cluster–robust VCE estimator requires that there are many clusters and the disturbances are uncorrelated across the clusters. The panel variable must be nested within the cluster variable because of the within-panel correlation that is generally induced by the random-effects transform when there is heteroskedasticity or within-panel serial correlation in the idiosyncratic errors.

References


**Also see**

[XT] *xtpoisson postestimation* — Postestimation tools for *xtpoisson*

[XT] *quadchk* — Check sensitivity of quadrature approximation

[XT] *xtgee* — Fit population-averaged panel-data models by using GEE

[XT] *xtnbreg* — Fixed-effects, random-effects, & population-averaged negative binomial models

[XT] *xtset* — Declare data to be panel data

[ME] *mepoisson* — Multilevel mixed-effects Poisson regression

[ME] *meqrpoisson* — Multilevel mixed-effects Poisson regression (QR decomposition)

[MI] *estimation* — Estimation commands for use with mi estimate

[R] *poisson* — Poisson regression

[U] 20 Estimation and postestimation commands
Description

The following postestimation commands are available after *xtpoisson*:

<table>
<thead>
<tr>
<th>Command</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>contrast</td>
<td>contrasts and ANOVA-style joint tests of estimates</td>
</tr>
<tr>
<td>estat ic</td>
<td>Akaike’s and Schwarz’s Bayesian information criteria (AIC and BIC)</td>
</tr>
<tr>
<td>estat summarize</td>
<td>summary statistics for the estimation sample</td>
</tr>
<tr>
<td>estat vce</td>
<td>variance–covariance matrix of the estimators (VCE)</td>
</tr>
<tr>
<td>estimates</td>
<td>cataloging estimation results</td>
</tr>
<tr>
<td>forecast</td>
<td>dynamic forecasts and simulations</td>
</tr>
<tr>
<td>hausman</td>
<td>Hausman’s specification test</td>
</tr>
<tr>
<td>lincom</td>
<td>point estimates, standard errors, testing, and inference for linear</td>
</tr>
<tr>
<td></td>
<td>combinations of coefficients</td>
</tr>
<tr>
<td>lrtest</td>
<td>likelihood-ratio test</td>
</tr>
<tr>
<td>margins</td>
<td>marginal means, predictive margins, marginal effects, and average</td>
</tr>
<tr>
<td></td>
<td>marginal effects</td>
</tr>
<tr>
<td>marginsplot</td>
<td>graph the results from margins (profile plots, interaction plots, etc.)</td>
</tr>
<tr>
<td>nlcom</td>
<td>point estimates, standard errors, testing, and inference for nonlinear</td>
</tr>
<tr>
<td></td>
<td>combinations of coefficients</td>
</tr>
<tr>
<td>predict</td>
<td>predictions, residuals, influence statistics, and other diagnostic measures</td>
</tr>
<tr>
<td>predictnl</td>
<td>point estimates, standard errors, testing, and inference for generalized</td>
</tr>
<tr>
<td></td>
<td>predictions</td>
</tr>
<tr>
<td>pwcompare</td>
<td>pairwise comparisons of estimates</td>
</tr>
<tr>
<td>test</td>
<td>Wald tests of simple and composite linear hypotheses</td>
</tr>
<tr>
<td>testnl</td>
<td>Wald tests of nonlinear hypotheses</td>
</tr>
</tbody>
</table>

1 `estat ic` is not appropriate after *xtpoisson*, pa.
2 `forecast` is not appropriate with *mi* estimation results.

Syntax for predict

**Random-effects (RE) and fixed-effects (FE) models**

```
predict [type] newvar [if] [in] [, RE/FE_statistic nooffset]
```

**Population-averaged (PA) model**

```
predict [type] newvar [if] [in] [, PA_statistic nooffset]
```
RE/FE.statistic  Description

Main
xb  linear prediction; the default
stdp standard error of the linear prediction
nu0 predicted number of events; assumes fixed or random effect is zero
iru0 predicted incidence rate; assumes fixed or random effect is zero
pr0(n) probability Pr(y_j = n) assuming the random effect is zero;
     only allowed after xtpoisson, re
pr0(a,b) probability Pr(a ≤ y_j ≤ b) assuming the random effect is zero;
     only allowed after xtpoisson, re

PA.statistic  Description

Main
mu  predicted number of events; considers the offset(); the default
rate predicted number of events
xb  linear prediction
pr(n) probability Pr(y_j = n)
pr(a,b) probability Pr(a ≤ y_j ≤ b)
stdp standard error of the linear prediction
score first derivative of the log likelihood with respect to \( x_j \beta \)

These statistics are available both in and out of sample; type predict ... if e(sample) ... if wanted only for the estimation sample.

Menu for predict

Statistics > Postestimation > Predictions, residuals, etc.

Options for predict

Main
xb calculates the linear prediction. This is the default for the random-effects and fixed-effects models.
mu and rate both calculate the predicted number of events. mu takes into account the offset(), and
rate ignores those adjustments. mu and rate are equivalent if you did not specify offset(). mu
is the default for the population-averaged model.

stdp calculates the standard error of the linear prediction.

nu0 calculates the predicted number of events, assuming a zero random or fixed effect.
iru0 calculates the predicted incidence rate, assuming a zero random or fixed effect.
pr0(n) calculates the probability Pr(y_j = n) assuming the random effect is zero, where n is a
nonnegative integer that may be specified as a number or a variable (only allowed after xtpoisson, re).
\texttt{pr0(a,b)} calculates the probability $\Pr(a \leq y_j \leq b)$ assuming the random effect is zero, where $a$ and $b$ are nonnegative integers that may be specified as numbers or variables (only allowed after \texttt{xtpoisson, re});

$b$ missing ($b \geq .$) means $+\infty$;

\texttt{pr0(20,.)} calculates $\Pr(y_j \geq 20)$;

\texttt{pr0(20,b)} calculates $\Pr(y_j \geq 20)$ in observations for which $b \geq .$ and calculates $\Pr(20 \leq y_j \leq b)$ elsewhere.

\texttt{pr0(.b)} produces a syntax error. A missing value in an observation of the variable $a$ causes a missing value in that observation for \texttt{pr0(a,b)}.

\texttt{pr(n)} calculates the probability $\Pr(y_j = n)$, where $n$ is a nonnegative integer that may be specified as a number or a variable (only allowed after \texttt{xtpoisson, pa}).

\texttt{pr(a,b)} calculates the probability $\Pr(a \leq y_j \leq b)$ (only allowed after \texttt{xtpoisson, pa}). The syntax for this option is analogous to that used with \texttt{pr0(a,b)}.

\texttt{score} calculates the equation-level score, $u_j = \frac{\partial \ln L_j(x_j \beta)}{\partial (x_j \beta)}$.

\texttt{nooffset} is relevant only if you specified \texttt{offset(varname)} for \texttt{xtpoisson}. It modifies the calculations made by \texttt{predict} so that they ignore the offset variable; the linear prediction is treated as $x_{it} \beta$ rather than $x_{it} \beta + \text{offset}_{it}$.

### Remarks and examples

#### Example 1

In example 1 of [XT] \texttt{xtpoisson}, we fit a random-effects model of the number of accidents experienced by five different types of ships on the basis of when the ships were constructed and operated. Here we obtain the predicted number of accidents for each observation, assuming that the random effect for each panel is zero:

\begin{verbatim}
. use http://www.stata-press.com/data/r13/ships
. xtpoisson accident op_75_79 co_65_69 co_70_74 co_75_79, exposure(service) irr
   (output omitted)
. predict n_acc, nu0
   (6 missing values generated)
. summarize n_acc

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>n_acc</td>
<td>34</td>
<td>13.52307</td>
<td>23.15885</td>
<td>.0617592</td>
<td>83.31905</td>
</tr>
</tbody>
</table>
\end{verbatim}

From these results, you may be tempted to conclude that some types of ships are safe, with a predicted number of accidents close to zero, whereas others are dangerous, because 1 observation is predicted to have more than 83 accidents.

However, when we fit the model, we specified the \texttt{exposure(service)} option. The variable \texttt{service} records the total number of months of operation for each type of ship constructed in and operated during particular years. Because ships experienced different utilization rates and thus were exposed to different levels of accident risk, we included \texttt{service} as our exposure variable. When comparing different types of ships, we must therefore predict the number of accidents, assuming that all ships faced the same exposure to risk. To do that, we use the \texttt{iru0} option with \texttt{predict}:
The probabilities calculated using the `pr0(n)` option are the probability \( \Pr(y_{it} = n) \) for a RE model assuming the random effect is zero. Define \( \mu_{it} = \exp(x_{it} \beta + \text{offset}_{it}) \). The probabilities in `pr0(n)` are calculated as the probability that \( y_{it} = n \), where \( y_{it} \) has a Poisson distribution with mean \( \mu_{it} \). Specifically,

\[
\Pr(y_{it} = n) = (n!)^{-1} \exp(-\mu_{it})(\mu_{it})^{n}
\]

Probabilities calculated using the `pr(n)` option after fitting a PA model are also calculated as described above.

**Also see**

[XT] `xtpoisson` — Fixed-effects, random-effects, and population-averaged Poisson models

[U] 20 Estimation and postestimation commands
Syntax

Random-effects (RE) model

```
xtprobit depvar [indepvars] [if] [in] [weight] [re RE_options]
```

Population-averaged (PA) model

```
xtprobit depvar [indepvars] [if] [in] [weight] [pa [PA_options]]
```

**RE_options**

<table>
<thead>
<tr>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
</tr>
<tr>
<td><code>noconstant</code></td>
</tr>
<tr>
<td><code>re</code></td>
</tr>
<tr>
<td><code>offset(varname)</code></td>
</tr>
<tr>
<td><code>constraints(constraints)</code></td>
</tr>
<tr>
<td><code>collinear</code></td>
</tr>
<tr>
<td><code>asis</code></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SE/Robust</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>vce(vcetype)</code></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Reporting</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>level(#)</code></td>
</tr>
<tr>
<td><code>noskip</code></td>
</tr>
<tr>
<td><code>noconsreport</code></td>
</tr>
<tr>
<td><code>display_options</code></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Integration</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>intmethod(intmethod)</code></td>
</tr>
<tr>
<td><code>intpoints(#)</code></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Maximization</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>maximize_options</code></td>
</tr>
<tr>
<td><code>coeflegend</code></td>
</tr>
</tbody>
</table>
## \textit{xtprobit} — Random-effects and population-averaged probit models

### PA\_options

<table>
<thead>
<tr>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>\textbf{Model}</td>
</tr>
<tr>
<td>\textbf{noconstant}</td>
</tr>
<tr>
<td>\textbf{pa}</td>
</tr>
<tr>
<td>\textbf{offset(\textit{varname})}</td>
</tr>
<tr>
<td>\textbf{asis}</td>
</tr>
<tr>
<td>\textbf{Correlation}</td>
</tr>
<tr>
<td>\textbf{corr(\textit{correlation})}</td>
</tr>
<tr>
<td>\textbf{force}</td>
</tr>
<tr>
<td>\textbf{SE/Robust}</td>
</tr>
<tr>
<td>\textbf{vce(\textit{vcetype})}</td>
</tr>
<tr>
<td>\textbf{ncmp}</td>
</tr>
<tr>
<td>\textbf{scale(\textit{parm})}</td>
</tr>
<tr>
<td>\textbf{Reporting}</td>
</tr>
<tr>
<td>\textbf{level(#)}</td>
</tr>
<tr>
<td>\textbf{display_options}</td>
</tr>
<tr>
<td>\textbf{Optimization}</td>
</tr>
<tr>
<td>\textbf{optimize_options}</td>
</tr>
<tr>
<td>\textbf{coeflegend}</td>
</tr>
</tbody>
</table>

### correlation

<table>
<thead>
<tr>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>\textbf{exchangeable}</td>
</tr>
<tr>
<td>\textbf{independent}</td>
</tr>
<tr>
<td>\textbf{unstructured}</td>
</tr>
<tr>
<td>\textbf{fixed matname}</td>
</tr>
<tr>
<td>\textbf{ar #}</td>
</tr>
<tr>
<td>\textbf{stationary #}</td>
</tr>
<tr>
<td>\textbf{nonstationary #}</td>
</tr>
</tbody>
</table>

A panel variable must be specified. For \textit{xtprobit}, \textit{pa}, correlation structures other than \textit{exchangeable} and \textit{independent} require that a time variable also be specified. Use \textit{xtset}; see [XT] \textit{xtset}.

\textit{indepvars} may contain factor variables; see [U] 11.4.3 Factor variables.

\textit{depvar} and \textit{indepvars} may contain time-series operators; see [U] 11.4.4 Time-series varlists.

\textit{by}, \textit{mi estimate}, and \textit{statsby} are allowed; see [U] 11.1.10 Prefix commands. \textit{fp} is allowed for the random-effects model.

\textit{vce(bootstrap)} and \textit{vce(jackknife)} are not allowed with the \textit{mi estimate} prefix; see [MI] \textit{mi estimate}.

\textit{iweights}, \textit{fweights}, and \textit{pweights} are allowed for the population-averaged model, and \textit{iweights} are allowed for the random-effects model; see [U] 11.1.6 weight. Weights must be constant within panel.

\textit{coeflegend} does not appear in the dialog box.

See [U] 20 Estimation and postestimation commands for more capabilities of estimation commands.
Menu
Statistics > Longitudinal/panel data > Binary outcomes > Probit regression (RE, PA)

Description

*_xtprobit_* fits random-effects and population-averaged probit models. There is no command for a conditional fixed-effects model, as there does not exist a sufficient statistic allowing the fixed effects to be conditioned out of the likelihood. Unconditional fixed-effects probit models may be fit with the *probit* command with indicator variables for the panels. However, unconditional fixed-effects estimates are biased.

By default, the population-averaged model is an equal-correlation model; *xtprobit, pa* assumes *corr(exchangeable)*. See [*XT* *xtgee*] for information about how to fit other population-averaged models.

See [*R* *logistic*] for a list of related estimation commands.

Options for RE model

---

**Model**

`noconstant;` see [*R* *estimation options*].

`re` requests the random-effects estimator. `re` is the default if neither `re` nor `pa` is specified.

`offset(varname), constraints(constraints), collinear;` see [*R* *estimation options*].

`asis` forces retention of perfect predictor variables and their associated, perfectly predicted observations and may produce instabilities in maximization; see [*R* *probit*].

---

**SE/Robust**

`vce(vcetype)` specifies the type of standard error reported, which includes types that are derived from asymptotic theory (`oim`), that are robust to some kinds of misspecification (`robust`), that allow for intragroup correlation (`cluster clustvar`), and that use bootstrap or jackknife methods (`bootstrap`, `jackknife`); see [*XT* *vce_options*].

Specifying `vce(robust)` is equivalent to specifying `vce(cluster panelvar)`; see _xtprobit, re and the robust VCE estimator_ in *Methods and formulas*.

---

**Reporting**

`level(#), noskip;` see [*R* *estimation options*].

`nocnsreport;` see [*R* *estimation options*].

`display_options:` `noomitted, vsquish, noemptycells, baselevels, allbaselevels, nofvlabel, fvwrap(#), fvwrapon(style), cformat(%fmt), pformat(%fmt), sformat(%fmt),` and `nolstretch;` see [*R* *estimation options*].

---

**Integration**

`intmethod(intmethod), intpoints(#);` see [*R* *estimation options*].
Maximization

`maximize_options: difficult, technique(algorithm_spec), iterate(#), [no]log, trace, gradient, showstep, hessian, showtolerance, tolerance(#), ltolerance(#), nrtolerance(#), nonrtolerance, and from(init_specs);` see [R] `maximize`. These options are seldom used.

The following option is available with `xtprobit` but is not shown in the dialog box:
`coeflegend`; see [R] `estimation options`.

Options for PA model

`noconstant`; see [R] `estimation options`.
`pa` requests the population-averaged estimator.
`offset(varname)`; see [R] `estimation options`.
`asis` forces retention of perfect predictor variables and their associated, perfectly predicted observations and may produce instabilities in maximization; see [R] `probit`.

`corr(correlation)` specifies the within-panel correlation structure; the default corresponds to the equal-correlation model, `corr(exchangeable)`.

When you specify a correlation structure that requires a lag, you indicate the lag after the structure’s name with or without a blank; for example, `corr(ar 1)` or `corr(ar1)`.

If you specify the fixed correlation structure, you specify the name of the matrix containing the assumed correlations following the word `fixed`, for example, `corr(fixed myr)`.
`force` specifies that estimation be forced even though the time variable is not equally spaced. This is relevant only for correlation structures that require knowledge of the time variable. These correlation structures require that observations be equally spaced so that calculations based on lags correspond to a constant time change. If you specify a time variable indicating that observations are not equally spaced, the (time dependent) model will not be fit. If you also specify `force`, the model will be fit, and it will be assumed that the lags based on the data ordered by the time variable are appropriate.

`vce(vcetype)` specifies the type of standard error reported, which includes types that are derived from asymptotic theory (`conventional`), that are robust to some kinds of misspecification (`robust`), and that use bootstrap or jackknife methods (`bootstrap, jackknife`); see [XT] `vce_options`.

`vce(conventional)`, the default, uses the conventionally derived variance estimator for generalized least-squares regression.
`nmp, scale(x2 | dev | phi | #)`; see [XT] `vce_options`.

`level(#)`; see [R] `estimation options`.  

---

**xtprobit — Random-effects and population-averaged probit models**

---

**xtprobit — Random-effects and population-averaged probit models**
display_options: noomitted, vsquish, noemptycells, baselevels, allbaselevels, nofvlabel, fvwrap(#), fvwrappon(style), cformat(\%fmt), pformat(\%fmt), sformat(\%fmt), and nolstretch; see [R] estimation options.

Optimization

Optimization options control the iterative optimization process. These options are seldom used.

iterate(#) specifies the maximum number of iterations. When the number of iterations equals #, the optimization stops and presents the current results, even if convergence has not been reached. The default is iterate(100).

tolerance(#) specifies the tolerance for the coefficient vector. When the relative change in the coefficient vector from one iteration to the next is less than or equal to #, the optimization process is stopped. tolerance(1e-6) is the default.

nolog suppresses display of the iteration log.

trace specifies that the current estimates be printed at each iteration.

The following option is available with xtprobit but is not shown in the dialog box: coeflegend; see [R] estimation options.

Remarks and examples

xtprobit is a convenience command for obtaining the population-averaged model. Typing

    . xtprobit ... pa ...

is equivalent to typing

    . xtgee ..., ..., family(binomial) link(probit) corr(exchangeable)

See also [XT] xtgee for information about xtprobit.

By default or when re is specified, xtprobit fits via maximum likelihood the random-effects model

\[ \Pr(y_{it} \neq 0|x_{it}) = \Phi(x_{it}\beta + \nu_i) \]

for \( i = 1, \ldots, n \) panels, where \( t = 1, \ldots, n_i \), \( \nu_i \) are i.i.d., \( N(0, \sigma^2_{\nu}) \), and \( \Phi \) is the standard normal cumulative distribution function.

Underlying this model is the variance components model

\[ y_{it} \neq 0 \iff x_{it}\beta + \nu_i + \epsilon_{it} > 0 \]

where \( \epsilon_{it} \) are i.i.d. Gaussian distributed with mean zero and variance \( \sigma^2_\epsilon = 1 \), independently of \( \nu_i \).

Example 1: Random-effects model

We are studying unionization of women in the United States and are using the union dataset; see [XT] xt. We wish to fit a random-effects model of union membership:
. use http://www.stata-press.com/data/r13/union
(NLS Women 14-24 in 1968)
. xtprobit union age grade i.not_smsa south##c.year

Fitting comparison model:
Iteration 0:  log likelihood =  -13864.23
Iteration 1:  log likelihood =  -13545.541
Iteration 2:  log likelihood =  -13544.385
Iteration 3:  log likelihood =  -13544.385

Fitting full model:
rho   =  0.0  log likelihood =  -13544.385
rho   =  0.1  log likelihood =  -12237.655
rho   =  0.2  log likelihood =  -11590.282
rho   =  0.3  log likelihood =  -11211.185
rho   =  0.4  log likelihood =  -10981.319
rho   =  0.5  log likelihood =  -10852.793
rho   =  0.6  log likelihood =  -10808.759
rho   =  0.7  log likelihood =  -10865.57
Iteration 0:  log likelihood =  -10807.712
Iteration 1:  log likelihood =  -10599.332
Iteration 2:  log likelihood =  -10552.287
Iteration 3:  log likelihood =  -10552.225
Iteration 4:  log likelihood =  -10552.225

Random-effects probit regression
Number of obs = 26200
Group variable: idcode  Number of groups = 4434
Random effects u_i ~ Gaussian
Obs per group:  min = 1
               avg = 5.9
               max = 12
Integration method: mvaghermite
Integration points = 12
Log likelihood = -10552.225
Wald chi2(6)   =   220.91
Prob > chi2    =   0.0000

|           | Coef.  | Std. Err. | z      | P>|z|  | [95% Conf. Interval] |
|-----------|--------|-----------|--------|-----|----------------------|
| union     | age    | .0082967  | .0084599 | 0.98 | 0.327   | -.0082843 , 0.0248778 |
|           | grade  | .0482731  | .0099469 | 4.85 | 0.000   | .0287776 , .0677686  |
|           | 1.not_smsa | -.139657  | .0460548 | -3.03 | 0.002   | -.2299227 , -.0493913 |
|           | 1.south | -1.584394 | .358473  | -4.42 | 0.000   | -.2869899 , -.8818002 |
|           | year   | -.0039854 | .0088399 | -0.45 | 0.652   | -.0213113 , .0133406  |
|           | south#c.year | .0134017  | .0044622 | 3.00 | 0.003   | .0046559 , .0221475  |
|           | _cons  | -1.668202 | .4751819 | -3.51 | 0.000   | -.2599542 , -.7368628 |
| /lnsig2u  |        | .6103616  | .0458783 | -3.51 | 0.000   | .5204418 , .7002814  |
| sigma_u   |        | 1.35687   | .0311255 | 1.297217 | 4.19267 | 1.297217 , 1.419267 |
| rho       |        | .6480233  | .0104643 | 6.272511 | 6.682502 |

Likelihood-ratio test of rho=0:  chibar2(01) = 5984.32 Prob >= chibar2 = 0.000

The output includes the additional panel-level variance component, which is parameterized as the log of the variance $\ln(\sigma^2_{\nu})$ (labeled $\lnsig2u$ in the output). The standard deviation $\sigma_{\nu}$ is also included in the output (labeled $sigma_u$) together with $\rho$ (labeled rho), where

$$\rho = \frac{\sigma^2_{\nu}}{\sigma^2_{\nu} + 1}$$
which is the proportion of the total variance contributed by the panel-level variance component.

When \( \rho \) is zero, the panel-level variance component is unimportant, and the panel estimator is not different from the pooled estimator. A likelihood-ratio test of this is included at the bottom of the output. This test formally compares the pooled estimator (probit) with the panel estimator.

#### Technical note

The random-effects model is calculated using quadrature, which is an approximation whose accuracy depends partially on the number of integration points used. We can use the quadchk command to see if changing the number of integration points affects the results. If the results change, the quadrature approximation is not accurate given the number of integration points. Try increasing the number of integration points using the `intpoints()` option and run `quadchk` again. Do not attempt to interpret the results of estimates when the coefficients reported by `quadchk` differ substantially.

```
. quadchk, nooutput
Refitting model intpoints() = 8
Refitting model intpoints() = 16

<table>
<thead>
<tr>
<th>Quadrature check</th>
<th>Fitted quadrature</th>
<th>Comparison quadrature</th>
<th>Comparison quadrature</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>12 points</td>
<td>8 points</td>
<td>16 points</td>
</tr>
<tr>
<td>Log likelihood</td>
<td>-10552.225</td>
<td>-10554.496</td>
<td>-10552.399</td>
</tr>
<tr>
<td>union:</td>
<td>.00829671</td>
<td>.00828745</td>
<td>.00831488</td>
</tr>
<tr>
<td>age</td>
<td>-.9265e-06</td>
<td>.00011167</td>
<td>.0018987</td>
</tr>
<tr>
<td>union:</td>
<td>.0482731</td>
<td>.04860277</td>
<td>.04826287</td>
</tr>
<tr>
<td>grade</td>
<td>.00329677</td>
<td>-.0001023</td>
<td>-.00021188</td>
</tr>
<tr>
<td>union:</td>
<td>-.13965702</td>
<td>-.1405741</td>
<td>-.13953521</td>
</tr>
<tr>
<td>1.not_smsa</td>
<td>-.0091739</td>
<td>-.00012181</td>
<td>-.0087218</td>
</tr>
<tr>
<td>union:</td>
<td>-1.5843944</td>
<td>-1.5909857</td>
<td>-1.5843375</td>
</tr>
<tr>
<td>1.south</td>
<td>-.00659135</td>
<td>.00005689</td>
<td>-.0003591</td>
</tr>
<tr>
<td>union:</td>
<td>-.00398535</td>
<td>-.00397811</td>
<td>-.00400181</td>
</tr>
<tr>
<td>year</td>
<td>7.237e-06</td>
<td>-.0003591</td>
<td>-.00012181</td>
</tr>
<tr>
<td>union:</td>
<td>.01340169</td>
<td>.01344457</td>
<td>.01340388</td>
</tr>
<tr>
<td>1.south#c.r</td>
<td>.00004288</td>
<td>2.193e-06</td>
<td>.001636</td>
</tr>
<tr>
<td>union:</td>
<td>-1.6682022</td>
<td>-1.6757524</td>
<td>-1.6665327</td>
</tr>
<tr>
<td>_cons</td>
<td>.00452957</td>
<td>-.00100077</td>
<td>-.00100077</td>
</tr>
<tr>
<td>lnSIG2U:</td>
<td>.61036163</td>
<td>.61780789</td>
<td>.60974814</td>
</tr>
<tr>
<td>_cons</td>
<td>.00744626</td>
<td>-.00061349</td>
<td>-.00100077</td>
</tr>
<tr>
<td></td>
<td>.01219976</td>
<td>-.00100077</td>
<td>-.00100077</td>
</tr>
</tbody>
</table>
```
The results obtained for 12 quadrature points were closer to the results for 16 points than to the results for eight points. Although the relative and absolute differences are a bit larger than we would like, they are not large. We can increase the number of quadrature points with the `intpoints()` option; if we choose `intpoints(20)` and do another `quadchk` we will get acceptable results, with relative differences around 0.01%.

This is not the case if we use nonadaptive quadrature. Then the results we obtain are

\[
\text{. xtprobit union age grade i.not_smsa south##c.year, intmethod(ghermite)}
\]

Fitting comparison model:

<table>
<thead>
<tr>
<th>Iteration</th>
<th>log likelihood</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>-13864.23</td>
</tr>
<tr>
<td>1</td>
<td>-13545.541</td>
</tr>
<tr>
<td>2</td>
<td>-13544.385</td>
</tr>
<tr>
<td>3</td>
<td>-13544.385</td>
</tr>
</tbody>
</table>

Fitting full model:

<table>
<thead>
<tr>
<th>rho</th>
<th>log likelihood</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>-13544.385</td>
</tr>
<tr>
<td>0.1</td>
<td>-12237.655</td>
</tr>
<tr>
<td>0.2</td>
<td>-11590.282</td>
</tr>
<tr>
<td>0.3</td>
<td>-11211.185</td>
</tr>
<tr>
<td>0.4</td>
<td>-10981.319</td>
</tr>
<tr>
<td>0.5</td>
<td>-10852.793</td>
</tr>
<tr>
<td>0.6</td>
<td>-10808.759</td>
</tr>
<tr>
<td>0.7</td>
<td>-10865.57</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Iteration</th>
<th>log likelihood</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>-10808.759</td>
</tr>
<tr>
<td>1</td>
<td>-10594.349</td>
</tr>
<tr>
<td>2</td>
<td>-10560.913</td>
</tr>
<tr>
<td>3</td>
<td>-10560.876</td>
</tr>
<tr>
<td>4</td>
<td>-10560.876</td>
</tr>
</tbody>
</table>

Random-effects probit regression

| Coef.         | Std. Err. | z     | P>|z| | [95% Conf. Interval] |
|---------------|-----------|-------|-------|----------------------|
| union age     | 0.0093488 | 0.0083385 | 1.12 | 0.262 | -0.0069945 to 0.025692 |
| grade         | 0.0488014 | 0.0101168 | 4.82 | 0.000 | 0.0289728 to 0.068632 |
| 1.not_smsa    | -0.1364862 | 0.0462831 | -2.95 | 0.003 | -0.2271995 to -0.045773 |
| 1.south       | -1.592711 | 0.3576715 | -4.45 | 0.000 | -2.293734 to -0.8916877 |
| year          | -0.0053723 | 0.0087219 | -0.62 | 0.538 | -0.0224668 to 0.0117223 |
| south#c.year 1| 0.0136764 | 0.0044532 | 3.07 | 0.002 | 0.0049482 to 0.0224046 |
| _cons         | -1.575539 | 0.4639881 | -3.40 | 0.001 | -2.484939 to -0.6661388 |
| /lnsig2u      | 0.5615976 | 0.0432021 | 12.78 | 0.000 | 0.476923 to 0.6462722 |
| sigma_u       | 1.324187 | 0.0286038 | 45.92 | 0.000 | 1.269295 to 1.381453 |
| rho           | 0.6368221 | 0.0099918 | 63.75 | 0.000 | 0.617021 to 0.6561699 |

Likelihood-ratio test of rho=0: chibar2(01) = 5967.02 Prob >= chibar2 = 0.000

We now check the stability of the quadrature technique for this nonadaptive quadrature model. We expect it to be less stable.
Once again, the results obtained for 12 quadrature points were closer to the results for 16 points than to the results for eight points. However, here the convergence point seems to be sensitive to the number of quadrature points, so we should not trust these results. We should increase the number of quadrature points with the `intpoints()` option and then use `quadchk` again. We should not use the results of a random-effects specification when there is evidence that the numeric technique for calculating the model is not stable (as shown by `quadchk`).

Generally, the relative differences in the coefficients should not change by more than 1% if the quadrature technique is stable. See [XT] `quadchk` for details. Increasing the number of quadrature points can often improve the stability, and for models with high `rho` we may need many. We can also switch between adaptive and nonadaptive quadrature. As a rule, adaptive quadrature, which is the default integration method, is much more flexible and robust.

<table>
<thead>
<tr>
<th>Quadrature check</th>
<th>Fitted quadrature</th>
<th>Comparison quadrature</th>
<th>Comparison quadrature</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>12 points</td>
<td>8 points</td>
<td>16 points</td>
</tr>
<tr>
<td>Log likelihood</td>
<td>-10560.876</td>
<td>-10574.239</td>
<td>-10555.792</td>
</tr>
<tr>
<td></td>
<td>13.362535</td>
<td>5.0839579</td>
<td>.00126529</td>
</tr>
<tr>
<td></td>
<td>-.0004814</td>
<td></td>
<td></td>
</tr>
<tr>
<td>union:</td>
<td>.00934876</td>
<td>.01264615</td>
<td>.00731888</td>
</tr>
<tr>
<td>age</td>
<td>.0032974</td>
<td>-.00202987</td>
<td></td>
</tr>
<tr>
<td></td>
<td>.35270966</td>
<td>-.21712744</td>
<td></td>
</tr>
<tr>
<td>union:</td>
<td>.04880139</td>
<td>.05710089</td>
<td>.04432417</td>
</tr>
<tr>
<td>grade</td>
<td>.00829951</td>
<td>-.00447722</td>
<td></td>
</tr>
<tr>
<td></td>
<td>.17006703</td>
<td>-.09174372</td>
<td></td>
</tr>
<tr>
<td>union:</td>
<td>-.13648624</td>
<td>-.13327724</td>
<td>-.14094541</td>
</tr>
<tr>
<td>1.not_smsa</td>
<td>-.003209</td>
<td>-.00445917</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-.0235115</td>
<td>.03267123</td>
<td></td>
</tr>
<tr>
<td>union:</td>
<td>-1.592711</td>
<td>-1.5275627</td>
<td>-1.6059143</td>
</tr>
<tr>
<td>1.south</td>
<td>.06514823</td>
<td>-.01320331</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-.04090399</td>
<td>.00828983</td>
<td></td>
</tr>
<tr>
<td>union:</td>
<td>-.00537226</td>
<td>-.00867673</td>
<td>-.00307042</td>
</tr>
<tr>
<td>year</td>
<td>-.00330447</td>
<td>.00230184</td>
<td></td>
</tr>
<tr>
<td></td>
<td>.61509968</td>
<td>-.4284678</td>
<td></td>
</tr>
<tr>
<td>union:</td>
<td>.01367641</td>
<td>.01278071</td>
<td>.01369009</td>
</tr>
<tr>
<td>1.south#c._r</td>
<td>-.0008957</td>
<td>.00001368</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-.06549266</td>
<td>.00100054</td>
<td></td>
</tr>
<tr>
<td>union:</td>
<td>-1.5755388</td>
<td>-1.4886646</td>
<td>-1.6505526</td>
</tr>
<tr>
<td>_cons</td>
<td>.08667418</td>
<td>-.0750138</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-.0550124</td>
<td>.04761152</td>
<td></td>
</tr>
<tr>
<td>ln.sig2u:</td>
<td>.56159763</td>
<td>.49290978</td>
<td>.58068904</td>
</tr>
<tr>
<td>_cons</td>
<td>-.06868786</td>
<td>.0190914</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-.12230795</td>
<td>.03399481</td>
<td></td>
</tr>
</tbody>
</table>
Because the `xtprobit`, re likelihood function is calculated by Gauss–Hermite quadrature, on large problems the computations can be slow. Computation time is roughly proportional to the number of points used for the quadrature.

> Example 2: Equal-correlation model

As an alternative to the random-effects specification, we can fit an equal-correlation probit model:

```
. xtprobit union age grade i.not_smsa south##c.year, pa
```

```
Iteration 1: tolerance =  .12544249
Iteration 2: tolerance =  .0034686
Iteration 3: tolerance =  .00017448
Iteration 4: tolerance =  8.382e-06
Iteration 5: tolerance =  3.997e-07
```

GEE population-averaged model Number of obs = 26200
Group variable: idcode Number of groups = 4434
Link: probit Obs per group: min = 1
Family: binomial avg = 5.9
Correlation: exchangeable max = 12

Wald chi2(6) = 242.57
Scale parameter: 1 Prob > chi2 = 0.0000

|                | Coef.     | Std. Err. | z      | P>|z| | [95% Conf. Interval] |
|----------------|-----------|-----------|--------|------|----------------------|
| union          |           |           |        |      |                      |
| age            | .0089699  | .0053208  | 1.69   | 0.092| -.0014586 .0193985   |
| grade          | .0333174  | .0062352  | 5.34   | 0.000| .0210966 .0455382    |
| 1.not_smsa     | -.0715717 | .027543   | -2.60  | 0.009| -.1255551 -.0175884  |
| 1.south        | -1.017368 | .207931   | -4.89  | 0.000| -1.424905 -.6098308  |
| year           | -.0062708 | .0055314  | -1.13  | 0.257| -.0171122 .0045706   |

| south#c.year   |           |           |        |      |                      |
| 1              | .0086294  | .00258    | 3.34   | 0.001| .0035727 .013686     |
| _cons          | -.8670997 | .294771   | -2.94  | 0.003| -1.44484 -.2893592   |

> Example 3: Population-averaged model

In example 3 of `R` probit, we showed the above results and compared them with probit, vce(cluster id). `xtprobit` with the pa option allows a vce(robust) option (the random-effects estimator does not allow the vce(robust) specification), so we can obtain the population-averaged probit estimator with the robust variance calculation by typing
. xtprobit union age grade i.not_smsa south##c.year, pa vce(robust) nolog

GEE population-averaged model

Number of obs = 26200
Group variable: idcode Number of groups = 4434
Link: probit Obs per group: min = 1
Family: binomial avg = 5.9
Correlation: exchangeable max = 12
Wald chi2(6) = 156.33
Scale parameter: 1 Prob > chi2 = 0.0000

(Std. Err. adjusted for clustering on idcode)

|                | Coef.  | Std. Err. |   z   | P>|z| | [95% Conf. Interval] |
|----------------|--------|-----------|------|-----|----------------------|
| union          |        |           |      |     |                      |
| age            | 0.0089699 | 0.0051169 | 1.75 | 0.080 | -0.001059  | 0.0189988 |
| grade          | 0.0333174 | 0.0076425 | 4.36 | 0.000 | 0.0183383  | 0.0482965 |
| 1.not_smsa     | -0.0715717 | 0.0348659 | -2.05 | 0.040 | -0.1399076 | -0.0032359 |
| 1.south        | -1.017368 | 0.3026981 | -3.36 | 0.001 | -1.610645  | -0.4240906 |
| year           | -0.0062708 | 0.0055745 | -1.12 | 0.261 | -0.0171965 | 0.0046549 |
| south#c.year   |       |           |      |     |                      |
| 1              | 0.0086294 | 0.0037866 | 2.28 | 0.023 | 0.0012078  | 0.0160509 |
| _cons          | -0.8670997 | 0.3243959 | -2.67 | 0.008 | -1.502904  | -0.2312955 |

These standard errors are similar to those shown for probit, vce(cluster id) in [R] probit.

Example 4: Random-effects model with stable quadrature

In a previous example, we showed how quadchk indicated that the quadrature technique was numerically unstable. Here we present an example in which the quadrature is stable.

In this example, we have (synthetic) data on whether workers complain to managers at fast-food restaurants. The covariates are age (in years of the worker), grade (years of schooling completed by the worker), south (equal to 1 if the restaurant is located in the South), tenure (the number of years spent on the job by the worker), gender (of the worker), race (of the worker), income (in thousands of dollars by the restaurant), genderm (gender of the manager), burger (equal to 1 if the restaurant specializes in hamburgers), and chicken (equal to 1 if the restaurant specializes in chicken). The model is given by
. use http://www.stata-press.com/data/r13/chicken
. xtprobit complain age grade south tenure gender race income genderm burger 
> chicken, nolog
Random-effects probit regression Number of obs = 2763
Group variable: restaurant Number of groups = 500
Random effects u_i ~ Gaussian Obs per group: min = 3
Integration method: mvaghermite Integration points = 12
Wald chi2(10) = 126.59
Log likelihood = -1318.2088 Prob > chi2 = 0.0000

| complain  | Coef.  | Std. Err. | z    | P>|z|   | [95% Conf. Interval] |
|-----------|--------|-----------|------|-------|---------------------|
| age       | -.0430409 | 0.0130211 | -3.31 | 0.001 | -.0685617 -.01752 |
| grade     | .0330934  | 0.0264572 | 1.25  | 0.211 | -.0187618 .0849486 |
| south     | .1012     | 0.0707196 | 1.43  | 0.152 | -.037408 .2398079 |
| tenure    | -.0440079 | 0.0987099 | -0.45 | 0.656 | -.2374758 .14946 |
| gender    | .3318499  | 0.0601382 | 5.52  | 0.000 | .2139812 .4497185 |
| race      | .3417901  | 0.0382251 | 8.94  | 0.000 | .2668703 .4167098 |
| income    | -.0022702 | 0.0008885 | -2.56 | 0.011 | -.0040117 -.0005288 |
| genderm   | .0524577  | 0.0706582 | 0.74  | 0.458 | -.0860305 .1909459 |
| burger    | .0448931  | 0.0956151 | 0.47  | 0.639 | -.1425091 .2322953 |
| chicken   | .1904714  | 0.0953067 | 2.00  | 0.046 | .0936737 .3772691 |
| _cons     | -.2145311 | .0953067  | -2.14 | 0.034 | -.4034758 .074413 |
| /lnsig2u  | -1.704494 | .2502057  | -2.14 | 0.034 | -.4034758 .074413 |
| sigma_u   | .4264557  | .0533508  | .333723 | .5449563 |
| rho       | .1538793  | .0325769  | .1002105 | .2289765 |

Likelihood-ratio test of rho=0: chibar2(01) = 29.91 Prob >= chibar2 = 0.000

Again we would like to check the stability of the quadrature technique of the model before interpreting the results. Given the estimate of $\rho$ and the small size of the panels (between 3 and 8), we should find that the quadrature technique is numerically stable.
. quadchk, nooutput
Refitting model intpoints() = 8
Refitting model intpoints() = 16

<table>
<thead>
<tr>
<th>Quadrature check</th>
<th>Fitted quadrature</th>
<th>Comparison quadrature</th>
<th>Comparison quadrature</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>12 points</td>
<td>8 points</td>
<td>16 points</td>
</tr>
<tr>
<td>Log likelihood</td>
<td>-1318.2088</td>
<td>-1318.2088</td>
<td>-1318.2088</td>
</tr>
<tr>
<td>complain:</td>
<td>-0.04304086</td>
<td>-0.04304086</td>
<td>-0.04304086</td>
</tr>
<tr>
<td>age</td>
<td>-3.896e-10</td>
<td>-2.625e-12</td>
<td>9.051e-09</td>
</tr>
<tr>
<td>complain:</td>
<td>0.0330934</td>
<td>2.208e-11</td>
<td>2.673e-10</td>
</tr>
<tr>
<td>grade</td>
<td>10119998</td>
<td>2.369e-09</td>
<td>3.957e-11</td>
</tr>
<tr>
<td>complain:</td>
<td>-0.04400789</td>
<td>-0.04400793</td>
<td>-0.04400789</td>
</tr>
<tr>
<td>tenure</td>
<td>-3.362e-09</td>
<td>-2.205e-11</td>
<td>7.640e-08</td>
</tr>
<tr>
<td>complain:</td>
<td>0.33184986</td>
<td>3.190e-09</td>
<td>9.612e-09</td>
</tr>
<tr>
<td>gender</td>
<td>34179006</td>
<td>3.801e-09</td>
<td>1.112e-08</td>
</tr>
<tr>
<td>complain:</td>
<td>-0.00227021</td>
<td>-0.00227021</td>
<td>-0.00227021</td>
</tr>
<tr>
<td>income</td>
<td>-4.468e-11</td>
<td>-9.252e-13</td>
<td>1.968e-08</td>
</tr>
<tr>
<td>complain:</td>
<td>0.05245769</td>
<td>1.963e-09</td>
<td>3.742e-08</td>
</tr>
<tr>
<td>genderm</td>
<td>0.04489311</td>
<td>4.173e-10</td>
<td>9.296e-09</td>
</tr>
<tr>
<td>complain:</td>
<td>0.19047138</td>
<td>3.096e-09</td>
<td>1.625e-08</td>
</tr>
<tr>
<td>_cons</td>
<td>-0.21453112</td>
<td>-1.281e-08</td>
<td>-5.972e-08</td>
</tr>
<tr>
<td>ln.sig2u:</td>
<td>-1.7044935</td>
<td>-1.7044934</td>
<td>-1.7044935</td>
</tr>
</tbody>
</table>
The relative and absolute differences are all small between the default 12 quadrature points and the result with 16 points. We do not have any coefficients that have a large difference between the default 12 quadrature points and eight quadrature points.

We conclude that the quadrature technique is stable. Because the differences here are so small, we would plan on using and interpreting these results rather than trying to rerun with more quadrature points.

Stored results

_xtprobit, re_ stores the following in _e()_:  

Scalars  
- **e(N)**: number of observations  
- **e(N_g)**: number of groups  
- **e(N_cd)**: number of completely determined observations  
- **e(k)**: number of parameters  
- **e(k_aux)**: number of auxiliary parameters  
- **e(k_eq)**: number of equations in _e(b)_  
- **e(k_eq_model)**: number of equations in overall model test  
- **e(k_dv)**: number of dependent variables  
- **e(df_m)**: model degrees of freedom  
- **e(ll)**: log likelihood  
- **e(ll_0)**: log likelihood, constant-only model  
- **e(ll_c)**: log likelihood, comparison model  
- **e(chi2)**: \( \chi^2 \)  
- **e(chi2_c)**: \( \chi^2 \) for comparison test  
- **e(N_clust)**: number of clusters  
- **e(rho)**: \( \rho \)  
- **e(sigma_u)**: panel-level standard deviation  
- **e(n_quad)**: number of quadrature points  
- **e(g_min)**: smallest group size  
- **e(g_avg)**: average group size  
- **e(g_max)**: largest group size  
- **e(p)**: significance  
- **e(rank)**: rank of _e(V)_  
- **e(rank0)**: rank of _e(V)_ for constant-only model  
- **e(ic)**: number of iterations  
- **e(rc)**: return code  
- **e(converged)**: 1 if converged, 0 otherwise
xtprobit — Random-effects and population-averaged probit models

Macros

- `e(cmd)` – `xtprobit`
- `e(cmdline)` – command as typed
- `e(depvar)` – name of dependent variable
- `e(ivar)` – variable denoting groups
- `e(model)` – `re`
- `e(wtype)` – weight type
- `e(wexp)` – weight expression
- `e(title)` – title in estimation output
- `e(clustvar)` – name of cluster variable
- `e(offset)` – linear offset variable
- `e(chi2type)` – Wald or LR; type of model $\chi^2$ test
- `e(chi2_ct)` – Wald or LR; type of model $\chi^2$ test corresponding to `e(chi2)`
- `e(vce)` – vcetype specified in `vce()`
- `e(vcetype)` – title used to label Std. Err.
- `e(intmethod)` – integration method
- `e(distrib)` – Gaussian; the distribution of the random effect
- `e(which)` – `max` or `min`; whether optimizer is to perform maximization or minimization
- `e(ml_method)` – type of ml method
- `e(user)` – name of likelihood-evaluator program
- `e(technique)` – maximization technique
- `e(properties)` – `b` `V`
- `e(predict)` – program used to implement `predict`
- `e(asbalanced)` – factor variables `fvset` as `asbalanced`
- `e(asobserved)` – factor variables `fvset` as `asobserved`

Matrices

- `e(b)` – coefficient vector
- `e(Cns)` – constraints matrix
- `e(ilog)` – iteration log
- `e(gradient)` – gradient vector
- `e(V)` – variance–covariance matrix of the estimators
- `e(V_modelbased)` – model-based variance

Functions

- `e(sample)` – marks estimation sample
xtprobit, pa stores the following in e():

Scalars
- e(N) number of observations
- e(N_g) number of groups
- e(df_m) model degrees of freedom
- e(chi2) \( \chi^2 \)
- e(p) significance
- e(df_pear) degrees of freedom for Pearson \( \chi^2 \)
- e(chi2_dev) \( \chi^2 \) test of deviance
- e(chi2_dis) \( \chi^2 \) test of deviance dispersion
- e(deviance) deviance
- e(dispers) deviance dispersion
- e(phi) scale parameter
- e(g_min) smallest group size
- e(g_avg) average group size
- e(g_max) largest group size
- e(rank) rank of e(V)
- e(tol) target tolerance
- e(dif) achieved tolerance
- e(rc) return code

Macros
- e(cmd) xtgee
- e(cmd2) xtprobit
- e(cmdline) command as typed
- e(depvar) name of dependent variable
- e(ivar) variable denoting groups
- e(tvar) variable denoting time within groups
- e(model) pa
- e(family) binomial
- e(link) probit; link function
- e(corr) correlation structure
- e(scale) x2, dev, phi, or #; scale parameter
- e(type) weight type
- e(wexp) weight expression
- e(offset) linear offset variable
- e(chi2type) Wald; type of model \( \chi^2 \) test
- e(vce) vcetype specified in vce()
- e(vcetype) title used to label Std. Err.
- e(nmp) nmp, if specified
- e(properties) b V
- e(predict) program used to implement predict
- e(marginsnotok) predictions disallowed by margins
- e(asbalanced) factor variables fvset as asbalanced
- e(asobserved) factor variables fvset as asobserved

Matrices
- e(b) coefficient vector
- e(Cns) constraints matrix
- e(R) estimated working correlation matrix
- e(V) variance–covariance matrix of the estimators
- e(V_modelbased) model-based variance

Functions
- e(sample) marks estimation sample

**Methods and formulas**

xtprobit reports the population-averaged results obtained by using xtgee, family(binomial) link(probit) to obtain estimates.
Assuming a normal distribution, \( N(0, \sigma^2_\nu) \), for the random effects \( \nu_i \)

\[
\Pr(y_{i1}, \ldots, y_{in_i} | x_{i1}, \ldots, x_{in_i}) = \int_{-\infty}^{\infty} \frac{e^{-\nu_i^2/2\sigma^2_\nu}}{\sqrt{2\pi}\sigma_\nu} \left\{ \prod_{t=1}^{n_i} F(y_{it}, x_{it} \beta + \nu_i) \right\} d\nu_i
\]

where

\[
F(y, z) = \begin{cases}
\Phi(z) & \text{if } y \neq 0 \\
1 - \Phi(z) & \text{otherwise}
\end{cases}
\]

where \( \Phi \) is the cumulative normal distribution.

The panel-level likelihood \( l_i \) is given by

\[
l_i = \int_{-\infty}^{\infty} \frac{e^{-\nu_i^2/2\sigma^2_\nu}}{\sqrt{2\pi}\sigma_\nu} \left\{ \prod_{t=1}^{n_i} F(y_{it}, x_{it} \beta + \nu_i) \right\} d\nu_i
\]

\[
= \int_{-\infty}^{\infty} g(y_{it}, x_{it}, \nu_i) d\nu_i
\]

This integral can be approximated with \( M \)-point Gauss–Hermite quadrature

\[
\int_{-\infty}^{\infty} e^{-x^2} h(x) dx \approx \sum_{m=1}^{M} w_m^* h(a_m^*)
\]

This is equivalent to

\[
\int_{-\infty}^{\infty} f(x) dx \approx \sum_{m=1}^{M} w_m^* \exp \left\{ (a_m^*)^2 \right\} f(a_m^*)
\]

where the \( w_m^* \) denote the quadrature weights and the \( a_m^* \) denote the quadrature abscissas. The log likelihood, \( L \), is the sum of the logs of the panel-level likelihoods \( l_i \).

The default approximation of the log likelihood is by adaptive Gauss–Hermite quadrature, which approximates the panel-level likelihood with

\[
l_i \approx \sqrt{2\hat{\sigma}_i} \sum_{m=1}^{M} w_m^* \exp \left\{ (a_m^*)^2 \right\} g(y_{it}, x_{it}, \sqrt{2\hat{\sigma}_i} a_m^* + \hat{\mu}_i)
\]

where \( \hat{\sigma}_i \) and \( \hat{\mu}_i \) are the adaptive parameters for panel \( i \). Therefore, with the definition of \( g(y_{it}, x_{it}, \nu_i) \), the total log likelihood is approximated by
\[
L \approx \sum_{i=1}^{n} w_i \log \left[ \sqrt{2\hat{\sigma}_i} \sum_{m=1}^{M} w_m^* \exp \left\{ (a_m^*)^2 \right\} \exp \left\{ -\left( \sqrt{2\hat{\sigma}_i} a_m^* + \hat{\mu}_i \right)^2 / 2\sigma_\nu^2 \right\} \prod_{t=1}^{n_i} F(y_{it}, x_{it}\beta + \sqrt{2\hat{\sigma}_i} a_m^* + \hat{\mu}_i) \right]
\]

where \( w_i \) is the user-specified weight for panel \( i \); if no weights are specified, \( w_i = 1 \).

The default method of adaptive Gauss–Hermite quadrature is to calculate the posterior mean and variance and use those parameters for \( \hat{\mu}_i \) and \( \hat{\sigma}_i \) by following the method of Naylor and Smith (1982), further discussed in Skrondal and Rabe-Hesketh (2004). We start with \( \hat{\sigma}_{i,0} = 1 \) and \( \hat{\mu}_{i,0} = 0 \), and the posterior means and variances are updated in the \( k \)th iteration. That is, at the \( k \)th iteration of the optimization for \( l_i \), we use

\[
\hat{\mu}_{i,k} = \frac{1}{\sqrt{2\hat{\sigma}_{i,k-1}} w_m^* \exp \left\{ (a_m^*)^2 \right\} g(y_{it}, x_{it}, \sqrt{2\hat{\sigma}_{i,k-1}} a_m^* + \hat{\mu}_{i,k-1})}
\]

Letting

\[
\tau_{i,m,k-1} = \sqrt{2\hat{\sigma}_{i,k-1}} a_m^* + \hat{\mu}_{i,k-1}
\]

\[
\hat{\mu}_{i,k} = \sqrt{\frac{\tau_{i,m,k-1}}{\sum_{m=1}^{M} \tau_{i,m,k-1}}} \prod_{m=1}^{M} \frac{\sqrt{2\hat{\sigma}_{i,k-1}} w_m^* \exp \left\{ (a_m^*)^2 \right\} g(y_{it}, x_{it}, \tau_{i,m,k-1})}{l_{i,k}}
\]

and

\[
\hat{\sigma}_{i,k} = \frac{1}{\sum_{m=1}^{M} \tau_{i,m,k-1}} \left( \frac{\tau_{i,m,k-1}}{w_m^* \exp \left\{ (a_m^*)^2 \right\} g(y_{it}, x_{it}, \tau_{i,m,k-1})} \right)^{1/2} - (\hat{\mu}_{i,k})^2
\]

and this is repeated until \( \hat{\mu}_{i,k} \) and \( \hat{\sigma}_{i,k} \) have converged for this iteration of the maximization algorithm. This adaptation is applied on every iteration until the log-likelihood change from the preceding iteration is less than a relative difference of \( 1e^{-6} \); after this, the quadrature parameters are fixed.

The log likelihood can also be calculated by nonadaptive Gauss–Hermite quadrature, the int-
metho(dghermite) option, where \( \rho = \sigma_\nu^2 / (\sigma_\nu^2 + 1) \):

\[
L = \sum_{i=1}^{n} w_i \log \left\{ \Pr(y_{i1}, \ldots, y_{in_i} | x_{i1}, \ldots, x_{in_i}) \right\}
\]

\[
\approx \sum_{i=1}^{n} w_i \log \left[ \frac{1}{\sqrt{\pi}} \sum_{m=1}^{M} w_m^* \prod_{t=1}^{n_i} F(y_{it}, x_{it}\beta + a_m^* \left( \frac{2\rho}{1 - \rho} \right)^{1/2}) \right]
\]

Both quadrature formulas require that the integrated function be well approximated by a polynomial of degree equal to the number of quadrature points. The number of periods (panel size) can affect whether

\[
\prod_{t=1}^{n_i} F(y_{it}, x_{it}\beta + \nu_i)
\]
is well approximated by a polynomial. As panel size and $\rho$ increase, the quadrature approximation can become less accurate. For large $\rho$, the random-effects model can also become unidentified. Adaptive quadrature gives better results for correlated data and large panels than nonadaptive quadrature; however, we recommend that you use the quadchk command (see [XT] quadchk) to verify the quadrature approximation used in this command, whichever approximation you choose.

**xtprobit, re and the robust VCE estimator**

Specifying `vce(robust)` or `vce(cluster clustvar)` causes the Huber/White/sandwich VCE estimator to be calculated for the coefficients estimated in this regression. See [P] _robust, particularly Introduction and Methods and formulas. Wooldridge (2013) and Arellano (2003) discuss this application of the Huber/White/sandwich VCE estimator. As discussed by Wooldridge (2013), Stock and Watson (2008), and Arellano (2003), specifying `vce(robust)` is equivalent to specifying `vce(cluster panelvar)`, where `panelvar` is the variable that identifies the panels.

Clustering on the panel variable produces a consistent VCE estimator when the disturbances are not identically distributed over the panels or there is serial correlation in $\epsilon_{it}$.

The cluster–robust VCE estimator requires that there are many clusters and the disturbances are uncorrelated across the clusters. The panel variable must be nested within the cluster variable because of the within-panel correlation that is generally induced by the random-effects transform when there is heteroskedasticity or within-panel serial correlation in the idiosyncratic errors.

**References**


Also see

[XT] *xtprobit* postestimation — Postestimation tools for *xtprobit*

[XT] *quadchk* — Check sensitivity of quadrature approximation

[XT] *xtcloglog* — Random-effects and population-averaged cloglog models

[XT] *xtgee* — Fit population-averaged panel-data models by using GEE

[XT] *xtlogit* — Fixed-effects, random-effects, and population-averaged logit models

[XT] *xtset* — Declare data to be panel data

[ME] *meprobit* — Multilevel mixed-effects probit regression

[MI] *estimation* — Estimation commands for use with *mi estimate*

[R] *probit* — Probit regression

[U] 20 Estimation and postestimation commands
Description

The following postestimation commands are available after `xtprobit`:

<table>
<thead>
<tr>
<th>Command</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>contrast</td>
<td>contrasts and ANOVA-style joint tests of estimates</td>
</tr>
<tr>
<td>ic¹</td>
<td>Akaike’s and Schwarz’s Bayesian information criteria (AIC and BIC)</td>
</tr>
<tr>
<td>estat ic</td>
<td>Akaike’s and Schwarz’s Bayesian information criteria (AIC and BIC)</td>
</tr>
<tr>
<td>estat summarize</td>
<td>summary statistics for the estimation sample</td>
</tr>
<tr>
<td>estat vce</td>
<td>variance–covariance matrix of the estimators (VCE)</td>
</tr>
<tr>
<td>estimates</td>
<td>cataloging estimation results</td>
</tr>
<tr>
<td>forecast²</td>
<td>dynamic forecasts and simulations</td>
</tr>
<tr>
<td>lincom</td>
<td>point estimates, standard errors, testing, and inference for linear combinations of coefficients</td>
</tr>
<tr>
<td>lrtest</td>
<td>likelihood-ratio test</td>
</tr>
<tr>
<td>margins</td>
<td>marginal means, predictive margins, marginal effects, and average marginal effects</td>
</tr>
<tr>
<td>marginsplot</td>
<td>graph the results from margins (profile plots, interaction plots, etc.)</td>
</tr>
<tr>
<td>nlcom</td>
<td>point estimates, standard errors, testing, and inference for nonlinear combinations of coefficients</td>
</tr>
<tr>
<td>predict</td>
<td>predictions, residuals, influence statistics, and other diagnostic measures</td>
</tr>
<tr>
<td>predictnl</td>
<td>point estimates, standard errors, testing, and inference for generalized predictions</td>
</tr>
<tr>
<td>pwcompare</td>
<td>pairwise comparisons of estimates</td>
</tr>
<tr>
<td>test</td>
<td>Wald tests of simple and composite linear hypotheses</td>
</tr>
<tr>
<td>testnl</td>
<td>Wald tests of nonlinear hypotheses</td>
</tr>
</tbody>
</table>

¹ `estat ic` is not appropriate after `xtprobit, pa`.
² `forecast` is not appropriate with mi estimation results.

Syntax for predict

**Random-effects model**

`predict [type] newvar [if] [in] [, RE_statistic nooffset]`

**Population-averaged model**

`predict [type] newvar [if] [in] [, PA_statistic nooffset]`
**RE_statistic**

<table>
<thead>
<tr>
<th>Main</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>xb</td>
<td>linear prediction; the default</td>
</tr>
<tr>
<td>pu0</td>
<td>probability of a positive outcome</td>
</tr>
<tr>
<td>stdp</td>
<td>standard error of the linear prediction</td>
</tr>
</tbody>
</table>

**PA_statistic**

<table>
<thead>
<tr>
<th>Main</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>mu</td>
<td>probability of <em>depvar</em>; considers the <code>offset()</code>; the default</td>
</tr>
<tr>
<td>rate</td>
<td>probability of <em>depvar</em></td>
</tr>
<tr>
<td>xb</td>
<td>linear prediction</td>
</tr>
<tr>
<td>stdp</td>
<td>standard error of the linear prediction</td>
</tr>
<tr>
<td>score</td>
<td>first derivative of the log likelihood with respect to $x_j\beta$</td>
</tr>
</tbody>
</table>

These statistics are available both in and out of sample; type `predict ... if e(sample)` ... if wanted only for the estimation sample.

**Menu for predict**

Statistics > Postestimation > Predictions, residuals, etc.

**Options for predict**

* **xb** calculates the linear prediction. This is the default for the random-effects model.*

* **pu0** calculates the probability of a positive outcome, assuming that the random effect for that observation’s panel is zero ($\nu = 0$). This probability may not be similar to the proportion of observed outcomes in the group.*

* **stdp** calculates the standard error of the linear prediction.*

* **mu and rate** both calculate the predicted probability of *depvar*. *mu* takes into account the `offset()`, and **rate** ignores those adjustments. **mu** and **rate** are equivalent if you did not specify `offset()`.*

* **mu** is the default for the population-averaged model.*

* **score** calculates the equation-level score, $u_j = \frac{\partial \ln L_j(x_j\beta)}{\partial (x_j\beta)}$.*

* **nooffset** is relevant only if you specified `offset(varname)` for `xtprobit`. It modifies the calculations made by `predict` so that they ignore the offset variable; the linear prediction is treated as $x_{it}\beta$ rather than $x_{it}\beta + \text{offset}_{it}$.
Remarks and examples

Example 1

In example 2 of [XT] xtprobit, we fit a population-averaged model of union status on the woman’s age and level of schooling, whether she lived in an urban area, whether she lived in the south, and the year observed. Here we compute the average marginal effects from that fitted model on the probability of being in a union.

```
. use http://www.stata-press.com/data/r13/union
(NLS Women 14–24 in 1968)
. xtprobit union age grade i.not_smsa south##c.year, pa
(output omitted)
. margins, dydx(*)
Average marginal effects Number of obs = 26200
Model VCE : Conventional
Expression : Pr(union != 0), predict()
dy/dx w.r.t. : age grade 1.not_smsa 1.south

|            | dy/dx | Std. Err. | z    | P>|z|    | [95% Conf. Interval] |
|------------|-------|-----------|------|--------|---------------------|
| age        | .0025337 | .0015035  | 1.69 | 0.092  | -.0004132 .0054805  |
| grade      | .0094109 | .0017566  | 5.36 | 0.000  | .005968 .0128537    |
| 1.not_smsa | -.0199744 | .0075879  | -2.63| 0.008  | -.0348464 -.0051023 |
| 1.south    | -.0910805 | .0073315  | -12.42| 0.000  | -.10545 -.076711    |
| year       | -.000938  | .0015413  | -0.61| 0.543  | -.0039589 .0020828  |

Note: dy/dx for factor levels is the discrete change from the base level.
```

On average, not living in a metropolitan area (not_smsa = 0) lowers the probability of being in a union by about two percentage points.

Also see

[XT] xtprobit — Random-effects and population-averaged probit models

[U] 20 Estimation and postestimation commands
Syntax

\texttt{xtrc depvar indepvars [if] [in] [ , options]}

<table>
<thead>
<tr>
<th>options</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>\texttt{noconstant}</td>
<td>suppress constant term</td>
</tr>
<tr>
<td>\texttt{offset(varname)}</td>
<td>include \textit{varname} in model with coefficient constrained to 1</td>
</tr>
<tr>
<td>\texttt{vce(vcetype)}</td>
<td>\textit{vcetype} may be \textit{conventional}, \textit{bootstrap}, or \textit{jackknife}</td>
</tr>
<tr>
<td>\texttt{level(#)}</td>
<td>set confidence level; default is \texttt{level(95)}</td>
</tr>
<tr>
<td>\texttt{betas}</td>
<td>display \textit{group-specific best linear predictors}</td>
</tr>
<tr>
<td>\texttt{display_options}</td>
<td>control column formats, row spacing, line width, display of omitted variables and base and empty cells, and factor-variable labeling</td>
</tr>
<tr>
<td>\texttt{coeflegend}</td>
<td>display legend instead of statistics</td>
</tr>
</tbody>
</table>

A panel variable must be specified; use \texttt{xtset}; see \texttt{[XT] xtset}.
\texttt{indepvars} may contain factor variables; see \texttt{[U 11.4.3 Factor variables]}.
by, \texttt{mi estimate}, and \texttt{statsby} are allowed; see \texttt{[U 11.1.10 Prefix commands]}.
\texttt{vce(bootstrap)} and \texttt{vce(jackknife)} are not allowed with the \texttt{mi estimate} prefix; see \texttt{[MI] mi estimate}.
\texttt{coeflegend} does not appear in the dialog box.
See \texttt{[U 20 Estimation and postestimation commands]} for more capabilities of estimation commands.

Menu

Statistics \textgreater{} Longitudinal/panel data \textgreater{} Random-coefficients regression by GLS

Description

\texttt{xtrc} fits the Swamy (1970) random-coefficients linear regression model.

Options

\texttt{noconstant, offset(varname)}; see \texttt{[R] estimation options}
vce(vcetype) specifies the type of standard error reported, which includes types that are derived from asymptotic theory (conventional) and that use bootstrap or jackknife methods (bootstrap, jackknife); see [XT] vce_options.

vce(conventional), the default, uses the conventionally derived variance estimator for generalized least-squares regression.

Reporting level(#); see [R] estimation options.

betas requests that the group-specific best linear predictors also be displayed.

display_options: noomitted, vsquish, noemptycells, baselevels, allbaselevels, nofvlabel, fvwrap(#), fvwrapon(style), cformat(%fmt), pformat(%fmt), sformat(%fmt), and nolstretch; see [R] estimation options.

The following option is available with xtrc but is not shown in the dialog box:
coeflegend; see [R] estimation options.

Remarks and examples

In random-coefficients models, we wish to treat the parameter vector as a realization (in each panel) of a stochastic process. xtrc fits the Swamy (1970) random-coefficients model, which is suitable for linear regression of panel data. See Greene (2012, chap. 11) and Poi (2003) for more information about this and other panel-data models.

Example 1

Greene (2012, 1112) reprints data from a classic study of investment demand by Grunfeld and Griliches (1960). In [XT] xtgls, we use this dataset to illustrate many of the possible models that may be fit with the xtgls command. Although the models included in the xtgls command offer considerable flexibility, they all assume that there is no parameter variation across firms (the cross-sectional units).

To take a first look at the assumption of parameter constancy, we should reshape our data so that we may fit a simultaneous-equation model with sureg; see [R] sureg. Because there are only five panels here, this is not too difficult.

}. use http://www.stata-press.com/data/r13/invest2
}. reshape wide invest market stock, i(time) j(company)
(note: j = 1 2 3 4 5)

Data long -> wide

<table>
<thead>
<tr>
<th></th>
<th>long</th>
<th>→</th>
<th>wide</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of obs.</td>
<td>100</td>
<td>→</td>
<td>20</td>
</tr>
<tr>
<td>Number of variables</td>
<td>5</td>
<td>→</td>
<td>16</td>
</tr>
<tr>
<td>j variable (5 values)</td>
<td>company → (dropped)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>xij variables:</td>
<td>invest → invest1 invest2 ... invest5</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>market → market1 market2 ... market5</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>stock → stock1 stock2 ... stock5</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Here we instead fit a random-coefficients model:

```
. use http://www.stata-press.com/data/r13/invest2
. xtrc invest market stock
```

Random-coefficients regression

|             | Coef. | Std. Err. | z     | P>|z|  | [95% Conf. Interval] |
|--------------|-------|-----------|-------|------|----------------------|
| invest       |       |           |       |      |                      |
| market       | .0807646 | .0250829 | 3.22  | 0.001| .0347935             |
| stock        | .2839914 | .0412016 | 6.86  | 0.000| .1968516             |
| _cons        | 85.42324 | 111.8774 | 0.76  | 0.445| -133.8525            |

Test of parameter constancy: chi2(12) = 603.99 Prob > chi2 = 0.0000
Just as the results of our simultaneous-equation model do not support the assumption of parameter constancy, the test included with the random-coefficients model also indicates that the assumption is not valid for these data. With large panel datasets, we would not want to take the time to look at a simultaneous-equations model (aside from the fact that our doing so was subjective).

### Stored results

*xtrc* stores the following in **e()**:  

**Scalars**  
- `e(N)` number of observations  
- `e(N_g)` number of groups  
- `e(df_m)` model degrees of freedom  
- `e(chi2)` $\chi^2$  
- `e(chi2_c)` $\chi^2$ for comparison test  
- `e(df_chi2c)` degrees of freedom for comparison $\chi^2$ test  
- `e(g_min)` smallest group size  
- `e(g_avg)` average group size  
- `e(g_max)` largest group size  
- `e(rank)` rank of $e(V)$  

**Macros**  
- `e(cmd)` `xtrc`  
- `e(cmdline)` command as typed  
- `e(depvar)` name of dependent variable  
- `e(ivar)` variable denoting groups  
- `e(tvar)` variable denoting time within groups  
- `e(title)` title in estimation output  
- `e(offset)` linear offset variable  
- `e(chi2type)` Wald; type of model $\chi^2$ test  
- `e(vce)` `vcetype` specified in `vce()`  
- `e(vcetype)` title used to label Std. Err.  
- `e(properties)` $b$ $V$  
- `e(predict)` program used to implement `predict`  
- `e(marginsnotok)` predictions disallowed by `margins`  
- `e(asbalanced)` factor variables `fvset` as `asbalanced`  
- `e(asobserved)` factor variables `fvset` as `asobserved`  

**Matrices**  
- `e(b)` coefficient vector  
- `e(Sigma)` $\Sigma$ matrix  
- `e(beta_ps)` matrix of best linear predictors  
- `e(V)` variance–covariance matrix of the estimators  
- `e(V_ps)` matrix of variances for the best linear predictors; row $i$ contains vec of variance matrix for group $i$ predictor  

**Functions**  
- `e(sample)` marks estimation sample

### Methods and formulas

In a random-coefficients model, the parameter heterogeneity is treated as stochastic variation. Assume that we write  

$$ y_i = X_i \beta_i + \epsilon_i $$

where $i = 1, \ldots, m$, and $\beta_i$ is the coefficient vector ($k \times 1$) for the $i$th cross-sectional unit, such that  

$$ \beta_i = \beta + \nu_i \quad E(\nu_i) = 0 \quad E(\nu_i \nu_i') = \Sigma $$

Our goal is to find $\widehat{\beta}$ and $\widehat{\Sigma}$. 
The derivation of the estimator assumes that the cross-sectional specific coefficient vector $\beta_i$ is the outcome of a random process with mean vector $\beta$ and covariance matrix $\Sigma$,

$$y_i = X_i\beta_i + \epsilon_i = X_i(\beta + \nu_i) + \epsilon_i = X_i\beta + (X_i\nu_i + \epsilon_i) = X_i\beta + \omega_i$$

where $E(\omega_i) = 0$ and

$$E(\omega_i \omega_i') = E\{(X_i\nu_i + \epsilon_i)(X_i\nu_i + \epsilon_i)\}' = E(\epsilon_i\epsilon_i') + X_iE(\nu_i\nu_i')X_i' = \sigma_i^2 I + X_i\Sigma X_i' = \Pi_i$$

Stacking the $m$ equations, we have

$$y = X\beta + \omega$$

where $\Pi \equiv E(\omega\omega')$ is a block diagonal matrix with $\Pi_i, i = 1...m$, along the main diagonal and zeros elsewhere. The GLS estimator of $\hat{\beta}$ is then

$$\hat{\beta} = \left(\sum_i X_i'\Pi_i^{-1}X_i\right)^{-1}\sum_i X_i'\Pi_i^{-1}y_i = \sum_{i=1}^m W_i b_i$$

where

$$W_i = \left\{\sum_{i=1}^m (\Sigma + V_i)^{-1}\right\}^{-1} \left(\Sigma + V_i\right)^{-1}$$

$$b_i = (X_i'X_i)^{-1}X_i'y_i$$ and $$V_i = \sigma_i^2 (X_i'X_i)^{-1},$$ showing that the resulting GLS estimator is a matrix-weighted average of the panel-specific OLS estimators. The variance of $\hat{\beta}$ is

$$\text{Var}(\hat{\beta}) = \sum_{i=1}^m (\Sigma + V_i)^{-1}$$

To calculate the above estimator $\hat{\beta}$ for the unknown $\Sigma$ and $V_i$ parameters, we use the two-step approach suggested by Swamy (1970):

$$b_i = \text{OLS panel-specific estimator}$$

$$\hat{\sigma}_i^2 = \frac{\epsilon_i'\epsilon_i}{n_i - k}$$

$$\hat{V}_i = \hat{\sigma}_i^2 (X_i'X_i)^{-1}$$

$$\bar{b} = \frac{1}{m} \sum_{i=1}^m b_i$$

$$\hat{\Sigma} = \frac{1}{m - 1} \left(\sum_{i=1}^m b_ib_i' - m\bar{b}\bar{b}'\right) - \frac{1}{m} \sum_{i=1}^m \hat{V}_i$$

The two-step procedure begins with the usual OLS estimates of $\beta_i$. With those estimates, we may proceed by obtaining estimates of $\hat{V}_i$ and $\hat{\Sigma}$ (and thus $\hat{W}_i$) and then obtain an estimate of $\beta$. 
Swamy (1970) further points out that the matrix $\hat{\Sigma}$ may not be positive definite and that because the second term is of order $1/(mT)$, it is negligible in large samples. A simple and asymptotically expedient solution is simply to drop this second term and instead use

$$\hat{\Sigma} = \frac{1}{m-1} \left( \sum_{i=1}^{m} b_i b_i' - m \bar{b} \bar{b}' \right)$$

As discussed by Judge et al. (1985, 541), the feasible best linear predictor of $\beta_i$ is given by

$$\hat{\beta}_i = \hat{\beta} + \hat{\Sigma} X_i' \left( X_i \hat{\Sigma} X_i' + \hat{\sigma}_i^2 I \right)^{-1} \left( y_i - X_i \hat{\beta} \right)$$

$$= \left( \hat{\Sigma}^{-1} + \hat{V}_i^{-1} \right)^{-1} \left( \hat{\Sigma}^{-1} \hat{\beta} + \hat{V}_i^{-1} b_i \right)$$

The conventional variance of $\hat{\beta}_i$ is given by

$$\text{Var}(\hat{\beta}_i) = \text{Var}(\hat{\beta}) + (I - A_i) \left\{ \hat{V}_i - \text{Var}(\hat{\beta}) \right\} (I - A_i)'$$

where

$$A_i = \left( \hat{\Sigma}^{-1} + \hat{V}_i^{-1} \right)^{-1} \hat{\Sigma}^{-1}$$

To test the model, we may look at the difference between the OLS estimate of $\beta$, ignoring the panel structure of the data and the matrix-weighted average of the panel-specific OLS estimators. The test statistic suggested by Swamy (1970) is given by

$$\chi^2_{k(m-1)} = \sum_{i=1}^{m} (b_i - \bar{\beta}^*)' \hat{V}_i^{-1} (b_i - \bar{\beta}^*) \quad \text{where} \quad \bar{\beta}^* = \left( \sum_{i=1}^{m} \hat{V}_i^{-1} \right)^{-1} \sum_{i=1}^{m} \hat{V}_i^{-1} b_i$$

Johnston and DiNardo (1997) have shown that the test is algebraically equivalent to testing

$$H_0 : \beta_1 = \beta_2 = \cdots = \beta_m$$

in the generalized (groupwise heteroskedastic) xtreg model, where $V$ is block diagonal with $i$th diagonal element $\Pi_i$. 
References


Also see

[XT] **xtrc postestimation** — Postestimation tools for xtrc

[XT] **xtreg** — Fixed-, between-, and random-effects and population-averaged linear models

[XT] **xtset** — Declare data to be panel data

[ME] **mixed** — Multilevel mixed-effects linear regression

[Mi] **estimation** — Estimation commands for use with mi estimate

[U] **20 Estimation and postestimation commands**
### Description

The following postestimation commands are available after `xtrc`:

<table>
<thead>
<tr>
<th>Command</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>contrast</td>
<td>contrasts and ANOVA-style joint tests of estimates</td>
</tr>
<tr>
<td>estat summarize</td>
<td>summary statistics for the estimation sample</td>
</tr>
<tr>
<td>estat vce</td>
<td>variance–covariance matrix of the estimators (VCE)</td>
</tr>
<tr>
<td>estimates</td>
<td>cataloging estimation results</td>
</tr>
<tr>
<td>forecast¹</td>
<td>dynamic forecasts and simulations</td>
</tr>
<tr>
<td>lincom</td>
<td>point estimates, standard errors, testing, and inference for linear combination of coefficients</td>
</tr>
<tr>
<td>margins</td>
<td>marginal means, predictive margins, marginal effects, and average marginal effects</td>
</tr>
<tr>
<td>marginsplot</td>
<td>graph the results from margins (profile plots, interaction plots, etc.)</td>
</tr>
<tr>
<td>nlcom</td>
<td>point estimates, standard errors, testing, and inference for nonlinear combinations of coefficients</td>
</tr>
<tr>
<td>predict</td>
<td>predictions, residuals, influence statistics, and other diagnostic measures</td>
</tr>
<tr>
<td>predictnl</td>
<td>point estimates, standard errors, testing, and inference for generalized predictions</td>
</tr>
<tr>
<td>pwcompare</td>
<td>pairwise comparisons of estimates</td>
</tr>
<tr>
<td>test</td>
<td>Wald tests of simple and composite linear hypotheses</td>
</tr>
<tr>
<td>testnl</td>
<td>Wald tests of nonlinear hypotheses</td>
</tr>
</tbody>
</table>

¹ `forecast` is not appropriate with `mi` estimation results.

### Syntax for predict

```
predict [type] newvar [if] [in] [ , statistic nooffset ]
```

<table>
<thead>
<tr>
<th>statistic</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>xb</td>
<td>linear prediction; the default</td>
</tr>
<tr>
<td>stdp</td>
<td>standard error of the linear prediction</td>
</tr>
<tr>
<td>group(group)</td>
<td>linear prediction based on group <code>group</code></td>
</tr>
</tbody>
</table>

These statistics are available both in and out of sample; type `predict ... if e(sample) ...` if wanted only for the estimation sample.
Menu for predict

Statistics  >  Postestimation  >  Predictions, residuals, etc.

Options for predict

xb, the default, calculates the linear prediction using the mean parameter vector.
stdp calculates the standard error of the linear prediction.

group(group) calculates the linear prediction using the best linear predictors for group group.

nooffset is relevant only if you specified offset(varname) for xtrc. It modifies the calculations made by predict so that they ignore the offset variable; the linear prediction is treated as x_{it}b rather than x_{it}b + offset_{it}.

Also see

[XT] xtrc — Random-coefficients model

[U] 20 Estimation and postestimation commands
xtreg — Fixed-, between-, and random-effects and population-averaged linear models

Syntax

GLS random-effects (RE) model

```
xtestreg depvar [indepvars] [if] [in] [, re RE_options]
```

Between-effects (BE) model

```
xtestreg depvar [indepvars] [if] [in] , be [BE_options]
```

Fixed-effects (FE) model

```
xtestreg depvar [indepvars] [if] [in] [weight] , fe [FE_options]
```

ML random-effects (MLE) model

```
xtestreg depvar [indepvars] [if] [in] [weight] , mle [MLE_options]
```

Population-averaged (PA) model

```
xtestreg depvar [indepvars] [if] [in] [weight] , pa [PA_options]
```

RE_options

<table>
<thead>
<tr>
<th>Model</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>re</td>
<td>use random-effects estimator; the default</td>
</tr>
<tr>
<td>sa</td>
<td>use Swamy–Arora estimator of the variance components</td>
</tr>
</tbody>
</table>

SE/Robust

```
vce(vcetype) vcetype may be conventional, robust, cluster clustvar, bootstrap, or jackknife
```

Reporting

```
level(#) set confidence level; default is level(95)
theta report \( \theta \)
display_options control column formats, row spacing, line width, display of omitted variables and base and empty cells, and factor-variable labeling
```

```
coeflegend display legend instead of statistics
```

359
### BE_options

<table>
<thead>
<tr>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Model</strong></td>
</tr>
<tr>
<td>be</td>
</tr>
<tr>
<td>wls</td>
</tr>
<tr>
<td><strong>SE</strong></td>
</tr>
<tr>
<td><code>vce(vcetype)</code></td>
</tr>
<tr>
<td><strong>Reporting</strong></td>
</tr>
<tr>
<td><code>level(#)</code></td>
</tr>
<tr>
<td><code>display_options</code></td>
</tr>
<tr>
<td><code>coeflegend</code></td>
</tr>
</tbody>
</table>

### FE_options

<table>
<thead>
<tr>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Model</strong></td>
</tr>
<tr>
<td>fe</td>
</tr>
<tr>
<td><strong>SE/Robust</strong></td>
</tr>
<tr>
<td><code>vce(vcetype)</code></td>
</tr>
<tr>
<td><strong>Reporting</strong></td>
</tr>
<tr>
<td><code>level(#)</code></td>
</tr>
<tr>
<td><code>display_options</code></td>
</tr>
<tr>
<td><code>coeflegend</code></td>
</tr>
</tbody>
</table>

### MLE_options

<table>
<thead>
<tr>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Model</strong></td>
</tr>
<tr>
<td><code>noconstant</code></td>
</tr>
<tr>
<td>mle</td>
</tr>
<tr>
<td><strong>SE</strong></td>
</tr>
<tr>
<td><code>vce(vcetype)</code></td>
</tr>
<tr>
<td><strong>Reporting</strong></td>
</tr>
<tr>
<td><code>level(#)</code></td>
</tr>
<tr>
<td><code>display_options</code></td>
</tr>
<tr>
<td><strong>Maximization</strong></td>
</tr>
<tr>
<td><code>maximize_options</code></td>
</tr>
<tr>
<td><code>coeflegend</code></td>
</tr>
</tbody>
</table>
### xtreg — Fixed-, between-, and random-effects and population-averaged linear models

<table>
<thead>
<tr>
<th><strong>Model</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><code>noconstant</code></td>
<td>suppress constant term</td>
</tr>
<tr>
<td><code>pa</code></td>
<td>use population-averaged estimator</td>
</tr>
<tr>
<td><code>offset(varname)</code></td>
<td>include <code>varname</code> in model with coefficient constrained to 1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Correlation</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><code>corr(corr)</code>, <code>force</code></td>
<td>within-panel correlation structure; estimate even if observations unequally spaced in time</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>SE/Robust</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><code>vce(vcetype)</code></td>
<td><code>vcetype</code> may be <code>conventional</code>, <code>robust</code>, <code>bootstrap</code>, or <code>jackknife</code></td>
</tr>
<tr>
<td><code>nmp</code></td>
<td>use divisor $N - P$ instead of the default $N$</td>
</tr>
<tr>
<td><code>rgf</code></td>
<td>multiply the robust variance estimate by $(N - 1)/(N - P)$</td>
</tr>
<tr>
<td><code>scale(parm)</code></td>
<td>overrides the default scale parameter; <code>parm</code> may be <code>x2</code>, <code>dev</code>, <code>phi</code>, or <code>#</code></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Reporting</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><code>level(#)</code>, <code>display_options</code></td>
<td>set confidence level; default is <code>level(95)</code>; control column formats, row spacing, line width, display of omitted variables and base and empty cells, and factor-variable labeling</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Optimization</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><code>optimize_options</code></td>
<td>control the optimization process; seldom used</td>
</tr>
<tr>
<td><code>coeflegend</code></td>
<td>display legend instead of statistics</td>
</tr>
</tbody>
</table>

---

### correlation

<table>
<thead>
<tr>
<th><strong>correlation</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><code>exchangeable</code>, <code>independent</code>, <code>unstructured</code>, <code>fixed matname</code>, <code>ar #</code>, <code>stationary #</code>, <code>nonstationary #</code></td>
<td>exchangeable, independent, unstructured, user-specified, autoregressive of order #, stationary of order #, nonstationary of order #</td>
</tr>
</tbody>
</table>

A panel variable must be specified. For `xtreg`, `pa`, correlation structures other than `exchangeable` and `independent` require that a time variable also be specified. Use `xtset`; see [XT] `xtset`.

`indepvars` may contain factor variables; see [U] 11.4.3 Factor variables.

`depvar` and `indepvars` may contain time-series operators; see [U] 11.4.4 Time-series varlists.

`by`, `mi estimate`, and `statsby` are allowed; see [U] 11.1.10 Prefix commands. `fp` is allowed for the between-effects, fixed-effects, and maximum-likelihood random-effects models.

`vce(bootstrap)` and `vce(jackknife)` are not allowed with the `mi estimate` prefix; see [MI] `mi estimate`.

`aweights`, `fweights`, and `pweights` are allowed for the fixed-effects model. `iweights`, `fweights`, and `pweights` are allowed for the population-averaged model. `iweights` are allowed for the maximum-likelihood random-effects (MLE) model. See [U] 11.1.6 weight. Weights must be constant within panel.

`coeflegend` does not appear in the dialog box.

See [U] 20 Estimation and postestimation commands for more capabilities of estimation commands.
xtreg fits regression models to panel data. In particular, xtreg with the be option fits random-effects models by using the between regression estimator; with the fe option, it fits fixed-effects models (by using the within regression estimator); and with the re option, it fits random-effects models by using the GLS estimator (producing a matrix-weighted average of the between and within results). See [XT] xtdta for a faster way to fit fixed- and random-effects models.

**Options for RE model**

- `re`, the default, requests the GLS random-effects estimator.
- `sa` specifies that the small-sample Swamy–Arora estimator individual-level variance component be used instead of the default consistent estimator. See xtreg, re in Methods and formulas for details.

- `vce(vcetype)` specifies the type of standard error reported, which includes types that are derived from asymptotic theory (conventional), that are robust to some kinds of misspecification (robust), that allow for intragroup correlation (cluster clustvar), and that use bootstrap or jackknife methods (bootstrap, jackknife); see [XT] vce_options.
  - `vce(robust)`, the default, uses the conventionally derived variance estimator for generalized least-squares regression.
  - Specifying `vce(robust)` is equivalent to specifying `vce(cluster panelvar)`; see xtreg, re in Methods and formulas.

- `level(#)`; see [R] estimation options.

- `theta` specifies that the output include the estimated value of θ used in combining the between and fixed estimators. For balanced data, this is a constant, and for unbalanced data, a summary of the values is presented in the header of the output.

  `display_options`: noomitted, vsquish, noemptycells, baselevels, allbaselevels, nolabel, fvwrap(#), fvwrapon(style), cformat(%fmt), pformat(%fmt), sformat(%fmt), and nolstretch; see [R] estimation options.

The following option is available with xtreg but is not shown in the dialog box:

- `coeflegend`; see [R] estimation options.
Options for BE model

`be` requests the between regression estimator.

`wls` specifies that, for unbalanced data, weighted least squares be used rather than the default OLS. Both methods produce consistent estimates. The true variance of the between-effects residual is \( \sigma^2_v + T_i \sigma^2_\epsilon \) (see `xtreg, be` in Methods and formulas below). WLS produces a “stabilized” variance of \( \sigma^2_v/T_i + \sigma^2_\epsilon \), which is also not constant. Thus the choice between OLS and WLS amounts to which is more stable.

Comment: `xtreg, be` is rarely used anyway, but between estimates are an ingredient in the random-effects estimate. Our implementation of `xtreg, re` uses the OLS estimates for this ingredient, based on our judgment that \( \sigma^2_v \) is large relative to \( \sigma^2_\epsilon \) in most models. Formally, only a consistent estimate of the between estimates is required.

`vce(vcetype)` specifies the type of standard error reported, which includes types that are derived from asymptotic theory (`conventional`) and that use bootstrap or jackknife methods (`bootstrap`, `jackknife`); see [XT] `vce_options`.

`vce(conventional)`, the default, uses the conventionally derived variance estimator for generalized least-squares regression.

Options for FE model

`fe` requests the fixed-effects (within) regression estimator.

`vce(vcetype)` specifies the type of standard error reported, which includes types that are derived from asymptotic theory (`conventional`), that are robust to some kinds of misspecification (`robust`), that allow for intragroup correlation (`cluster clustvar`), and that use bootstrap or jackknife methods (`bootstrap`, `jackknife`); see [XT] `vce_options`.

`vce(conventional)`, the default, uses the conventionally derived variance estimator for generalized least-squares regression.

Specifying `vce(robust)` is equivalent to specifying `vce(cluster panelvar)`; see `xtreg, fe` in Methods and formulas.
xtreg — Fixed-, between-, and random-effects and population-averaged linear models

Reporting

level(#) ; see [R] estimation options.

display_options: noomitted, vsquish, noemptycells, baselevels, allbaselevels, nofvlabel, fvwrap(#), fvwrappon(style), cformat(\%fmt), pformat(\%fmt), sformat(\%fmt), and nolstretch; see [R] estimation options.

The following option is available with xtreg but is not shown in the dialog box:
coeflegend; see [R] estimation options.

Options for MLE model

Model

noconstant; see [R] estimation options.
mle requests the maximum-likelihood random-effects estimator.

SE

vce(vcetype) specifies the type of standard error reported, which includes types that are derived from asymptotic theory (oim) and that use bootstrap or jackknife methods (bootstrap, jackknife); see [XT] vce_options.

Reporting

level(#) ; see [R] estimation options.

display_options: noomitted, vsquish, noemptycells, baselevels, allbaselevels, nofvlabel, fvwrap(#), fvwrappon(style), cformat(\%fmt), pformat(\%fmt), sformat(\%fmt), and nolstretch; see [R] estimation options.

Maximization

maximize_options: iterate(#), [no]log, trace, tolerance(#), ltolerance(#), and from(init_specs); see [R] maximize. These options are seldom used.

The following option is available with xtreg but is not shown in the dialog box:
coeflegend; see [R] estimation options.

Options for PA model

Model

noconstant; see [R] estimation options.
pa requests the population-averaged estimator. For linear regression, this is the same as a random-effects estimator (both interpretations hold).

xtreg, pa is equivalent to xtgee, family(gaussian) link(id) corr(exchangeable), which are the defaults for the xtgee command. xtreg, pa allows all the relevant xtgee options such as vce(robust). Whether you use xtreg, pa or xtgee makes no difference. See [XT] xtgee.
offset(varname); see [R] estimation options.
corr\( (\text{correlation}) \) specifies the within-panel correlation structure; the default corresponds to the equal-correlation model, corr\( (\text{exchangeable}) \).

When you specify a correlation structure that requires a lag, you indicate the lag after the structure’s name with or without a blank; for example, corr\( (\text{ar } 1) \) or corr\( (\text{ari}) \).

If you specify the fixed correlation structure, you specify the name of the matrix containing the assumed correlations following the word fixed, for example, corr\( (\text{fixed myr}) \).

force specifies that estimation be forced even though the time variable is not equally spaced. This is relevant only for correlation structures that require knowledge of the time variable. These correlation structures require that observations be equally spaced so that calculations based on lags correspond to a constant time change. If you specify a time variable indicating that observations are not equally spaced, the (time dependent) model will not be fit. If you also specify force, the model will be fit, and it will be assumed that the lags based on the data ordered by the time variable are appropriate.

vce\( (\text{vcetype}) \) specifies the type of standard error reported, which includes types that are derived from asymptotic theory (conventional), that are robust to some kinds of misspecification (robust), and that use bootstrap or jackknife methods (bootstrap, jackknife); see [XT] vce options.

vce\( (\text{conventional}) \), the default, uses the conventionally derived variance estimator for generalized least-squares regression.

rgf specifies that the robust variance estimate is multiplied by \( (N - 1)/(N - P) \), where \( N \) is the total number of observations and \( P \) is the number of coefficients estimated. This option can be used with family\( (\text{gaussian}) \) only when vce\( (\text{robust}) \) is either specified or implied by the use of pweights. Using this option implies that the robust variance estimate is not invariant to the scale of any weights used.

scale\( (x2 | \text{dev} | \phi | #) \); see [XT] vce options.

level\( (#) \); see [R] estimation options.

display options: noomitted, vsquish, noemptycells, baselevels, allbaselevels, nofvlabel, fvwrap(\#), fvwrapon(style), cformat\( (\%fmt) \), pformat\( (\%fmt) \), sformat\( (\%fmt) \), and nolstretch; see [R] estimation options.

optimize options control the iterative optimization process. These options are seldom used.

iterate\( (#) \) specifies the maximum number of iterations. When the number of iterations equals \#, the optimization stops and presents the current results, even if convergence has not been reached. The default is iterate\( (100) \).

tolerance\( (#) \) specifies the tolerance for the coefficient vector. When the relative change in the coefficient vector from one iteration to the next is less than or equal to \#, the optimization process is stopped. tolerance\( (1e-6) \) is the default.

nolog suppresses display of the iteration log.

trace specifies that the current estimates be printed at each iteration.
The following option is available with `xtreg` but is not shown in the dialog box: `coeflegend`; see [R] estimation options.

Remarks and examples

If you have not read [XT] **xt**, please do so.

See Baltagi (2013, chap. 2) and Wooldridge (2013, chap. 14) for good overviews of fixed-effects and random-effects models. Allison (2009) provides perspective on the use of fixed- versus random-effects estimators and provides many examples using Stata.

Consider fitting models of the form

\[ y_{it} = \alpha + x_{it}\beta + \nu_i + \epsilon_{it} \]  

(1)

In this model, \( \nu_i + \epsilon_{it} \) is the error term that we have little interest in; we want estimates of \( \beta \). \( \nu_i \) is the unit-specific error term; it differs between units, but for any particular unit, its value is constant.

In the pulmonary data of [XT] **xt**, a person who exercises less would presumably have a lower forced expiratory volume (FEV) year after year and so would have a negative \( \nu_i \).

\( \epsilon_{it} \) is the “usual” error term with the usual properties (mean 0, uncorrelated with itself, uncorrelated with \( x \), uncorrelated with \( \nu \), and homoskedastic), although in a more thorough development, we could decompose \( \epsilon_{it} = \upsilon_t + \omega_{it} \), assume that \( \omega_{it} \) is a conventional error term, and better describe \( \upsilon_t \).

Before making the assumptions necessary for estimation, let’s perform some useful algebra on (1). Whatever the properties of \( \nu_i \) and \( \epsilon_{it} \), if (1) is true, it must also be true that

\[ \bar{y}_i = \alpha + \bar{x}_i\beta + \nu_i + \bar{\epsilon}_i \]  

(2)

where \( \bar{y}_i = \sum_t y_{it}/T_i \), \( \bar{x}_i = \sum_t x_{it}/T_i \), and \( \bar{\epsilon}_i = \sum_t \epsilon_{it}/T_i \). Subtracting (2) from (1), it must be equally true that

\[ (y_{it} - \bar{y}_i) = (x_{it} - \bar{x}_i)\beta + (\epsilon_{it} - \bar{\epsilon}_i) \]  

(3)

These three equations provide the basis for estimating \( \beta \). In particular, `xtreg, fe` provides what is known as the fixed-effects estimator—also known as the within estimator—and amounts to using OLS to perform the estimation of (3). `xtreg, be` provides what is known as the between estimator and amounts to using OLS to perform the estimation of (2). `xtreg, re` provides the random-effects estimator and is a (matrix) weighted average of the estimates produced by the between and within estimators. In particular, the random-effects estimator turns out to be equivalent to estimation of

\[ (y_{it} - \theta \bar{y}_i) = (1 - \theta)\alpha + (x_{it} - \theta \bar{x}_i)\beta + \{(1 - \theta)\nu_i + (\epsilon_{it} - \theta \bar{\epsilon}_i)\} \]  

(4)

where \( \theta \) is a function of \( \sigma_\nu^2 \) and \( \sigma_\epsilon^2 \). If \( \sigma_\nu^2 = 0 \), meaning that \( \nu_i \) is always 0, \( \theta = 0 \) and (1) can be estimated by OLS directly. Alternatively, if \( \sigma_\epsilon^2 = 0 \), meaning that \( \epsilon_{it} \) is 0, \( \theta = 1 \) and the within estimator returns all the information available (which will, in fact, be a regression with an \( R^2 \) of 1).

For more reasonable cases, few assumptions are required to justify the fixed-effects estimator of (3). The estimates are, however, conditional on the sample in that the \( \nu_i \) are not assumed to have a distribution but are instead treated as fixed and estimable. This statistical fine point can lead to difficulty when making out-of-sample predictions, but that aside, the fixed-effects estimator has much to recommend it.
More is required to justify the between estimator of (2), but the conditioning on the sample is not assumed because $\nu_i + \bar{\epsilon}_i$ is treated as an error term. Newly required is that we assume that $\nu_i$ and $\bar{X}_i$ are uncorrelated. This follows from the assumptions of the OLS estimator but is also transparent: were $\nu_i$ and $\bar{X}_i$ correlated, the estimator could not determine how much of the change in $\bar{Y}_i$, associated with an increase in $X_i$, to assign to $\beta$ versus how much to attribute to the unknown correlation. (This, of course, suggests the use of an instrumental-variable estimator, $\bar{Z}_i$, which is correlated with $X_i$ but uncorrelated with $\nu_i$, though that approach is not implemented here.)

The random-effects estimator of (4) requires the same no-correlation assumption. In comparison with the between estimator, the random-effects estimator produces more efficient results, albeit ones with unknown small-sample properties. The between estimator is less efficient because it discards the over-time information in the data in favor of simple means; the random-effects estimator uses both the within and the between information.

All this would seem to leave the between estimator of (2) with no role (except for a minor, technical part it plays in helping to estimate $\sigma^2_{1v}$ and $\sigma^2_{\epsilon}$, which are used in the calculation of $\theta$, on which the random-effects estimates depend). Let’s, however, consider a variation on (1):

$$y_{it} = \alpha + \bar{X}_i \beta_1 + (x_{it} - \bar{X}_i) \beta_2 + \nu_i + \epsilon_{it} \tag{1'}$$

In this model, we postulate that changes in the average value of $x$ for an individual have a different effect from temporary departures from the average. In an economic situation, $y$ might be purchases of some item and $x$ income; a change in average income should have more effect than a transitory change. In a clinical situation, $y$ might be a physical response and $x$ the level of a chemical in the brain; the model allows a different response to permanent rather than transitory changes.

The variations of (2) and (3) corresponding to (1') are

$$\bar{y}_i = \alpha + \bar{X}_i \beta_1 + \nu_i + \bar{\epsilon}_i \tag{2'}$$

$$(y_{it} - \bar{y}_i) = (x_{it} - \bar{X}_i) \beta_2 + (\epsilon_{it} - \bar{\epsilon}_i) \tag{3'}$$

That is, the between estimator estimates $\beta_1$ and the within $\beta_2$, and neither estimates the other. Thus even when estimating equations like (1), it is worth comparing the within and between estimators. Differences in results can suggest models like (1'), or at least some other specification error.

Finally, it is worth understanding the role of the between and within estimators with regressors that are constant over time or constant over units. Consider the model

$$y_{it} = \alpha + x_{it} \beta_1 + s_i \beta_2 + z_t \beta_3 + \nu_i + \epsilon_{it} \tag{1''}$$

This model is the same as (1), except that we explicitly identify the variables that vary over both time and $i$ ($x_{it}$, such as output or FEV); variables that are constant over time ($s_i$, such as race or sex); and variables that vary solely over time ($z_t$, such as the consumer price index or age in a cohort study). The corresponding between and within equations are

$$\bar{y}_i = \alpha + \bar{X}_i \beta_1 + s_i \beta_2 + \bar{z}_t \beta_3 + \nu_i + \bar{\epsilon}_i \tag{2''}$$

$$(y_{it} - \bar{y}_i) = (x_{it} - \bar{X}_i) \beta_1 + (z_t - \bar{z}) \beta_3 + (\epsilon_{it} - \bar{\epsilon}_i) \tag{3''}$$

In the between estimator of (2''), no estimate of $\beta_3$ is possible because $\bar{z}$ is a constant across the $i$ observations; the regression-estimated intercept will be an estimate of $\alpha + \bar{z} \beta_3$. On the other hand, it can provide estimates of $\beta_1$ and $\beta_2$. It can estimate effects of factors that are constant over time, such as race and sex, but to do so it must assume that $\nu_i$ is uncorrelated with those factors.
Assessing goodness of fit

 Assessing goodness of fit

Assessing goodness of fit

$R^2$ is a popular measure of goodness of fit in ordinary regression. In our case, given $\hat{\alpha}$ and $\hat{\beta}$ estimates of $\alpha$ and $\beta$, we can assess the goodness of fit with respect to (1), (2), or (3). The prediction equations are, respectively,

$$\hat{y}_{it} = \hat{\alpha} + x_{it}\hat{\beta} \quad (1''')$$

$$\hat{\gamma}_i = \hat{\alpha} + \bar{x}_i\hat{\beta} \quad (2''')$$

$$\hat{\gamma}_i = (\hat{y}_{it} - \hat{\gamma}_i) = (x_{it} - \bar{x}_i)\hat{\beta} \quad (3''')$$

`xtreg` reports “$R$-squares” corresponding to these three equations. $R$-squares is in quotes because the $R$-squares reported do not have all the properties of the OLS $R^2$.

The ordinary properties of $R^2$ include being equal to the squared correlation between $\hat{y}$ and $y$ and being equal to the fraction of the variation in $y$ explained by $\hat{y}$—formally defined as $\text{Var}(\hat{y})/\text{Var}(y)$.

The identity of the definitions is from a special property of the OLS estimates; in general, given a prediction $\hat{y}$ for $y$, the squared correlation is not equal to the ratio of the variances, and the ratio of the variances is not required to be less than 1.

`xtreg` reports $R^2$ values calculated as correlations squared, calling them $R^2$ overall, corresponding to (1'''); $R^2$ between, corresponding to (2'''); and $R^2$ within, corresponding to (3'''). In fact, you can think of each of these three numbers as having all the properties of ordinary $R^2$s, if you bear in mind that the prediction being judged is not $\hat{y}_{it}$, $\hat{\gamma}_i$, and $\hat{\gamma}_i$, but $\gamma_1\hat{y}_{it}$ from the regression $y_{it} = \gamma_1\hat{y}_{it}$; $\gamma_2\hat{\gamma}_i$ from the regression $\bar{y}_i = \gamma_2\hat{\gamma}_i$; and $\gamma_3\hat{y}_{it}$ from $y_{it} = \gamma_3\hat{y}_{it}$.

In particular, `xtreg`, `fe` obtains its estimates by performing OLS on (2), and therefore its reported $R^2$ between is an ordinary $R^2$. The other two reported $R^2$s are merely correlations squared, or, if you prefer, $R^2$s from the second-round regressions $y_{it} = \gamma_{11}\hat{y}_{it}$ and $\hat{y}_{it} = \gamma_{13}\hat{y}_{it}$.

`xtreg`, `fe` obtains its estimates by performing OLS on (3), so its reported $R^2$ within is an ordinary $R^2$. As with `be`, the other $R^2$s are correlations squared, or, if you prefer, $R^2$s from the second-round regressions $y_{it} = \gamma_{22}\hat{y}_{it}$ and, as with `be`, $\hat{y}_{it} = \gamma_{23}\hat{y}_{it}$.

`xtreg`, `re` obtains its estimates by performing OLS on (4); none of the $R^2$s corresponding to (1'''), (2'''), or (3''') correspond directly to this estimator (the “relevant” $R^2$ is the one corresponding to (4)). All three reported $R^2$s are correlations squared, or, if you prefer, from second-round regressions.
Example 1: Between-effects model

Using nlswork.dta described in [XT] xt, we will model ln_wage in terms of completed years of schooling (grade), current age and age squared, current years worked (experience) and experience squared, current years of tenure on the current job and tenure squared, whether black (race = 2), whether residing in an area not designated a standard metropolitan statistical area (SMSA), and whether residing in the South.

To obtain the between-effects estimates, we use xtreg, be. nlswork.dta has previously been xtset idcode year because that is what is true of the data, but for running xtreg, it would have been sufficient to have xtset idcode by itself.

The between-effects regression is estimated on person-averages, so the “n = 4697” result is relevant. xtreg, be reports the “number of observations” and group-size information: describe in [XT] xt showed that we have 28,534 “observations”—person-years, really—of data. If we take the subsample that has no missing values in ln_wage, grade, ..., south leaves us with 28,091 observations on person-years, reflecting 4,697 persons, each observed for an average of 6.0 years.

For goodness of fit, the $R^2$ between is directly relevant; our $R^2$ is 0.4900. If, however, we use these estimates to predict the within model, we have an $R^2$ of 0.1591. If we use these estimates to fit the overall data, our $R^2$ is 0.3695.
The $F$ statistic tests that the coefficients on the regressors grade, age, ..., south are all jointly zero. Our model is significant.

The root mean squared error of the fitted regression, which is an estimate of the standard deviation of $\nu_i + \epsilon_i$, is 0.3036.

For our coefficients, each year of schooling increases hourly wages by 6.1%; age increases wages up to age 26.9 and thereafter decreases them (because the quadratic $ax^2 + bx + c$ turns over at $x = -b/2a$, which for our age and c.age#c.age coefficients is $0.0323158/(2 \times 0.0005997) \approx 26.9$); total experience increases wages at an increasing rate (which is surprising and bothersome); tenure on the current job increases wages up to a tenure of 12.1 years and thereafter decreases them; wages of blacks are, these things held constant, (approximately) 5.6% below that of nonblacks (approximately because 2.race is an indicator variable); residing in a non-SMSA (rural area) reduces wages by 18.6%; and residing in the South reduces wages by 9.9%.
Example 2: Fixed-effects model

To fit the same model with the fixed-effects estimator, we specify the `fe` option.

```
. xtreg ln_w grade age c.age#c.age ttl_exp c.ttl_exp#c.ttl_exp tenure
> c.tenure#c.tenure 2.race not_smsa south, fe
note: grade omitted because of collinearity
note: 2.race omitted because of collinearity
```

```
Fixed-effects (within) regression
Group variable: idcode

Number of obs = 28091
Number of groups = 4697

Obs per group:
min = 1
avg = 6.0
max = 15

R-sq: within = 0.1727  overall = 0.2625
        between = 0.3505  F(8,23386) = 610.12
        overall = 0.2625

corr(u_i, Xb) = 0.1936  Prob > F = 0.0000
```

```
              Coef.  Std. Err.     t     P>|t|    [95% Conf. Interval]
-------------  --------  --------  ------  -------  ----------------------
ln_wage      grade      0 (omitted)     -       .
              age       .0359987    .0033864  10.63  0.000  .0293611 -.0426362
              c.age       -.000723    .0000533 -13.58  0.000 -.0008274 -.0006186
              ttl_exp    .0334668    .0029653  11.29  0.000  .0276545  .039279
              c.ttl_exp    .0002163    .0001277  1.69  0.090  -.0000341  .0004666
              tenure     .0357539    .0018487  19.34  0.000  .0321303  .0393775
              c.tenure     -.0019701    .0001250 -15.76  0.000  -.0022151  -.0017251
              race  black      0 (omitted)     -       .
              not_smsa   -.0890108    .0095316  -9.34  0.000  -.1076933  -.0703282
              south     -.0606309    .0109319  -5.55  0.000  -.0820582  -.0392036
              _cons      1.03732     .0485546  21.36  0.000   .9421496  1.13249
sigma_u    .35562203
sigma_e    .29068923
rho      .59946283  (fraction of variance due to u_i)
```

The observation summary at the top is the same as for the between-effects model, although this time it is the “Number of obs” that is relevant.

Our three $R^2$s are not too different from those reported previously; the $R^2$ within is slightly higher (0.1727 versus 0.1591), and the $R^2$ between is a little lower (0.3505 versus 0.4900), as expected, because the between estimator maximizes $R^2$ between and the within estimator $R^2$ within. In terms of overall fit, these estimates are somewhat worse (0.2625 versus 0.3695).

`xtreg`, `fe` can estimate $\sigma_\nu$ and $\sigma_\epsilon$, although how you interpret these estimates depends on whether you are using `xtreg` to fit a fixed-effects model or random-effects model. To clarify this fine point, in the fixed-effects model, $\nu_i$ are formally fixed—they have no distribution. If you subscribe to this view, think of the reported $\hat{\sigma}_\nu$ as merely an arithmetic way to describe the range of the estimated but fixed $\nu_i$. If, however, you are using the fixed-effects estimator of the random-effects model, 0.355622 is an estimate of $\sigma_\nu$ or would be if there were no omitted variables.
Here both `grade` and `2.race` were omitted from the model because they do not vary over time. Because `grade` and `2.race` are time invariant, our estimate $u_i$ is an estimate of $\nu_i$ plus the effects of `grade` and `2.race`, so our estimate of the standard deviation is based on the variation in $\nu_i$, `grade`, and `2.race`. On the other hand, had `2.race` and `grade` been omitted merely because they were collinear with the other regressors in our model, $u_i$ would be an estimate of $\nu_i$, and $0.355622$ would be an estimate of $\sigma_{\nu}$. (`xtsum` and `xttab` allow you to determine whether a variable is time invariant; see [XT] `xtsum` and [XT] `xttab`.)

Regardless of the status of $u_i$, our estimate of the standard deviation of $\epsilon_{it}$ is valid (and, in fact, is the estimate that would be used by the random-effects estimator to produce its results).

Our estimate of the correlation of $u_i$ with $x_{it}$ suffers from the problem of what $u_i$ measures. We find correlation but cannot say whether this is correlation of $\nu_i$ with $x_{it}$ or merely correlation of `grade` and `2.race` with $x_{it}$. In any case, the fixed-effects estimator is robust to such a correlation, and the other estimates it produces are unbiased.

So, although this estimator produces no estimates of the effects of `grade` and `2.race`, it does predict that age has a positive effect on wages up to age 24.9 years (compared with 26.9 years estimated by the between estimator); that total experience still increases wages at an increasing rate (which is still bothersome); that tenure increases wages up to 9.1 years (compared with 12.1); that living in a non-SMSA reduces wages by 8.9% (compared with a more drastic 18.6%); and that living in the South reduces wages by 6.1% (as compared with 9.9%).

Example 3: Fixed-effects models with robust standard errors

If we suspect that there is heteroskedasticity or within-panel serial correlation in the idiosyncratic error term $\epsilon_{it}$, we could specify the `vce(robust)` option:
```stata
.xtreg ln_w grade c.age#c.age ttl_exp c.ttl_exp#c.ttl_exp tenure > c.tenure#c.tenure 2.race not_smsa south, fe vce(robust)
```

Note: grade omitted because of collinearity

Note: 2.race omitted because of collinearity

<table>
<thead>
<tr>
<th>Fixed-effects (within) regression</th>
<th>Number of obs = 28091</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group variable: idcode</td>
<td>Number of groups = 4697</td>
</tr>
<tr>
<td>R-sq: within = 0.1727</td>
<td>Obs per group: min = 1</td>
</tr>
<tr>
<td>between = 0.3505</td>
<td>avg = 6.0</td>
</tr>
<tr>
<td>overall = 0.2625</td>
<td>max = 15</td>
</tr>
<tr>
<td>F(8,4696) = 273.86</td>
<td>Prob &gt; F = 0.0000</td>
</tr>
<tr>
<td>corr(u_i, Xb) = 0.1936</td>
<td>(Std. Err. adjusted for 4697 clusters in idcode)</td>
</tr>
</tbody>
</table>

| ln_wage | Robust Coef. | Std. Err. | t | P>|t| | [95% Conf. Interval] |
|---------|--------------|-----------|---|------|-----------------------|
| grade   | 0 (omitted)  |           |   |      |                       |
| age     | .0359987     | .0052407  | 6.87 | 0.000 | .0257243 .046273    |
| c.age#c.age | -.000723 | .0000845 | -8.56 | 0.000 | -.0008887 -.0005573|
| ttl_exp | .0334668     | .004069   | 8.22 | 0.000 | .0254896 .0414439   |
| c.ttl_exp#c.ttl_exp | .0002163 | .0001763 | 1.23 | 0.220 | -.0001294 .0005619  |
| tenure  | .0357539     | .0024683  | 14.49 | 0.000 | .0309148 .040593    |
| c.tenure#c.tenure | -.0019701 | .0001696 | -11.62 | 0.000 | -.0023026 -.0016376 |
| race    | 0 (omitted)  |           |   |      |                       |
| black   | -.0890108    | .0137629  | -6.47 | 0.000 | -.1159926 -.062029  |
| not_smsa| -.0606309    | .0163366  | -3.71 | 0.000 | -.0926583 -.0286035 |
| south   | 1.03732      | .0739644  | 14.02 | 0.000 | .8923149 1.182325   |
| _cons   | 1.03732      | .0739644  | 14.02 | 0.000 | .8923149 1.182325   |

Although the estimated coefficients are the same with and without the vce(robust) option, the robust estimator produced larger standard errors and a p-value for c.ttl_exp#c.ttl_exp above the conventional 10%. The F test of $\nu_i = 0$ is suppressed because it is too difficult to compute the robust form of the statistic when there are more than a few panels.

Technical note

The robust standard errors reported above are identical to those obtained by clustering on the panel variable idcode. Clustering on the panel variable produces an estimator of the VCE that is robust to cross-sectional heteroskedasticity and within-panel (serial) correlation that is asymptotically equivalent to that proposed by Arellano (1987). Although the example above applies the fixed-effects estimator, the robust and cluster–robust VCE estimators are also available for the random-effects estimator. Wooldridge (2013) and Arellano (2003) discuss these robust and cluster–robust VCE estimators for the fixed-effects and random-effects estimators. More details are available in Methods and formulas.
Example 4: Random-effects model

Refitting our log-wage model with the random-effects estimator, we obtain

```
xtreg ln_w grade age c.age#c.age ttl_exp c.ttl_exp#c.ttl_exp tenure 
> c.tenure#c.tenure 2.race not_smsa south, re theta
```

Random-effects GLS regression

<table>
<thead>
<tr>
<th>Number of obs</th>
<th>=</th>
<th>28091</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group variable: idcode</td>
<td>Number of groups</td>
<td>=</td>
</tr>
<tr>
<td>R-sq: within</td>
<td>=</td>
<td>0.1715</td>
</tr>
<tr>
<td>Obs per group: min</td>
<td>=</td>
<td>1</td>
</tr>
<tr>
<td>between</td>
<td>=</td>
<td>0.4784</td>
</tr>
<tr>
<td>avg</td>
<td>=</td>
<td>6.0</td>
</tr>
<tr>
<td>overall</td>
<td>=</td>
<td>0.3708</td>
</tr>
<tr>
<td>max</td>
<td>=</td>
<td>15</td>
</tr>
<tr>
<td>Wald chi2(10)</td>
<td>=</td>
<td>9244.74</td>
</tr>
<tr>
<td>corr(u_i, X) = 0 (assumed)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prob &gt; chi2</td>
<td>=</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>theta</th>
</tr>
</thead>
<tbody>
<tr>
<td>min</td>
</tr>
<tr>
<td>5%</td>
</tr>
<tr>
<td>median</td>
</tr>
<tr>
<td>95%</td>
</tr>
<tr>
<td>max</td>
</tr>
</tbody>
</table>

| ln_wage | Coef. | Std. Err. | z | P>|z| | [95% Conf. Interval] |
|---------|-------|-----------|---|-------|---------------------|
| grade | .0646499 | .0017812 | 36.30 | 0.000 | .0611589 | .0681409 |
| age | .0368059 | .0031195 | 11.80 | 0.000 | .0306918 | .0429201 |
| c.age#c.age | -.0007133 | .00005 | -14.27 | 0.000 | -.0008113 | -.0006153 |
| ttl_exp | .0290208 | .002422 | 11.98 | 0.000 | .0242739 | .0337678 |
| c.ttl_exp#c.ttl_exp | .0003049 | .0001162 | 2.62 | 0.009 | .000077 | .0005327 |
| tenure | .0392519 | .0017554 | 22.36 | 0.000 | .0358113 | .0426925 |
| c.tenure#c.tenure | -.0020035 | .0001193 | -16.80 | 0.000 | -.0022373 | -.0017697 |
| race | black | -.053053 | .0099926 | -5.31 | 0.000 | -.0726381 | -.0334679 |
| not_smsa | -.1308252 | .0071751 | -18.23 | 0.000 | -.1448881 | -.1167622 |
| south | -.0868922 | .0073032 | -11.90 | 0.000 | -.1012062 | -.0725781 |
| _cons | .2387207 | .049469 | 4.83 | 0.000 | .1417633 | .3356781 |

Sigma_u | .25790526 |
Sigma_e | .29068923 |
Rho | .44045273 (fraction of variance due to u_i) |

According to the $R^2$s, this estimator performs worse within than the within fixed-effects estimator and worse between than the between estimator, as it must, and slightly better overall.

We estimate that $\sigma_\nu$ is 0.2579 and $\sigma_\epsilon$ is 0.2907 and, by assertion, assume that the correlation of $\nu$ and $x$ is zero.

All that is known about the random-effects estimator is its asymptotic properties, so rather than reporting an $F$ statistic for overall significance, `xtreg, re` reports a $\chi^2$. Taken jointly, our coefficients are significant.

`xtreg, re` also reports a summary of the distribution of $\theta_i$, an ingredient in the estimation of (4). $\theta$ is not a constant here because we observe women for unequal periods.

We estimate that schooling has a rate of return of 6.5% (compared with 6.1% between and no estimate within); that the increase of wages with age turns around at 25.8 years (compared with 26.9 between and 24.9 within); that total experience yet again increases wages increasingly; that the effect
of job tenure turns around at 9.8 years (compared with 12.1 between and 9.1 within); that being black reduces wages by 5.3% (compared with 5.6% between and no estimate within); that living in a non-SMSA reduces wages 13.1% (compared with 18.6% between and 8.9% within); and that living in the South reduces wages 8.7% (compared with 9.9% between and 6.1% within).

Example 5: Random-effects model fit using ML

We could also have fit this random-effects model with the maximum likelihood estimator:

```
. xtab ln_w grade age c.age#c.age ttl_exp c.ttl_exp#c.ttl_exp tenure > c.tenure#c.tenure 2.race not_smsa south, mle
Fitting constant-only model:
Iteration 0: log likelihood = -13690.161
Iteration 1: log likelihood = -12819.317
Iteration 2: log likelihood = -12662.039
Iteration 3: log likelihood = -12649.744
Iteration 4: log likelihood = -12649.614
Iteration 5: log likelihood = -12649.614
Fitting full model:
Iteration 0: log likelihood = -8922.145
Iteration 1: log likelihood = -8853.6409
Iteration 2: log likelihood = -8853.4255
Iteration 3: log likelihood = -8853.4254
Random-effects ML regression Number of obs = 28091
Group variable: idcode Number of groups = 4697
Random effects u_i ~ Gaussian Obs per group: min = 1
                             avg = 6.0
                             max = 15
LR chi2(10) = 7592.38
Log likelihood = -8853.4254 Prob > chi2 = 0.0000

| ln_wage          | Coef.  | Std. Err. |      z  | P>|z| | [95% Conf. Interval] |
|------------------|--------|-----------|---------|------|----------------------|
| grade            | .0646093 | .0017372 | 37.19   | 0.000| .0612044 .0680142    |
| age              | .0368531 | .0031226 | 11.80   | 0.000| .030733 .0429732     |
| c.age#c.age      | -.0007132 | .0000501 | -14.24  | 0.000| -.0008113 -.000615   |
| ttl_exp          | .0288196 | .0024143 | 11.94   | 0.000| .0240877 .0335515    |
| c.ttl_exp#c.ttl_exp | .000309 | .0001163 | 2.66    | 0.008| .0000811 .0005369    |
| tenure           | .0394371 | .0017604 | 22.40   | 0.000| .0359868 .0428875    |
| c.tenure#c.tenure | -.0020052 | .0001195 | -16.77  | 0.000| -.0022395 -.0017709  |
| race             |         |           |         |      |                      |
| black            | -.0533394 | .0097338 | -5.48   | 0.000| -.0724172 -.0342615  |
| not_smsa         | -.1323433 | .0071322 | -18.56  | 0.000| -.1463221 -.1183644  |
| south            | -.0875959 | .0072143 | -12.14  | 0.000| -.1016998 -.0734201  |
| _cons            | .2390837 | .0491902 | 4.86    | 0.000| .1426727 .3354947    |
| /sigma_u         | .2485556 | .0035017 |         |      | .2417863 .2555144    |
| /sigma_e         | .2918458 | .001352  |         |      | .289208 .2945076     |
| rho              | .4204033 | .0074828 |         |      | .4057959 .4351212    |
```

Likelihood-ratio test of sigma_u=0: chibar2(01)= 7339.84 Prob>=chibar2 = 0.000
The estimates are nearly the same as those produced by \texttt{xtreg, re}—the GLS estimator. For instance, \texttt{xtreg, re} estimated the coefficient on \textit{grade} to be 0.0646499, \texttt{xtreg, mle} estimated 0.0646093, and the ratio is 0.0646499/0.0646093 = 1.001 to three decimal places. Similarly, the standard errors are nearly equal: 0.0017811/0.0017372 = 1.025. Below we compare all 11 coefficients:

<table>
<thead>
<tr>
<th>Estimator</th>
<th>Coefficient ratio</th>
<th>SE ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mean</td>
<td>min.</td>
</tr>
<tr>
<td>\texttt{xtreg, mle} (ML)</td>
<td>1.</td>
<td>1.</td>
</tr>
<tr>
<td>\texttt{xtreg, re} (GLS)</td>
<td>.997</td>
<td>.987</td>
</tr>
</tbody>
</table>

Example 6: Population-averaged model

We could also have fit this model with the population-averaged estimator:

\begin{verbatim}
. xtregr ln_w grade age c.age#c.age ttl_exp c.ttl_exp#c.ttl_exp tenure > c.tenure#c.tenure 2.race not_smsa south, pa
\end{verbatim}

Iteration 1: tolerance = .0310561
Iteration 2: tolerance = .00074898
Iteration 3: tolerance = .0000147
Iteration 4: tolerance = 2.880e-07

GEE population-averaged model

<table>
<thead>
<tr>
<th>Obs per group:</th>
<th>Number of obs = 28091</th>
<th>Group variable:</th>
<th>idcode</th>
<th>Number of groups = 4697</th>
</tr>
</thead>
<tbody>
<tr>
<td>Link:</td>
<td>identity</td>
<td>Obs per group:</td>
<td>min = 1</td>
<td>avg = 6.0</td>
</tr>
<tr>
<td>Family:</td>
<td>Gaussian</td>
<td>max = 15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Correlation:</td>
<td>exchangeable</td>
<td>Wald chi2(10)</td>
<td>9598.89</td>
<td></td>
</tr>
<tr>
<td>Scale parameter:</td>
<td>.1436709</td>
<td>Prob &gt; chi2</td>
<td>0.0000</td>
<td></td>
</tr>
</tbody>
</table>

\begin{verbatim}
ln_wage     Coef.  Std. Err.  z     P>|z|     [95% Conf. Interval]  
grade       .0645427  .0016829  38.35  0.000   .0612442   .0678412  
age       .036932   .0031509  11.72  0.000   .0307564   .0431076  
c.age#c.age -.0007129  .0000506 -14.10  0.000  -.0008121  -.0006138  
ttl_exp    .0284878  .0024169  11.79  0.000   .0237508   .0332248  
c.ttl_exp# \texttt{c.ttl_exp}  .0003158  .0001172  2.69  0.007     .000086   .0005456  
tenure     .0397468  .0017779  22.36  0.000   .0362621   .0432315  
c.tenure# \texttt{c.tenure}  -.002008  .0001209 -16.61  0.000  -.0022449  -.0017711  
race       -black  -.0538314  .0094086  -5.72  0.000  -.0722720  -.0353909  
not_smsa   -.1347788  .0070543  -19.11  0.000  -.1486049  -.1209526  
south      -.0885969  .0071132  -12.46  0.000  -.1025386  -.0746552  
_cons     .2396286  .0491465  4.88  0.000   .1433034   .3359539  
\end{verbatim}

These results differ from those produced by \texttt{xtreg, re} and \texttt{xtreg, mle}. Coefficients are larger and standard errors smaller. \texttt{xtreg, pa} is simply another way to run the \texttt{xtgee} command. That is, we would have obtained the same output had we typed
See [XT] xtgee. In the language of xtgee, the random-effects model corresponds to an exchangeable correlation structure and identity link, and xtgee also allows other correlation structures. Let’s stay with the random-effects model, however. xtgee will also produce robust estimates of variance, and we refit this model that way by typing

\[
.\text{xtgee ln_w grade age c.age#c.age ttl_exp c.ttl_exp#c.ttl_exp} \\
> \text{tenure c.tenure#c.tenure 2.race not_smsa south, vce(robust)} \\
\text{(output omitted, coefficients the same, standard errors different)}
\]

In the previous example, we presented a table comparing xtreg, re with xtreg, mle. Below we add the results from the estimates shown and the ones we did with xtgee, vce(robust):

<table>
<thead>
<tr>
<th>Estimator</th>
<th>Coefficient ratio</th>
<th>SE ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mean</td>
<td>min.</td>
</tr>
<tr>
<td>xtreg, mle</td>
<td>(ML)</td>
<td>1.</td>
</tr>
<tr>
<td>xtreg, re</td>
<td>(GLS)</td>
<td>.997</td>
</tr>
<tr>
<td>xtreg, pa</td>
<td>(PA)</td>
<td>1.060</td>
</tr>
<tr>
<td>xtgee, vce(robust)</td>
<td>(PA)</td>
<td>1.060</td>
</tr>
</tbody>
</table>

So, which are right? This is a real dataset, and we do not know. However, in example 2 in [XT] xtreg postestimation, we will present evidence that the assumptions underlying the xtreg, re and xtreg, mle results are not met.
xtreg, re stores the following in e():

Scalars

- e(N)  number of observations
- e(N_g) number of groups
- e(df_m) model degrees of freedom
- e(g_min) smallest group size
- e(g_avg) average group size
- e(g_max) largest group size
- e(Tcon) 1 if $T$ is constant
- e(sigma) ancillary parameter ($\text{gamma}, \text{lnormal}$)
- e(sigma_u) panel-level standard deviation
- e(sigma_e) standard deviation of $\epsilon_{it}$
- e(r2_w) $R^2$-squared for within model
- e(r2_o) $R^2$-squared for overall model
- e(r2_b) $R^2$-squared for between model
- e(N_clust) number of clusters
- e(chi2) $\chi^2$
- e(p) significance
- e(rho) $\rho$
- e(omega_min) minimum $\theta$
- e(omega_5) $\theta$, 5th percentile
- e(omega_50) $\theta$, 50th percentile
- e(omega_95) $\theta$, 95th percentile
- e(omega_max) maximum $\theta$
- e(rmse) root mean squared error of GLS regression
- e(Tbar) harmonic mean of group sizes
- e(rank) rank of e(V)

Macros

- e(cmd) xtreg
- e(cmdline) command as typed
- e(depvar) name of dependent variable
- e(ivar) variable denoting groups
- e(model) re
- e(clustvar) name of cluster variable
- e(chi2type) Wald; type of model $\chi^2$ test
- e(vce) vcetype specified in vce()
- e(vcetype) title used to label Std. Err.
- e(sa) Swamy–Arora estimator of the variance components (sa only)
- e(properties) b V
- e(predict) program used to implement predict
- e(marginsnotok) predictions disallowed by margins
- e(asbalanced) factor variables fvset as asbalanced
- e(asobserved) factor variables fvset as asobserved

Matrices

- e(b) coefficient vector
- e(bf) coefficient vector for fixed-effects model
- e(theta) $\theta$
- e(V) variance–covariance matrix of the estimators
- e(VCEf) VCE for fixed-effects model

Functions

- e(sample) marks estimation sample
**xtreg**, be stores the following in \texttt{e()}:

**Scalars**

- \texttt{e(N)}  number of observations
- \texttt{e(N_g)}  number of groups
- \texttt{e(typ)}  \(\hat{W}LS\), if \(\hat{W}\hat{LS}\) specified
- \texttt{e(mss)}  model sum of squares
- \texttt{e(df_m)}  model degrees of freedom
- \texttt{e(rss)}  residual sum of squares
- \texttt{e(df_r)}  residual degrees of freedom
- \texttt{e(ll)}  log likelihood
- \texttt{e(ll_0)}  log likelihood, constant-only model
- \texttt{e(g_min)}  smallest group size
- \texttt{e(g_avg)}  average group size
- \texttt{e(g_max)}  largest group size
- \texttt{e(Tcon)}  1 if \(T\) is constant
- \texttt{e(r2)}  \(R^2\)
- \texttt{e(r2_a)}  adjusted \(R^2\)
- \texttt{e(r2_w)}  \(R^2\)-squared for within model
- \texttt{e(r2_o)}  \(R^2\)-squared for overall model
- \texttt{e(r2_b)}  \(R^2\)-squared for between model
- \texttt{e(F)}  \(F\) statistic
- \texttt{e(rmse)}  root mean squared error
- \texttt{e(Tbar)}  harmonic mean of group sizes
- \texttt{e(rank)}  rank of \texttt{e(V)}

** Macros**

- \texttt{e(cmd)}  \texttt{xtreg}
- \texttt{e(cmdline)}  command as typed
- \texttt{e(depvar)}  name of dependent variable
- \texttt{e(ivar)}  variable denoting groups
- \texttt{e(model)}  be
- \texttt{e(title)}  title in estimation output
- \texttt{e(vcetype)}  \texttt{vcetype} specified in \texttt{vce()}
- \texttt{e(vce)}  \texttt{vcetype} specified in \texttt{vce()}
- \texttt{e(properties)}  \texttt{b V}
- \texttt{e(predict)}  program used to implement predict
- \texttt{e(marginsok)}  predictions allowed by \texttt{margins}
- \texttt{e(marginsnotok)}  predictions disallowed by \texttt{margins}
- \texttt{e(asbalanced)}  factor variables \texttt{fvset} as \texttt{asbalanced}
- \texttt{e(asobserved)}  factor variables \texttt{fvset} as \texttt{asobserved}

** Matrices**

- \texttt{e(b)}  coefficient vector
- \texttt{e(V)}  variance–covariance matrix of the estimators

** Functions**

- \texttt{e(sample)}  marks estimation sample
xtreg, fe stores the following in e():

Scalars
- e(N)  
  number of observations
- e(N_g)  
  number of groups
- e(mss)  
  model sum of squares
- e(df_m)  
  model degrees of freedom
- e(rss)  
  residual sum of squares
- e(df_r)  
  residual degrees of freedom
- e(tss)  
  total sum of squares
- e(g_min)  
  smallest group size
- e(g_avg)  
  average group size
- e(g_max)  
  largest group size
- e(Tcon)  
  1 if T is constant
- e(sigma)  
  ancillary parameter (gamma, lnormal)
- e(corr)  
  corr(u_i, Xb)
- e(sigma_u)  
  panel-level standard deviation
- e(sigma_e)  
  standard deviation of \( \epsilon_{it} \)
- e(r2)  
  R-squared
- e(r2_a)  
  adjusted R-squared
- e(r2_w)  
  R-squared for within model
- e(r2_o)  
  R-squared for overall model
- e(r2_b)  
  R-squared for between model
- e(ll)  
  log likelihood
- e(ll_0)  
  log likelihood, constant-only model
- e(N_cluster)  
  number of clusters
- e(rho)  
  \( \rho \)
- e(F)  
  F statistic
- e(F_f)  
  F for \( u_i = 0 \)
- e(df_a)  
  degrees of freedom for absorbed effect
- e(df_b)  
  numerator degrees of freedom for F statistic
- e(rmse)  
  root mean squared error
- e(Tbar)  
  harmonic mean of group sizes
- e(rank)  
  rank of e(V)

Macros
- e(cmd)  
  xtreg
- e(cmdline)  
  command as typed
- e(depvar)  
  name of dependent variable
- e(ivar)  
  variable denoting groups
- e(model)  
  fe
- e(wtype)  
  weight type
- e(wexp)  
  weight expression
- e(clustvar)  
  name of cluster variable
- e(vce)  
  vcetype specified in vce()
- e(vcetype)  
  title used to label Std. Err.
- e(properties)  
  b V
- e(predict)  
  program used to implement predict
- e(marginsnotok)  
  predictions disallowed by margins
- e(asbalanced)  
  factor variables fvset as asbalanced
- e(asobserved)  
  factor variables fvset as asobserved

Matrices
- e(b)  
  coefficient vector
- e(Cns)  
  constraints matrix
- e(V)  
  variance–covariance matrix of the estimators

Functions
- e(sample)  
  marks estimation sample
xtreg, mle stores the following in `e()`: 

Scalars 
- `e(N)` number of observations 
- `e(N_g)` number of groups 
- `e(df_m)` model degrees of freedom 
- `e(g_min)` smallest group size 
- `e(g_avg)` average group size 
- `e(g_max)` largest group size 
- `e(sigma_u)` panel-level standard deviation 
- `e(sigma_e)` standard deviation of $\epsilon_{it}$ 
- `e(ll)` log likelihood 
- `e(ll_0)` log likelihood, constant-only model 
- `e(ll_c)` log likelihood, comparison model 
- `e(chi2)` $\chi^2$ 
- `e(chi2_c)` $\chi^2$ for comparison test 
- `e(rho)` $\rho$ 
- `e(rank)` rank of `e(V)` 

Macros 
- `e(cmd)` `xtreg` 
- `e(cmdline)` command as typed 
- `e(depvar)` name of dependent variable 
- `e(ivar)` variable denoting groups 
- `e(model)` ml 
- `e(wtype)` weight type 
- `e(wexp)` weight expression 
- `e(title)` title in estimation output 
- `e(vce)` `vcetype` specified in `vce()` 
- `e(vcetype)` title used to label Std. Err. 
- `e(chi2type)` Wald or LR; type of model $\chi^2$ test 
- `e(chi2_ct)` Wald or LR; type of model $\chi^2$ test corresponding to `e(chi2_c)` 
- `e(distrib)` Gaussian; the distribution of the RE 
- `e(properties)` b V 
- `e(predict)` program used to implement predict 
- `e(marginsnotok)` predictions disallowed by `margins` 
- `e(asbalanced)` factor variables `fvset` as `asbalanced` 
- `e(asobserved)` factor variables `fvset` as `asobserved` 

Matrices 
- `e(b)` coefficient vector 
- `e(V)` variance–covariance matrix of the estimators 

Functions 
- `e(sample)` marks estimation sample
xtreg, pa stores the following in e():

Scalars
- e(N): number of observations
- e(N_g): number of groups
- e(df_m): model degrees of freedom
- e(chi2): $\chi^2$
- e(p): significance
- e(df_pear): degrees of freedom for Pearson $\chi^2$
- e(chi2_dev): $\chi^2$ test of deviance
- e(chi2_dis): $\chi^2$ test of deviance dispersion
- e(deviance): deviance
- e(dispers): deviance dispersion
- e(phi): scale parameter
- e(g_min): smallest group size
- e(g_avg): average group size
- e(g_max): largest group size
- e(rank): rank of e(V)
- e(tol): target tolerance
- e(dif): achieved tolerance
- e(rc): return code

Macros
- e(cmd): xtgee
- e(cmd2): xtreg
- e(cmdline): command as typed
- e(depvar): name of dependent variable
- e(ivar): variable denoting groups
- e(tvar): variable denoting time within groups
- e(model): pa
- e(family): Gaussian
- e(link): identity; link function
- e(corr): correlation structure
- e(scale): x2, dev, phi, or #; scale parameter
- e(wtype): weight type
- e(wexp): weight expression
- e(offset): linear offset variable
- e(chi2type): Wald; type of model $\chi^2$ test
- e(vce): vcetype specified in vce()
- e(vcetype): title used to label Std. Err.
- e(rgf): rgf, if rgf specified
- e(nmp): nmp, if specified
- e(properties): b V
- e(predict): program used to implement predict
- e(marginsnotok): predictions disallowed by margins
- e(asbalanced): factor variables fvset as asbalanced
- e(asobserved): factor variables fvset as asobserved

Matrices
- e(b): coefficient vector
- e(R): estimated working correlation matrix
- e(V): variance–covariance matrix of the estimators
- e(V_modelbased): model-based variance

Functions
- e(sample): marks estimation sample

Methods and formulas

The model to be fit is

$$y_{it} = \alpha + x_{it}\beta + \nu_i + \epsilon_{it}$$

for $i = 1, \ldots, n$ and, for each $i, t = 1, \ldots, T$, of which $T_i$ periods are actually observed.
xtreg — Fixed-, between-, and random-effects and population-averaged linear models

Methods and formulas are presented under the following headings:
- xtreg, fe
- xtreg, be
- xtreg, re
- xtreg, mle
- xtreg, pa

xtreg, fe

_xtreg, fe produces estimates by running OLS on_

\[(y_{it} - \bar{y}_i + \bar{y}) = \alpha + (x_{it} - \bar{x} + \bar{x})\beta + (\epsilon_{it} - \bar{\epsilon}_i + \bar{\nu}) + \bar{\epsilon}\]

where \(\bar{y}_i = \sum_{t=1}^{T_i} y_{it}/T_i\), and similarly, \(\bar{y} = \sum_i \sum_t y_{it}/(nT_i)\). The conventional covariance matrix of the estimators is adjusted for the extra \(n-1\) estimated means, so results are the same as using OLS on (1) to estimate \(\nu_i\) directly. Specifying vce(robust) or vce(cluster clustvar) causes the Huber/White/sandwich VCE estimator to be calculated for the coefficients estimated in this regression. See [P] _robust, particularly Introduction and Methods and formulas. Wooldridge (2013) and Arellano (2003) discuss this application of the Huber/White/sandwich VCE estimator. As discussed by Wooldridge (2013), Stock and Watson (2008), and Arellano (2003), specifying vce(robust) is equivalent to specifying vce(cluster panelvar), where panelvar is the variable that identifies the panels.

Clustering on the panel variable produces a consistent VCE estimator when the disturbances are not identically distributed over the panels or there is serial correlation in \(\epsilon_{it}\).

The cluster–robust VCE estimator requires that there are many clusters and the disturbances are uncorrelated across the clusters. The panel variable must be nested within the cluster variable because of the within-panel correlation induced by the within transform. From the estimates \(\hat{\alpha}\) and \(\hat{\beta}\), estimates \(u_i\) of \(\nu_i\) are obtained as \(u_i = \bar{y}_i - \hat{\alpha} - \bar{x}_i\hat{\beta}\). Reported from the calculated \(u_i\) are its standard deviation and its correlation with \(\bar{x}_i\hat{\beta}\). Reported as the standard deviation of \(e_{it}\) is the regression’s estimated root mean squared error, \(s\), which is adjusted (as previously stated) for the \(n-1\) estimated means.

- Reported as \(R^2\) within is the \(R^2\) from the mean-deviated regression.
- Reported as \(R^2\) between is \(\text{corr}(\bar{x}_i\hat{\beta}, y_{it})^2\).
- Reported as \(R^2\) overall is \(\text{corr}(x_{it}\hat{\beta}, y_{it})^2\).

xtreg, be

_xtreg, be fits the following model:_

\[\bar{y}_i = \alpha + \bar{x}_i\beta + \nu_i + \epsilon_i\]

Estimation is via OLS unless \(T_i\) is not constant and the wls option is specified. Otherwise, the estimation is performed via WLS. The estimates and conventional VCE are obtained from regress for both cases, but for WLS, [aweight=T_i] is specified.

- Reported as \(R^2\) between is the \(R^2\) from the fitted regression.
- Reported as \(R^2\) within is \(\text{corr}\{(x_{it} - \bar{x}_i)\hat{\beta}, y_{it} - \bar{y}_i\}^2\).
- Reported as \(R^2\) overall is \(\text{corr}(x_{it}\hat{\beta}, y_{it})^2\).
xtreg, re

The key to the random-effects estimator is the GLS transform. Given estimates of the idiosyncratic component, \( \hat{\sigma}_e^2 \), and the individual component, \( \hat{\sigma}_u^2 \), the GLS transform of a variable \( z \) for the random-effects model is

\[
z_{it}^* = z_{it} - \hat{\theta}_i \bar{z}_i
\]

where \( \bar{z}_i = \frac{1}{T_i} \sum_{t} T_i z_{it} \) and

\[
\hat{\theta}_i = 1 - \sqrt{\frac{\hat{\sigma}_e^2}{T_i \hat{\sigma}_u^2 + \hat{\sigma}_e^2}}
\]

Given an estimate of \( \hat{\theta}_i \), one transforms the dependent and independent variables, and then the coefficient estimates and the conventional variance–covariance matrix come from an OLS regression of \( y_{it}^* \) on \( x_{it}^* \) and the transformed constant \( 1 - \hat{\theta}_i \). Specifying vce(robust) or vce(cluster clustvar) causes the Huber/White/sandwich VCE estimator to be calculated for the coefficients estimated in this regression. See [P] _robust; in particular, see Introduction and Methods and formulas. Wooldridge (2013) and Arellano (2003) discuss this application of the Huber/White/sandwich VCE estimator. As discussed by Wooldridge (2013), Stock and Watson (2008), and Arellano (2003), specifying vce(robust) is equivalent to specifying vce(cluster panelvar), where panelvar is the variable that identifies the panels.

Clustering on the panel variable produces a consistent VCE estimator when the disturbances are not identically distributed over the panels or there is serial correlation in \( \epsilon_{it} \).

The cluster–robust VCE estimator requires that there are many clusters and the disturbances are uncorrelated across the clusters. The panel variable must be nested within the cluster variable because of the within-panel correlation that is generally induced by the random-effects transform when there is heteroskedasticity or within-panel serial correlation in the idiosyncratic errors.

Stata has two implementations of the Swamy–Arora method for estimating the variance components. They produce the same results in balanced panels and share the same estimator of \( \hat{\sigma}_e^2 \). However, the two methods differ in their estimator of \( \hat{\sigma}_u^2 \) in unbalanced panels. We call the first \( \hat{\sigma}_{uT}^2 \) and the second \( \hat{\sigma}_{uSA}^2 \). Both estimators are consistent; however, \( \hat{\sigma}_{uSA}^2 \) has a more elaborate adjustment for small samples than \( \hat{\sigma}_{uT}^2 \). (See Baltagi [2013], Baltagi and Chang [1994], and Swamy and Arora [1972] for derivations of these methods.)

Both methods use the same function of within residuals to estimate the idiosyncratic error component \( \sigma_e \). Specifically,

\[
\hat{\sigma}_e^2 = \frac{\sum_i \sum_{t} e_{it}^2}{N - n - K + 1}
\]

where

\[
e_{it} = (y_{it} - \bar{y}_i + \bar{y}) - \hat{\alpha}_w - (x_{it} - \bar{x}_i + \bar{x})\hat{\beta}_w
\]

and \( \hat{\alpha}_w \) and \( \hat{\beta}_w \) are the within estimates of the coefficients and \( N = \sum_i T_i \). After passing the within residuals through the within transform, only the idiosyncratic errors are left.
The default method for estimating $\sigma_u^2$ is

$$\hat{\sigma}_{uT}^2 = \max \left\{ 0, \frac{SSR_b}{n-K} - \frac{\hat{\sigma}_e^2}{\bar{T}} \right\}$$

where

$$SSR_b = \sum_{i}^n \left( \bar{y}_i - \hat{\alpha}_b - \bar{x}_i \hat{\beta}_b \right)^2$$

$\hat{\alpha}_b$ and $\hat{\beta}_b$ are coefficient estimates from the between regression and $\bar{T}$ is the harmonic mean of $T_i$:

$$\bar{T} = \frac{n}{\sum_i^n \frac{1}{T_i}}$$

This estimator is consistent for $\sigma_u^2$ and is computationally less expensive than the second method. The sum of squared residuals from the between model estimate a function of both the idiosyncratic component and the individual component. Using our estimator of $\sigma_e^2$, we can remove the idiosyncratic component, leaving only the desired individual component.

The second method is the Swamy–Arora method for unbalanced panels derived by Baltagi and Chang (1994), which has a more precise small-sample adjustment. Using this method,

$$\hat{\sigma}_{uSA}^2 = \max \left\{ 0, \frac{SSR_b - (n-K)\hat{\sigma}_e^2}{N-tr} \right\}$$

where

$$tr = \text{trace} \left\{ (X'P X)^{-1} X' Z Z' X \right\}$$

$$P = \text{diag} \left\{ \left( \frac{1}{T_i} \right) \nu_{T_i} \nu_{T_i}' \right\}$$

$$Z = \text{diag} \left[ \nu_{T_i} \right]$$

$X$ is the $N \times K$ matrix of covariates, including the constant, and $\nu_{T_i}$ is a $T_i \times 1$ vector of ones.

The estimated coefficients $(\hat{\alpha}_r, \hat{\beta}_r)$ and their covariance matrix $V_r$ are reported together with the previously calculated quantities $\hat{\sigma}_e$ and $\hat{\sigma}_u$. The standard deviation of $\nu_i + e_{it}$ is calculated as $\sqrt{\hat{\sigma}_e^2 + \hat{\sigma}_u^2}$.

Reported as $R^2$ between is $\text{corr}(\bar{x}_i \hat{\beta}, \bar{y}_i)^2$.

Reported as $R^2$ within is $\text{corr} \left\{ (x_{it} - \bar{x}_i) \hat{\beta}, y_{it} - \bar{y}_i \right\}^2$.

Reported as $R^2$ overall is $\text{corr}(x_{it} \hat{\beta}, y_{it})^2$. 
xtreg, mle

The log likelihood for the \( i \)th unit is

\[
l_i = -\frac{1}{2} \left( \frac{1}{\sigma_e^2} \left[ \sum_{t=1}^{T_i} (y_{it} - x_{it}\beta)^2 - \frac{\sigma_u^2}{T_i\sigma_e^2 + \sigma_u^2} \left\{ \sum_{t=1}^{T_i} (y_{it} - x_{it}\beta) \right\}^2 \right]
+ \ln \left( \frac{T_i \sigma_u^2}{\sigma_e^2} + 1 \right) + T_i \ln(2\pi \sigma_e^2) \right)
\]

The \texttt{mle} and \texttt{re} options yield essentially the same results, except when total \( N = \sum_i T_i \) is small (200 or less) and the data are unbalanced.

xtreg, pa

See \texttt{[XT] xtgee} for details on the methods and formulas used to calculate the population-averaged model using a generalized estimating equations approach.

Acknowledgments

We thank Richard Goldstein, who wrote the first draft of the routine that fits random-effects regressions, Badi Baltagi of the Department of Economics at Syracuse University, and Manuelita Ureta of the Department of Economics at Texas A&M University, who assisted us in working our way through the literature.

References


xtreg — Fixed-, between-, and random-effects and population-averaged linear models  

Also see

[XT] xtreg postestimation — Postestimation tools for xtreg
[XT] xtgee — Fit population-averaged panel-data models by using GEE
[XT] xtgls — Fit panel-data models by using GLS
[XT] xtivreg — Instrumental variables and two-stage least squares for panel-data models
[XT] xtregar — Fixed- and random-effects linear models with an AR(1) disturbance
[XT] xtset — Declare data to be panel data
[ME] mixed — Multilevel mixed-effects linear regression
[MII] estimation — Estimation commands for use with mi estimate
[R] areg — Linear regression with a large dummy-variable set
[R] regress — Linear regression
[TS] forecast — Econometric model forecasting
[TS] prais — Prais–Winsten and Cochrane–Orcutt regression
[U] 20 Estimation and postestimation commands
The following postestimation commands are of special interest after \texttt{xtreg}:

<table>
<thead>
<tr>
<th>Command</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>\texttt{xttest0}</td>
<td>Breusch and Pagan LM test for random effects</td>
</tr>
</tbody>
</table>

The following standard postestimation commands are also available:

<table>
<thead>
<tr>
<th>Command</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>\texttt{contrast}</td>
<td>contrasts and ANOVA-style joint tests of estimates</td>
</tr>
<tr>
<td>\texttt{estat ic}¹</td>
<td>Akaike’s and Schwarz’s Bayesian information criteria (AIC and BIC)</td>
</tr>
<tr>
<td>\texttt{estat summarize}</td>
<td>summary statistics for the estimation sample</td>
</tr>
<tr>
<td>\texttt{estat vce}</td>
<td>variance–covariance matrix of the estimators (VCE)</td>
</tr>
<tr>
<td>\texttt{estimates}</td>
<td>cataloging estimation results</td>
</tr>
<tr>
<td>\texttt{forecast}²</td>
<td>dynamic forecasts and simulations</td>
</tr>
<tr>
<td>\texttt{hausman}</td>
<td>Hausman’s specification test</td>
</tr>
<tr>
<td>\texttt{lincom}</td>
<td>point estimates, standard errors, testing, and inference for linear combinations of coefficients</td>
</tr>
<tr>
<td>\texttt{lrtest}</td>
<td>likelihood-ratio test</td>
</tr>
<tr>
<td>\texttt{margins}</td>
<td>marginal means, predictive margins, marginal effects, and average marginal effects</td>
</tr>
<tr>
<td>\texttt{marginsplot}</td>
<td>graph the results from margins (profile plots, interaction plots, etc.)</td>
</tr>
<tr>
<td>\texttt{nlcom}</td>
<td>point estimates, standard errors, testing, and inference for nonlinear combinations of coefficients</td>
</tr>
<tr>
<td>\texttt{predict}</td>
<td>predictions, residuals, influence statistics, and other diagnostic measures</td>
</tr>
<tr>
<td>\texttt{predictnl}</td>
<td>point estimates, standard errors, testing, and inference for generalized predictions</td>
</tr>
<tr>
<td>\texttt{pwcompare}</td>
<td>pairwise comparisons of estimates</td>
</tr>
<tr>
<td>\texttt{test}</td>
<td>Wald tests of simple and composite linear hypotheses</td>
</tr>
<tr>
<td>\texttt{testnl}</td>
<td>Wald tests of nonlinear hypotheses</td>
</tr>
</tbody>
</table>

¹ \texttt{estat ic} is not appropriate after \texttt{xtreg} with the \texttt{be}, \texttt{pa}, or \texttt{re} option.
² \texttt{forecast} is not appropriate with \texttt{mi} estimation results.
Special-interest postestimation commands

`xttest0`, for use after `xtreg, re`, presents the Breusch and Pagan (1980) Lagrange multiplier test for random effects, a test that \( \text{Var}(\nu_i) = 0 \).

Syntax for predict

For all but the population-averaged model

```
predict [type] newvar [if] [in] [, statistic nooffset]
```

Population-averaged model

```
predict [type] newvar [if] [in] [, PA_statistic nooffset]
```

<table>
<thead>
<tr>
<th>statistic</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Main</td>
<td></td>
</tr>
<tr>
<td><code>xb</code></td>
<td>( x_jb ), fitted values; the default</td>
</tr>
<tr>
<td><code>stdp</code></td>
<td>standard error of the fitted values</td>
</tr>
<tr>
<td><code>ue</code></td>
<td>( u_i + e_{it} ), the combined residual</td>
</tr>
<tr>
<td><code>*xbu</code></td>
<td>( x_jb + u_i ), prediction including effect</td>
</tr>
<tr>
<td><code>*u</code></td>
<td>( u_i ), the fixed- or random-error component</td>
</tr>
<tr>
<td><code>*e</code></td>
<td>( e_{it} ), the overall error component</td>
</tr>
</tbody>
</table>

Unstarred statistics are available both in and out of sample; type `predict ... if e(sample) ...` if wanted only for the estimation sample. Starred statistics are calculated only for the estimation sample, even when `if e(sample)` is not specified.

<table>
<thead>
<tr>
<th>PA_statistic</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Main</td>
<td></td>
</tr>
<tr>
<td><code>mu</code></td>
<td>probability of <code>depvar</code>; considers the <code>offset()</code></td>
</tr>
<tr>
<td><code>rate</code></td>
<td>probability of <code>depvar</code></td>
</tr>
<tr>
<td><code>xb</code></td>
<td>linear prediction</td>
</tr>
<tr>
<td><code>stdp</code></td>
<td>standard error of the linear prediction</td>
</tr>
<tr>
<td><code>score</code></td>
<td>first derivative of the log likelihood with respect to ( x_j\beta )</td>
</tr>
</tbody>
</table>

These statistics are available both in and out of sample; type `predict ... if e(sample) ...` if wanted only for the estimation sample.

Menu for predict

Statistics > Postestimation > Predictions, residuals, etc.

Options for predict

```
xb calculates the linear prediction, that is, \( a + bx_{it} \). This is the default for all except the population-averaged model.
```
stdp calculates the standard error of the linear prediction. For the fixed-effects model, this excludes the variance due to uncertainty about the estimate of $u_i$.

mu and rate both calculate the predicted probability of $depvar$. mu takes into account the offset(), and rate ignores those adjustments. mu and rate are equivalent if you did not specify offset(). mu is the default for the population-averaged model.

ue calculates the prediction of $u_i + e_{it}$.

xbu calculates the prediction of $a + bx_{it} + u_i$, the prediction including the fixed or random component.

e calculates the prediction of $e_{it}$.

score calculates the equation-level score, $u_j = \partial \ln L_j(x_j\beta)/\partial(x_j\beta)$.

nooffset is relevant only if you specified offset(varname) for xtreg, pa. It modifies the calculations made by predict so that they ignore the offset variable; the linear prediction is treated as $x_{it}b$ rather than $x_{it}b + offset_{it}$.

**Syntax for xttest0**

```
xgtest0
```
Example 2

More importantly, after `xtreg`, `re` estimation, `hausman` will perform the Hausman specification test. If our model is correctly specified, and if $\nu_i$ is uncorrelated with $x_{it}$, the (subset of) coefficients that are estimated by the fixed-effects estimator and the same coefficients that are estimated here should not statistically differ:

```stata
     . xtreg ln_w grade age c.age#c.age ttl_exp c.ttl_exp#c.ttl_exp
     > tenure c.tenure#c.tenure 2.race not_smsa south, re
     (output omitted)

     . estimates store random_effects

     . xtreg ln_w grade age c.age#c.age ttl_exp c.ttl_exp#c.ttl_exp
     > tenure c.tenure#c.tenure 2.race not_smsa south, fe
     (output omitted)

     . hausman . random_effects
```

The following table shows the coefficients estimated by `xtreg` with random-effects (`random_effects`) and fixed-effects (`fe`) estimation:

<table>
<thead>
<tr>
<th>Coefficients</th>
<th>(b)</th>
<th>(B)</th>
<th>(b-B)</th>
<th>sqrt(diag(V_b-V_B))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random Effects</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>age</td>
<td>.0359987</td>
<td>.0368059</td>
<td>-.0008073</td>
<td>.0013177</td>
</tr>
<tr>
<td>c.age#c.age</td>
<td>-.000723</td>
<td>-.0007133</td>
<td>-9.68e-06</td>
<td>.000184</td>
</tr>
<tr>
<td>ttl_exp</td>
<td>.0334668</td>
<td>.0290208</td>
<td>.0044597</td>
<td>.001711</td>
</tr>
<tr>
<td>c.ttl_exp#c.tenure</td>
<td>.0002163</td>
<td>.00003049</td>
<td>-.0001866</td>
<td>.000053</td>
</tr>
<tr>
<td>c.tenure#c-e</td>
<td>.0357539</td>
<td>.0392519</td>
<td>-.003498</td>
<td>.0005797</td>
</tr>
<tr>
<td>not_smsa</td>
<td>-.0890108</td>
<td>-.1308252</td>
<td>.0418144</td>
<td>.0062745</td>
</tr>
<tr>
<td>south</td>
<td>-.0606309</td>
<td>-.0868922</td>
<td>.0262613</td>
<td>.0081345</td>
</tr>
</tbody>
</table>

The test statistic is:

```
chi2(8) = (b-B)'[(V_b-V_B)^(-1)](b-B)
= 149.43
Prob>chi2 = 0.0000
```

We can reject the hypothesis that the coefficients are the same. Before turning to what this means, note that `hausman` listed the coefficients estimated by the two models. It did not, however, list `grade` and `2.race`. `hausman` did not make a mistake; in the Hausman test, we compare only the coefficients estimated by both techniques.

What does this mean? We have an unpleasant choice: we can admit that our model is misspecified—that we have not parameterized it correctly—or we can hold that our specification is correct, in which case the observed differences must be due to the zero correlation of $\nu_i$ and the $x_{it}$ assumption.

Technical note

We can also mechanically explore the underpinnings of the test’s dissatisfaction. In the comparison table from `hausman`, it is the coefficients on `not_smsa` and `south` that exhibit the largest differences. In equation (1’) of [XT] `xtreg`, we showed how to decompose a model into within and between effects. Let’s do that with these two variables, assuming that changes in the average have one effect, whereas transitional changes have another:
```stata
. egen avgnsmsa = mean(not_smsa), by(id)
. generate devnsma = not_smsa -avgnsmsa
(8 missing values generated)
. egen avgsouth = mean(south), by(id)
. generate devsou = south - avgsouth
(8 missing values generated)

. xtreg ln_w grade age c.age#c.age ttl_exp c.ttl_exp#c.ttl_exp tenure c.tenure#
> c.tenure 2.race avgnsm devnsm avgsou devsou

Random-effects GLS regression  Number of obs = 28091
Group variable: idcode       Number of groups = 4697
R-sq: within = 0.1723   Obs per group: min =  1
between = 0.4809        avg =  6.0
overall = 0.3737        max =  15

Wald chi2(12) = 9319.56  corr(u_i, X) = 0 (assumed)
Prob > chi2 = 0.0000

| Coef. | Std. Err. | z    | P>|z|  | [95% Conf. Interval] |
|-------|-----------|------|------|----------------------|
| ln_w  |           |      |      |                      |
| grade | 0.0631716 | 0.0017903 | 35.29 | 0.000 | 0.0596627 - 0.0666805 |
| age   | 0.0375196 | 0.0031186 | 12.03 | 0.000 | 0.0314072 - 0.043632 |
| c.age#c.age | -0.0007248 | 0.00005 | -14.50 | 0.000 | -0.0008228 - 0.0006269 |
| ttl_exp | 0.0286543 | 0.0024207 | 11.84 | 0.000 | 0.0239098 - 0.0333989 |
| c.ttl_exp#c.ttl_exp | 0.0003222 | 0.0001162 | 2.77 | 0.006 | 0.0000945 - 0.0005499 |
| tenure | 0.0394423 | 0.001754 | 22.49 | 0.000 | 0.0360044 - 0.0428801 |
| c.tenure#c.tenure | -0.0020081 | 0.0001192 | -16.85 | 0.000 | -0.0022417 - -0.0017746 |
| race  |           |      |      |                      |
| black | -0.0545936 | 0.0102101 | -5.35 | 0.000 | -0.074605 - -0.0345821 |
| avgnsm | -0.1833237 | 0.0109339 | -16.77 | 0.000 | -0.2047537 - -16.18937 |
| devnsm | -0.0887596 | 0.0095071 | -9.34 | 0.000 | -0.1073931 - -0.070126 |
| avgsouth | -0.1011235 | 0.0098789 | -10.24 | 0.000 | -0.1204858 - -0.0817611 |
| devsouth | -0.0598538 | 0.0109054 | -5.49 | 0.000 | -0.081228 - -0.0384797 |
| _cons | 0.2682987 | 0.0495778 | 5.41 | 0.000 | 0.171128 - 0.3654694 |

sigma_u | .2579182 |
sigma_e | .29068923 |
rho     | .44047745 (fraction of variance due to u_i)
```

We will leave the reinterpretation of this model to you, except that if we were really going to sell this model, we would have to explain why the between and within effects are different. Focusing on residence in a non-SMSA, we might tell a story about rural people being paid less and continuing to get paid less when they move to the SMSA. Given our panel data, we could create variables to measure this (an indicator for moved from non-SMSA to SMSA) and to measure the effects. In our assessment of this model, we should think about women in the cities moving to the country and their relative productivity in a bucolic setting.
In any case, the Hausman test now is

\begin{verbatim}
. estimates store new_random_effects
. xtreg ln_w grade age c.age#c.age ttl_exp c.ttl_exp#c.ttl_exp
  > tenure c.tenure#c.tenure 2.race avgnsm devnsm avgsou devsou, fe
  (output omitted)
. hausman . new_random_effects
\end{verbatim}

<table>
<thead>
<tr>
<th>Coefficients</th>
<th>(b)</th>
<th>(B)</th>
<th>(b-B)</th>
<th>sqrt(diag(V_b-V_B))</th>
</tr>
</thead>
<tbody>
<tr>
<td>new_random_s</td>
<td>.0359987</td>
<td>.0375196</td>
<td>-.0015209</td>
<td>.0013198</td>
</tr>
<tr>
<td>c.age#c.age</td>
<td>-.000723</td>
<td>-.0007248</td>
<td>1.84e-06</td>
<td>.0000184</td>
</tr>
<tr>
<td>ttl_exp</td>
<td>.0334668</td>
<td>.0286543</td>
<td>.0048124</td>
<td>.0017127</td>
</tr>
<tr>
<td>c.ttl_exp#c.ttl_exp</td>
<td>.0357539</td>
<td>.0394423</td>
<td>-.0036884</td>
<td>.0005839</td>
</tr>
<tr>
<td>tenure</td>
<td>-.0890108</td>
<td>-.0887596</td>
<td>-.0002512</td>
<td>.000683</td>
</tr>
<tr>
<td>c.tenure#c.tenure</td>
<td>-.0606309</td>
<td>-.0598538</td>
<td>-.0007771</td>
<td>.0007618</td>
</tr>
</tbody>
</table>

\( b = \text{consistent under } H_0 \text{ and } H_A; \text{ obtained from } \text{xtreg} \)
\( B = \text{inconsistent under } H_A, \text{ efficient under } H_0; \text{ obtained from } \text{xtreg} \)

Test: \( H_0: \text{difference in coefficients not systematic} \)
\[
\chi^2(n) = (n - 1) \left( \sum_{i=1}^{n} \left( \sum_{t=1}^{T_i} v_{it}^2 \right) \right)
\]
\[
= 92.52
\]
Prob\( > \chi^2 = 0.0000
\]

We have mechanically succeeded in greatly reducing the \( \chi^2 \), but not by enough. The major differences now are in the age, experience, and tenure effects. We already knew this problem existed because of the ever-increasing effect of experience. More careful parameterization work rather than simply including squares needs to be done.

\section*{Methods and formulas}

\texttt{xttest0} reports the Lagrange multiplier test for random effects developed by Breusch and Pagan (1980) and as modified by Baltagi and Li (1990). The model
\[
y_{it} = \alpha + \mathbf{x}_{it} \beta + \nu_{it}
\]
is fit via OLS, and then the quantity
\[
\lambda_{LM} = \frac{(n - 1)^2}{2} \left( \frac{A_1^2}{\left( \sum_{i=1}^{n} T_i \right) - nT} \right)
\]
is calculated, where
\[
A_1 = 1 - \frac{\sum_{i=1}^{n} \left( \sum_{t=1}^{T_i} v_{it}^2 \right)}{\sum_{i} \sum_{t} v_{it}^2}
\]
The Baltagi and Li modification allows for unbalanced data and reduces to the standard formula

$$\lambda_{LM} = \begin{cases} \frac{nT}{2(T-1)} \left\{ \sum_i \left( \sum_t v_{it} \right)^2 - 1 \right\}^2, & \hat{\sigma}_u^2 \geq 0 \\ 0, & \hat{\sigma}_u^2 < 0 \end{cases}$$

when $T_i = T$ (balanced data). Under the null hypothesis, $\lambda_{LM}$ is distributed as a 50:50 mixture of a point mass at zero and $\chi^2(1)$.

References


Also see

[XT] xtreg — Fixed-, between-, and random-effects and population-averaged linear models

[U] 20 Estimation and postestimation commands
xtregar — Fixed- and random-effects linear models with an AR(1) disturbance

Syntax

**GLS random-effects (RE) model**

```
xtregar depvar [indepvars] [if] [in] [, re options]
```

**Fixed-effects (FE) model**

```
xtregar depvar [indepvars] [if] [in] [weight], fe [options]
```

### options

<table>
<thead>
<tr>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Model</strong></td>
</tr>
<tr>
<td>re</td>
</tr>
<tr>
<td>fe</td>
</tr>
<tr>
<td>rhotype(rhmethod)</td>
</tr>
<tr>
<td>rhof(#)</td>
</tr>
<tr>
<td>twostep</td>
</tr>
<tr>
<td><strong>Reporting</strong></td>
</tr>
<tr>
<td>level(#)</td>
</tr>
<tr>
<td>lbi</td>
</tr>
<tr>
<td><strong>display_options</strong></td>
</tr>
<tr>
<td><strong>coeflegend</strong></td>
</tr>
</tbody>
</table>

A panel variable and a time variable must be specified; use `xtset`; see [XT] `xtset`.

`indepvars` may contain factor variables; see [U] 11.4.3 Factor variables.

`depvar` and `indepvars` may contain time-series operators; see [U] 11.4.4 Time-series varlists.

by and `statsby` are allowed; see [U] 11.1.10 Prefix commands.

`fweight`s and `aweight`s are allowed for the fixed-effects model with `rhtype(regress)` or `rhtype(freg)`, or with a fixed rho; see [U] 11.1.6 `weight`. Weights must be constant within panel.

`coeflegend` does not appear in the dialog box.

See [U] 20 Estimation and postestimation commands for more capabilities of estimation commands.

Menu

Statistics > Longitudinal/panel data > Linear models > Linear regression with AR(1) disturbance (FE, RE)
Description

xtregar fits cross-sectional time-series regression models when the disturbance term is first-order autoregressive. xtregar offers a within estimator for fixed-effects models and a GLS estimator for random-effects models. Consider the model

\[ y_{it} = \alpha + x_{it}\beta + \nu_i + \epsilon_{it} \quad i = 1, \ldots, N; \quad t = 1, \ldots, T_i \]  

where

\[ \epsilon_{it} = \rho \epsilon_{i,t-1} + \eta_{it} \]  

and where \(|\rho| < 1\) and \(\eta_{it}\) is independent and identically distributed (i.i.d.) with mean 0 and variance \(\sigma_{\eta}^2\). If \(\nu_i\) are assumed to be fixed parameters, the model is a fixed-effects model. If \(\nu_i\) are assumed to be realizations of an i.i.d. process with mean 0 and variance \(\sigma_{\nu}^2\), it is a random-effects model. Whereas in the fixed-effects model, the \(\nu_i\) may be correlated with the covariates \(x_{it}\), in the random-effects model the \(\nu_i\) are assumed to be independent of the \(x_{it}\). On the other hand, any \(x_{it}\) that do not vary over \(t\) are collinear with the \(\nu_i\) and will be dropped from the fixed-effects model. In contrast, the random-effects model can accommodate covariates that are constant over time.

xtregar can accommodate unbalanced panels whose observations are unequally spaced over time. xtregar implements the methods derived in Baltagi and Wu (1999).

Options

**Model**

-re requests the GLS estimator of the random-effects model, which is the default.
-fe requests the within estimator of the fixed-effects model.
-rhotype(rhomethod) allows the user to specify any of the following estimators of \(\rho\):

- **dw** \(\rho_{dw} = 1 - d/2\), where \(d\) is the Durbin–Watson \(d\) statistic
- **regr** \(\rho_{reg} = \beta\) from the residual regression \(\epsilon_t = \beta \epsilon_{t-1}\)
- **freg** \(\rho_{freg} = \beta\) from the residual regression \(\epsilon_t = \beta \epsilon_{t+1}\)
- **tscorr** \(\rho_{tscorr} = \epsilon' \epsilon_{t-1}/\epsilon' \epsilon\), where \(\epsilon\) is the vector of residuals and \(\epsilon_{t-1}\) is the vector of lagged residuals
- **theil** \(\rho_{theil} = \rho_{tscorr}(N - k)/N\)
- **nagar** \(\rho_{nagar} = (\rho_{dw} N^2 + k^2)/(N^2 - k^2)\)
- **onestep** \(\rho_{onestep} = (n/m_c)(\epsilon' \epsilon_{t-1}/\epsilon' \epsilon)\), where \(\epsilon\) is the vector of residuals, \(n\) is the number of observations, and \(m_c\) is the number of consecutive pairs of residuals

- **dw** is the default method. Except for **onestep**, the details of these methods are given in [TS] prais. prais handles unequally spaced data. **onestep** is the one-step method proposed by Baltagi and Wu (1999). More details on this method are available below in Methods and formulas.

-rhof(#) specifies that the given number be used for \(\rho\) and that \(\rho\) not be estimated.

-twostep requests that a two-step implementation of the rhomethod estimator of \(\rho\) be used. Unless a fixed value of \(\rho\) is specified, \(\rho\) is estimated by running prais on the de-meaned data. When twostep is specified, prais will stop on the first iteration after the equation is transformed by \(\rho\)—the two-step efficient estimator. Although it is customary to iterate these estimators to convergence, they are efficient at each step. When twostep is not specified, the FGLS process iterates to convergence as described in the Methods and formulas of [TS] prais.
Reporting

level(#) \textit{R}estimation options.

\texttt{lb1} requests that the Baltagi–Wu (1999) locally best invariant (LBI) test statistic that $\rho = 0$ and a modified version of the Bhargava, Franzini, and Narendranathan (1982) Durbin–Watson statistic be calculated and reported. The default is not to report them. \textit{P}-values are not reported for either statistic. Although Bhargava, Franzini, and Narendranathan (1982) published critical values for their statistic, no tables are currently available for the Baltagi–Wu LBI. Baltagi and Wu (1999) derive a normalized version of their statistic, but this statistic cannot be computed for datasets of moderate size. You can also specify these options upon replay.

\textit{Display options}: noomitted, vsquish, noemptycells, baselevels, allbaselevels, nofvlabel, fvwrap(#), fvwrappon(style), cformat(\%)fmt, pformat(\%)fmt, sformat(\%)fmt, and nolstretch; see \textit{R}estimation options.

The following option is available with \texttt{xtregar} but is not shown in the dialog box: coeflegend; see \textit{R}estimation options.

Remarks and examples

Remarks are presented under the following headings:

- Introduction
- The fixed-effects model
- The random-effects model

Introduction

If you have not read \textit{[XT] xt}, please do so.

Consider a linear panel-data model described by (1) and (2). In the fixed-effects model, the $\nu_i$ are a set of fixed parameters to be estimated. Alternatively, the $\nu_i$ may be random and correlated with the other covariates, with inference conditional on the $\nu_i$ in the sample; see Mundlak (1978) and Hsiao (2003). In the random-effects model, also known as the variance-components model, the $\nu_i$ are assumed to be realizations of an i.i.d. process with mean 0 and variance $\sigma^2_\nu$. \texttt{xtregar} offers a within estimator for the fixed-effect model and the Baltagi–Wu (1999) GLS estimator of the random-effects model. The Baltagi–Wu (1999) GLS estimator extends the balanced panel estimator in Baltagi and Li (1991) to a case of exogenously unbalanced panels with unequally spaced observations. Both these estimators offer several estimators of $\rho$.

The data can be unbalanced and unequally spaced. Specifically, the dataset contains observations on individual $i$ at times $t_{ij}$ for $j = 1, \ldots, n_i$. The difference $t_{ij} - t_{i,j-1}$ plays an integral role in the estimation techniques used by \texttt{xtregar}. For this reason, you must \texttt{xtset} your data before using \texttt{xtregar}. For instance, if you have quarterly data, the “time” difference between the third and fourth quarter must be 1 month, not 3.

The fixed-effects model

Let’s examine the fixed-effect model first. The basic approach is common to all fixed-effects models. The $\nu_i$ are treated as nuisance parameters. We use a transformation of the model that removes the nuisance parameters and leaves behind the parameters of interest in an estimable form. Subtracting the group means from (1) removes the $\nu_i$ from the model.
\[ y_{itij} - \bar{y}_i = (\bar{x}_{itij} - \bar{x}_i) \beta + \epsilon_{itij} - \bar{\epsilon}_i \]  

(3)

where

\[ \bar{y}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{itij} \]
\[ \bar{x}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{itij} \]
\[ \bar{\epsilon}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} \epsilon_{itij} \]

After the transformation, (3) is a linear AR(1) model, potentially with unequally spaced observations. (3) can be used to estimate \( \rho \). Given an estimate of \( \rho \), we must do a Cochrane–Orcutt transformation on each panel and then remove the within-panel means and add back the overall mean for each variable. OLS on the transformed data will produce the within estimates of \( \alpha \) and \( \beta \).

Example 1: Fixed-effects model

Let's use the Grunfeld investment dataset to illustrate how `xtregar` can be used to fit the fixed-effects model. This dataset contains information on 10 firms' investment, market value, and the value of their capital stocks. The data were collected annually between 1935 and 1954. The following output shows that we have `xtset` our data and gives the results of running a fixed-effects model with investment as a function of market value and the capital stock.

```
use http://www.stata-press.com/data/r13/grunfeld
. xtset
    panel variable: company (strongly balanced)
    time variable: year, 1935 to 1954
    delta: 1 year
. xtregar invest mvalue kstock, fe
```

FE (within) regression with AR(1) disturbances
Number of obs = 190
Group variable: company
Number of groups = 10
R-sq: within = 0.5927
     between = 0.7989
     overall = 0.7904

F(2, 178) = 129.49
corr(u_i, Xb) = -0.0454
Prob > F = 0.0000

|             | Coef.    | Std. Err. | t     | P>|t|  | [95% Conf. Interval] |
|-------------|----------|-----------|-------|------|---------------------|
| invest      |          |           |       |      |                     |
| mvalue      | 0.0949999| 0.0091377 | 10.40 | 0.000| [0.0769677, 0.113032]|
| kstock      | 0.350161 | 0.029375  | 11.92 | 0.000| [0.2921935, 0.408129]|
| _cons       | -63.22022| 5.648271  | -11.19| 0.000| [-74.36641, -52.07402]|
| rho_ar      | 0.672106 | 0.00175   | 38.42 | 0.000| [0.652606, 0.691602]  |
| sigma_u     | 91.507609|           |       |      |                     |
| sigma_e     | 40.992469|           |       |      |                     |
| rho_fov     | 0.832864 | (fraction of variance because of u_i) |

F test that all u_i=0: F(9, 178) = 11.53
Prob > F = 0.0000

Because there are 10 groups, the panel-by-panel Cochrane–Orcutt method decreases the number of available observations from 200 to 190. The above example used the default `dw` estimator of \( \rho \). Using the `tscorr` estimator of \( \rho \) yields
Technical note

The `tscorr` estimator of $\rho$ is bounded in $[-1, 1]$. The other estimators of $\rho$ are not. In samples with short panels, the estimates of $\rho$ produced by the other estimators of $\rho$ may be outside $[-1, 1]$. If this happens, use the `tscorr` estimator. However, simulations have shown that the `tscorr` estimator is biased toward zero. `dw` is the default because it performs well in Monte Carlo simulations. In the example above, the estimate of $\rho$ produced by `tscorr` is much smaller than the one produced by `dw`.

Example 2: Using `xtset`

`xtregar` will complain if you try to run `xtregar` on a dataset that has not been `xtset`:

```
. xtset, clear
. xtregar invest mvalue kstock, fe
must specify panelvar and timevar; use `xtset r(459);`
```

You must `xtset` your data to ensure that `xtregar` understands the nature of your time variable. Suppose that our observations were taken quarterly instead of annually. We will get the same results with the quarterly variable `t2` that we did with the annual variable `year`.

```
. generate t = year - 1934
. generate t2 = tq(1934q4) + t
. format t2 %tq
```
In all the examples thus far, we have assumed that $\epsilon_{it}$ is first-order autoregressive. Testing the hypothesis of $\rho = 0$ in a first-order autoregressive process produces test statistics with extremely complicated distributions. Bhargava, Franzini, and Narendranathan (1982) extended the Durbin–Watson statistic to the case of balanced, equally spaced panel datasets. Baltagi and Wu (1999) modify their statistic to account for unbalanced panels with unequally spaced data. In the same article, Baltagi and Wu (1999) derive the locally best invariant test statistic of $\rho = 0$. Both these test statistics have extremely complicated distributions, although Bhargava, Franzini, and Narendranathan (1982) did publish some critical values in their article. Specifying the `lbi` option to `xtregar` causes Stata to calculate and report the modified Bhargava et al. Durbin–Watson and the Baltagi–Wu LBI.

**Example 3: Testing for autocorrelation**

In this example, we calculate the modified Bhargava et al. Durbin–Watson statistic and the Baltagi–Wu LBI. We exclude periods 9 and 10 from the sample, thereby reproducing the results of Baltagi and Wu (1999, 822). $p$-values are not reported for either statistic. Although Bhargava, Franzini, and Narendranathan (1982) published critical values for their statistic, no tables are currently available for the Baltagi–Wu (LBI). Baltagi and Wu (1999) did derive a normalized version of their statistic, but this statistic cannot be computed for datasets of moderate size.
. xtregar invest mvalue kstock if year !=1934 & year !=1944, fe lbi

FE (within) regression with AR(1) disturbances

|                  | Coef. | Std. Err. | t     | P>|t| | [95% Conf. Interval] |
|------------------|-------|-----------|-------|------|----------------------|
| invest           |       |           |       |      |                      |
| mvalue           | .0941122 | .0090926 | 10.35 | 0.000 | .0761617 -- .1120627 |
| kstock           | .3535872 | .0303562 | 11.65 | 0.000 | .2936584 -- .4135161 |
| _cons            | -64.82534 | 5.946885 | -10.90 | 0.000 | -76.56559 -- -53.08509 |
| rho_ar           | .6697198 |           |       |      |                      |
| sigma_u          | 93.320452 |          |       |      |                      |
| sigma_e          | 41.580712 |          |       |      |                      |
| rho_fov          | .83435413 |          |       |      |                      |

F test that all u_i=0: F(9,168) = 11.55 Prob > F = 0.0000
modified Bhargava et al. Durbin-Watson = .71380994
Baltagi-Wu LBI = 1.0134522

The random-effects model

In the random-effects model, the \( \nu_i \) are assumed to be realizations of an i.i.d. process with mean 0 and variance \( \sigma_{\nu}^2 \). Furthermore, the \( \nu_i \) are assumed to be independent of both the \( \epsilon_{it} \) and the covariates \( x_{it} \). The latter of these assumptions can be strong, but inference is not conditional on the particular realizations of the \( \nu_i \) in the sample. See Mundlak (1978) for a discussion of this point.

Example 4: Random-effects model

By specifying the re option, we obtain the Baltagi–Wu GLS estimator of the random-effects model. This estimator can accommodate unbalanced panels and unequally spaced data. We run this model on the Grunfeld dataset:
xtregar invest mvalue kstock if year != 1934 & year != 1944, re lbi

**RE GLS regression with AR(1) disturbances**

Number of obs = 190  
Number of groups = 10  
R-sq: within = 0.7707  
Obs per group: min = 19  
between = 0.8039  
avg = 19.0  
overall = 0.7958  
max = 19

Wald chi2(3) = 351.37  
corr(u_i, Xb) = 0 (assumed)  
Prob > chi2 = 0.0000

| invest   | Coef.  | Std. Err. | z      | P>|z|   | [95% Conf. Interval] |
|-----------|--------|-----------|--------|-------|---------------------|
| mvalue    | 0.0947714 | 0.0083691 | 11.32  | 0.000 | 0.0783683 - 0.1111746 |
| kstock    | 0.3223932 | 0.0263226 | 12.25  | 0.000 | 0.2708019 - 0.3739845 |
| _cons     | -45.21427 | 27.12492 | -1.67  | 0.096 | -98.37814 - 7.949603 |

rho_ar  .6697198 (estimated autocorrelation coefficient)
sigma_u  74.662876
sigma_e  42.253042
rho_fov  .75742494 (fraction of variance due to u_i)
theta    .66973313

modified Bhargava et al. Durbin–Watson = .71380994
Baltagi–Wu LBI = 1.0134522

The modified Bhargava et al. Durbin–Watson and the Baltagi–Wu LBI are the same as those reported for the fixed-effects model because the formulas for these statistics do not depend on fitting the fixed-effects model or the random-effects model.
Stored results

`xtregar, re` stores the following in `e()`:

Scalars

- `e(N)` number of observations
- `e(N_g)` number of groups
- `e(df_m)` model degrees of freedom
- `e(g_min)` smallest group size
- `e(g_avg)` average group size
- `e(g_max)` largest group size
- `e(d1)` Bhargava et al. Durbin–Watson
- `e(LBI)` Baltagi–Wu LBI statistic
- `e(N_LBI)` number of obs used in `e(LBI)`
- `e(Tcon)` 1 if $T$ is constant
- `e(sigma_u)` panel-level standard deviation
- `e(sigma_e)` standard deviation of $\eta_{it}$
- `e(r2_w)` $R$-squared for within model
- `e(r2_o)` $R$-squared for overall model
- `e(r2_b)` $R$-squared for between model
- `e(chi2)` $\chi^2$
- `e(rho_ar)` autocorrelation coefficient
- `e(rho_fov)` $u_i$ fraction of variance
- `e(theta_min)` minimum $\theta$
- `e(theta_5)` $\theta$, 5th percentile
- `e(theta_50)` $\theta$, 50th percentile
- `e(theta_95)` $\theta$, 95th percentile
- `e(theta_max)` maximum $\theta$
- `e(Tbar)` harmonic mean of group sizes
- `e(rank)` rank of $e(V)$

Macros

- `e(cmd)` `xtregar`
- `e(cmdline)` command as typed
- `e(depvar)` name of dependent variable
- `e(ivar)` variable denoting groups
- `e(tvar)` variable denoting time within groups
- `e(model)` re
- `e(rhotype)` method of estimating $\rho_{ar}$
- `e(dw)` LBI, if requested
- `e(chi2type)` Wald; type of model $\chi^2$ test
- `e(properties)` b V
- `e(predict)` program used to implement predict
- `e(asbalanced)` factor variables fvset as asbalanced
- `e(asobserved)` factor variables fvset as asobserved

Matrices

- `e(b)` coefficient vector
- `e(V)` VCE for random-effects model

Functions

- `e(sample)` marks estimation sample
**xtregar, fe** stores the following in **e()**:

**Scalars**

- **e(N)**: number of observations
- **e(N_g)**: number of groups
- **e(df_m)**: model degrees of freedom
- **e(mss)**: model sum of squares
- **e(rss)**: residual sum of squares
- **e(g_min)**: smallest group size
- **e(g_avg)**: average group size
- **e(g_max)**: largest group size
- **e(d1)**: Bhargava et al. Durbin–Watson
- **e(LBI)**: Baltagi–Wu LBI statistic
- **e(N_LBI)**: number of obs used in **e(LBI)**
- **e(Tcon)**: 1 if *T* is constant
- **e(corr)**: \( \text{corr}(u_i, Xb) \)
- **e(sigma_u)**: panel-level standard deviation
- **e(sigma_e)**: standard deviation of \( \epsilon_{it} \)
- **e(r2_a)**: adjusted *R*-squared
- **e(r2_w)**: *R*-squared for within model
- **e(r2_o)**: *R*-squared for overall model
- **e(r2_b)**: *R*-squared for between model
- **e(ll)**: log likelihood
- **e(ll_0)**: log likelihood, constant-only model
- **e(rho_ar)**: autocorrelation coefficient
- **e(rho_fov)**: \( u_i \) fraction of variance
- **e(F)**: *F* statistic
- **e(F_f)**: *F* for \( u_i = 0 \)
- **e(df_r)**: residual degrees of freedom
- **e(df_a)**: degrees of freedom for absorbed effect
- **e(df_b)**: numerator degrees of freedom for *F* statistic
- **e(rmse)**: root mean squared error
- **e(Tbar)**: harmonic mean of group sizes
- **e(rank)**: rank of **e(V)**

**Macros**

- **e(cmd)**: **xtregar**
- **e(cmdline)**: command as typed
- **e(depvar)**: name of dependent variable
- **e(ivar)**: variable denoting groups
- **e(tvar)**: variable denoting time within groups
- **e(wtype)**: weight type
- **e(wexp)**: weight expression
- **e(model)**: **fe**
- **e(rhotype)**: method of estimating \( \rho_{ar} \)
- **e(dw)**: LBI, if requested
- **e(properties)**: \( b \ V \)
- **e(predict)**: program used to implement **predict**
- **e(asbalanced)**: factor variables **fvset** as **asbalanced**
- **e(asobserved)**: factor variables **fvset** as **asobserved**

**Matrices**

- **e(b)**: coefficient vector
- **e(V)**: variance–covariance matrix of the estimators

**Functions**

- **e(sample)**: marks estimation sample
Methods and formulas

Consider a linear panel-data model described by (1) and (2). The data can be unbalanced and unequally spaced. Specifically, the dataset contains observations on individual \( i \) at times \( t_{ij} \) for \( j = 1, \ldots, n_i \).

Methods and formulas are presented under the following headings:

- Estimating \( \rho \)
- Transforming the data to remove the AR(1) component
- The within estimator of the fixed-effects model
- The Baltagi–Wu GLS estimator
- The test statistics

Estimating \( \rho \)

The estimate of \( \rho \) is always obtained after removing the group means. Let \( \tilde{y}_{it} = y_{it} - \overline{y}_i \), let \( \tilde{x}_{it} = x_{it} - \overline{x}_i \), and let \( \tilde{\epsilon}_{it} = \epsilon_{it} - \overline{\epsilon}_i \).

Then, except for the onestep method, all the estimates of \( \rho \) are obtained by running Stata’s prais on

\[
\tilde{y}_{it} = \tilde{x}_{it}\beta + \tilde{\epsilon}_{it}
\]

See [TS] prais for the formulas for each of the methods.

When onestep is specified, a regression is run on the above equation, and the residuals are obtained. Let \( e_{it_{ij}} \) be the residual used to estimate the error \( \tilde{\epsilon}_{it_{ij}} \). If \( t_{ij} - t_{i,j-1} > 1 \), \( e_{it_{ij}} \) is set to zero. Given this series of residuals

\[
\hat{\rho}_{onestep} = \frac{n}{m_c} \frac{\sum_{i=1}^N \sum_{t=2}^T e_{it} e_{i,t-1}}{\sum_{i=1}^N \sum_{t=1}^T e_{it}^2}
\]

where \( n \) is the number of nonzero elements in \( e \) and \( m_c \) is the number of consecutive pairs of nonzero \( e_{it} \)s.

Transforming the data to remove the AR(1) component

After estimating \( \rho \), Baltagi and Wu (1999) derive a transformation of the data that removes the AR(1) component. Their \( C_i(\rho) \) can be written as

\[
y_{it_{ij}}^* = \begin{cases} (1 - \rho^2)^{1/2} y_{it_{ij}} \\ (1 - \rho^2)^{1/2} \left[ y_{i,t_{ij}} \left\{ \frac{1}{1 - \rho^2(t_{ij} - t_{i,j-1})} \right\}^{1/2} - y_{i,t_{i,j-1}} \left\{ \frac{\rho^2(t_{ij} - t_{i,j-1})}{1 - \rho^2(t_{i,j} - t_{i,j-1})} \right\}^{1/2} \right] \end{cases} \quad \text{if } t_{ij} = 1 \quad \text{if } t_{ij} > 1
\]

Using the analogous transform on the independent variables generates transformed data without the AR(1) component. Performing simple OLS on the transformed data leaves behind the residuals \( \mu^* \).
The within estimator of the fixed-effects model

To obtain the within estimator, we must transform the data that come from the AR(1) transform. For the within transform to remove the fixed effects, the first observation of each panel must be dropped. Specifically, let

\[
\dot{y}_{itij} = y_{itij}^* - \bar{y}_t^* + \bar{y}^* \quad \forall j > 1
\]

\[
\dot{x}_{itij} = x_{itij}^* - \bar{x}_i^* + \bar{x}^* \quad \forall j > 1
\]

\[
\dot{\epsilon}_{itij} = \epsilon_{itij}^* - \bar{\epsilon}_i^* + \bar{\epsilon}^* \quad \forall j > 1
\]

where

\[
\bar{y}_i^* = \frac{\sum_{j=2}^{n_i-1} y_{itij}^*}{n_i - 1}
\]

\[
\bar{\bar{y}}^* = \frac{\sum_{i=1}^{N} \sum_{j=2}^{n_i-1} y_{itij}^*}{\sum_{i=1}^{N} n_i - 1}
\]

\[
\bar{x}_i^* = \frac{\sum_{j=2}^{n_i-1} x_{itij}^*}{n_i - 1}
\]

\[
\bar{x}^* = \frac{\sum_{i=1}^{N} \sum_{j=2}^{n_i-1} x_{itij}^*}{\sum_{i=1}^{N} n_i - 1}
\]

\[
\bar{\epsilon}_i^* = \frac{\sum_{j=2}^{n_i-1} \epsilon_{itij}^*}{n_i - 1}
\]

\[
\bar{\bar{\epsilon}}^* = \frac{\sum_{i=1}^{N} \sum_{j=2}^{n_i-1} \epsilon_{itij}^*}{\sum_{i=1}^{N} n_i - 1}
\]

The within estimator of the fixed-effects model is then obtained by running OLS on

\[
\dot{y}_{itij} = \alpha + \dot{x}_{itij} \beta + \dot{\epsilon}_{itij}
\]

Reported as $R^2$ within is the $R^2$ from the above regression.

Reported as $R^2$ between is $\left\{ \text{corr}(\bar{x}_i \hat{\beta}, \bar{y}_i) \right\}^2$.

Reported as $R^2$ overall is $\left\{ \text{corr}(x_{it} \hat{\beta}, y_{it}) \right\}^2$.

The Baltagi–Wu GLS estimator

The residuals $\mu^*$ can be used to estimate the variance components. Translating the matrix formulas given in Baltagi and Wu (1999) into summations yields the following variance-components estimators:
\[
\hat{\sigma}_\omega^2 = \sum_{i=1}^{N} \frac{(\mu'_i g_i)^2}{(g'_i g_i)}
\]
\[
\hat{\sigma}_\epsilon^2 = \frac{\left[ \sum_{i=1}^{N} (\mu'_i \mu' \mu_i) \right] - \sum_{i=1}^{N} \left\{ \frac{(\mu'_i g_i)^2}{(g'_i g_i)} \right\}}{\sum_{i=1}^{N} (n_i - 1)}
\]
\[
\hat{\sigma}_\mu^2 = \frac{\left[ \sum_{i=1}^{N} \left\{ \frac{(\mu'_i g_i)^2}{(g'_i g_i)} \right\} - N\hat{\sigma}_\epsilon^2 \right]}{\sum_{i=1}^{N} (g'_i g_i)}
\]

where

\[
g_i = \begin{bmatrix}
1, & \left\{ 1 - \rho(t_{i,2} - t_{i,1}) \right\} & , & \ldots, & \left\{ 1 - \rho(t_{i,n_i - t_{i,n_i - 1}}) \right\}
\end{bmatrix}'
\]

and \(\mu'_i\) is the \(n_i \times 1\) vector of residuals from \(\mu^*\) that correspond to person \(i\).

Then

\[
\hat{\theta}_i = 1 - \left( \frac{\hat{\sigma}_\mu}{\hat{\omega}_i} \right)
\]

where

\[
\hat{\omega}_i^2 = g'_i g_i \hat{\sigma}_\mu^2 + \hat{\sigma}_\epsilon^2
\]

With these estimates in hand, we can transform the data via

\[
z^{**}_{itij} = z^{*}_{itij} - \hat{\theta}_i g_{ij} \frac{\sum_{s=1}^{n_i} g_{is} z^{*}_{itis}}{\sum_{s=1}^{n_i} g_{is}^2}
\]

for \(z \in \{y, x\}\).

Running OLS on the transformed data \(y^{**}, x^{**}\) yields the feasible GLS estimator of \(\alpha\) and \(\beta\).

Reported as \(R^2\) between is \(\left\{ \text{corr}(\bar{x}_i \hat{\beta}, \bar{y}_i) \right\}^2\).

Reported as \(R^2\) within is \(\left\{ \text{corr}\left( (x_{it} - \bar{x}_i) \hat{\beta}, y_{it} - \bar{y}_i \right) \right\}^2\).

Reported as \(R^2\) overall is \(\left\{ \text{corr}(x_{it} \hat{\beta}, y_{it}) \right\}^2\).
The test statistics

The Baltagi–Wu LBI is the sum of terms

\[ d_* = d_1 + d_2 + d_3 + d_4 \]

where

\[ d_1 = \frac{\sum_{i=1}^{N} \sum_{j=1}^{n_i} \{ \tilde{z}_{it_i,j-1} - \tilde{z}_{it_ij} I(t_{ij} - t_{i,j-1} = 1) \}^2}{\sum_{i=1}^{N} \sum_{j=1}^{n_i} \tilde{z}_{it_ij}^2} \]

\[ d_2 = \frac{\sum_{i=1}^{N} \sum_{j=1}^{n_i-1} \tilde{z}_{it_i,j-1}^2 \{ 1 - I(t_{ij} - t_{i,j-1} = 1) \}^2}{\sum_{i=1}^{N} \sum_{j=1}^{n_i} \tilde{z}_{it_ij}^2} \]

\[ d_3 = \frac{\sum_{i=1}^{N} \tilde{z}_{it_i1}^2}{\sum_{i=1}^{N} \sum_{j=1}^{n_i} \tilde{z}_{it_ij}^2} \]

\[ d_4 = \frac{\sum_{i=1}^{N} \tilde{z}_{it_in_i}^2}{\sum_{i=1}^{N} \sum_{j=1}^{n_i} \tilde{z}_{it_ij}^2} \]

\( I() \) is the indicator function that takes the value of 1 if the condition is true and 0 otherwise. The \( \tilde{z}_{it_i,j-1} \) are residuals from the within estimator.

Baltagi and Wu (1999) also show that \( d_1 \) is the Bhargava et al. Durbin–Watson statistic modified to handle cases of unbalanced panels and unequally spaced data.

Acknowledgment

We thank Badi Baltagi of the Department of Economics at Syracuse University for his helpful comments.

References


Also see

[XT] **xtregar postestimation** — Postestimation tools for xtregar

[XT] **xtset** — Declare data to be panel data

[XT] **xtgee** — Fit population-averaged panel-data models by using GEE

[XT] **xtgls** — Fit panel-data models by using GLS

[XT] **xtreg** — Fixed-, between-, and random-effects and population-averaged linear models

[TS] **newey** — Regression with Newey–West standard errors

[TS] **prais** — Prais–Winsten and Cochrane–Orcutt regression

[U] **20 Estimation and postestimation commands**
Description

The following postestimation commands are available after `xtregar`:

<table>
<thead>
<tr>
<th>Command</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>contrast</code></td>
<td>contrasts and ANOVA-style joint tests of estimates</td>
</tr>
<tr>
<td><code>estat ic</code>¹</td>
<td>Akaike’s and Schwarz’s Bayesian information criteria (AIC and BIC)</td>
</tr>
<tr>
<td><code>estat summarize</code></td>
<td>summary statistics for the estimation sample</td>
</tr>
<tr>
<td><code>estat vce</code></td>
<td>variance–covariance matrix of the estimators (VCE)</td>
</tr>
<tr>
<td><code>estimates</code></td>
<td>cataloging estimation results</td>
</tr>
<tr>
<td><code>forecast</code></td>
<td>dynamic forecasts and simulations</td>
</tr>
<tr>
<td><code>hausman</code></td>
<td>Hausman’s specification test</td>
</tr>
<tr>
<td><code>lincom</code></td>
<td>point estimates, standard errors, testing, and inference for linear</td>
</tr>
<tr>
<td></td>
<td>combinations of coefficients</td>
</tr>
<tr>
<td><code>margins</code></td>
<td>marginal means, predictive margins, marginal effects, and average marginal</td>
</tr>
<tr>
<td></td>
<td>effects</td>
</tr>
<tr>
<td><code>marginsplot</code></td>
<td>graph the results from margins (profile plots, interaction plots, etc.)</td>
</tr>
<tr>
<td><code>nlcom</code></td>
<td>point estimates, standard errors, testing, and inference for nonlinear</td>
</tr>
<tr>
<td></td>
<td>combinations of coefficients</td>
</tr>
<tr>
<td><code>predict</code></td>
<td>predictions, residuals, influence statistics, and other diagnostic measures</td>
</tr>
<tr>
<td><code>predictnl</code></td>
<td>point estimates, standard errors, testing, and inference for generalized</td>
</tr>
<tr>
<td></td>
<td>predictions</td>
</tr>
<tr>
<td><code>pwcompare</code></td>
<td>pairwise comparisons of estimates</td>
</tr>
<tr>
<td><code>test</code></td>
<td>Wald tests of simple and composite linear hypotheses</td>
</tr>
<tr>
<td><code>testnl</code></td>
<td>Wald tests of nonlinear hypotheses</td>
</tr>
</tbody>
</table>

¹ `estat ic` is not appropriate after `xtregar, re`.

Syntax for predict

```
predict [type] newvar [if] [in] [, statistic]
```

<table>
<thead>
<tr>
<th>statistic</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>xb</code></td>
<td>( x_{it}\hat{b} ), linear prediction; the default</td>
</tr>
<tr>
<td><code>ue</code></td>
<td>( u_{it} + e_{it} ), the combined residual</td>
</tr>
<tr>
<td><code>*u</code></td>
<td>( u_{it} ), the fixed- or random-error component</td>
</tr>
<tr>
<td><code>*e</code></td>
<td>( e_{it} ), the overall error component</td>
</tr>
</tbody>
</table>

Unstarred statistics are available both in and out of sample; type `predict ... if e(sample) ...` if wanted only for the estimation sample. Starred statistics are calculated only for the estimation sample, even when `if e(sample)` is not specified.
Menu for predict

Statistics > Postestimation > Predictions, residuals, etc.

Options for predict

xb, the default, calculates the linear prediction, $x_{it}\beta$.
ue calculates the prediction of $u_i + e_{it}$.
u calculates the prediction of $u_i$, the estimated fixed or random effect.
e calculates the prediction of $e_{it}$.

Also see

[XT] xtregar — Fixed- and random-effects linear models with an AR(1) disturbance
[U] 20 Estimation and postestimation commands
xtset — Declare data to be panel data

Syntax

Declare data to be panel

    xtset panelvar
    xtset panelvar timevar [ , tsoptions ]

Display how data are currently xtset

    xtset

Clear xt settings

    xtset, clear

In the declare syntax, *panelvar* identifies the panels and the optional *timevar* identifies the times within panels. *tsoptions* concern *timevar*.

<table>
<thead>
<tr>
<th>tsoptions</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>unitoptions</td>
<td>specify units of <em>timevar</em></td>
</tr>
<tr>
<td>deltaoption</td>
<td>specify periodicity of <em>timevar</em></td>
</tr>
<tr>
<td>noquery</td>
<td>suppress summary calculations and output</td>
</tr>
</tbody>
</table>

*noquery* is not shown in the dialog box.

<table>
<thead>
<tr>
<th>unitoptions</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>(default)</td>
<td>*timevar’s units to be obtained from *timevar’s display format</td>
</tr>
<tr>
<td>clocktime</td>
<td>*timevar is %tc: 0 = 1jan1960 00:00:00.000, 1 = 1jan1960 00:00:00.001, ...</td>
</tr>
<tr>
<td>daily</td>
<td>*timevar is %td: 0 = 1jan1960, 1 = 2jan1960, ...</td>
</tr>
<tr>
<td>weekly</td>
<td>*timevar is %tw: 0 = 1960w1, 1 = 1960w2, ...</td>
</tr>
<tr>
<td>monthly</td>
<td>*timevar is %tm: 0 = 1960m1, 1 = 1960m2, ...</td>
</tr>
<tr>
<td>quarterly</td>
<td>*timevar is %tq: 0 = 1960q1, 1 = 1960q2, ...</td>
</tr>
<tr>
<td>halfyearly</td>
<td>*timevar is %th: 0 = 1960h1, 1 = 1960h2, ...</td>
</tr>
<tr>
<td>yearly</td>
<td>*timevar is %ty: 1960 = 1960, 1961 = 1961, ...</td>
</tr>
<tr>
<td>generic</td>
<td>*timevar is %tg: 0 = ?, 1 = ?, ...</td>
</tr>
<tr>
<td>format(%fmt)</td>
<td>specify *timevar’s format and then apply default rule</td>
</tr>
</tbody>
</table>

In all cases, negative *timevar* values are allowed.
**deltaoption** specifies the period between observations in *timevar* units and may be specified as:

<table>
<thead>
<tr>
<th>deltaoption</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>delta(#)</code></td>
<td><code>delta(1)</code> or <code>delta(2)</code></td>
</tr>
<tr>
<td><code>delta((exp))</code></td>
<td><code>delta((7*24))</code></td>
</tr>
<tr>
<td><code>delta(# units)</code></td>
<td><code>delta(7 days)</code> or <code>delta(15 minutes)</code> or <code>delta(7 days 15 minutes)</code></td>
</tr>
<tr>
<td><code>delta((exp) units)</code></td>
<td><code>delta((2+3) weeks)</code></td>
</tr>
</tbody>
</table>

Allowed units for %tc and %tC timevars are:
- **seconds**  secs  sec
- **minutes**  mins  min
- **hours**    hour
- **days**     day
- **weeks**    week

and for all other %t timevars are:
- **days**     day
- **weeks**    week

**Menu**

Statistics  >  Longitudinal/panel data  >  Setup and utilities  >  Declare dataset to be panel data

**Description**

`xtset` declares the data in memory to be a panel. You must `xtset` your data before you can use the other `xt` commands. If you save your data after `xtset`, the data will be remembered to be a panel and you will not have to `xtset` again.

There are two syntaxes for setting the data:

```
xset panelvar
xset panelvar timevar
```

In the first syntax—`xtset panelvar`—the data are set to be a panel and the order of the observations within panel is considered to be irrelevant. For instance, `panelvar` might be country and the observations within might be city.

In the second syntax—`xtset panelvar timevar`—the data are to be a panel and the order of observations within panel are considered ordered by *timevar*. For instance, in data collected from repeated surveying of the same people over various years, `panelvar` might be person and *timevar*, year. When you specify *timevar*, you may then use Stata’s time-series operators such as L. and F. (lag and lead) in other commands. The operators will be interpreted as lagged and lead values within panel.

`xtset` without arguments—`xtset`—displays how the data are currently `xtset`. If the data are set with a `panelvar` and a *timevar*, `xtset` also sorts the data by `panelvar` `timevar`. If the data are set with a `panelvar` only, the sort order is not changed.

`xtset`, clear is a rarely used programmer’s command to declare that the data are no longer to be considered a panel.
Options

unitoptions clocktime, daily, weekly, monthly, quarterly, halfyearly, yearly, generic, and format(%fmt) specify the units in which timevar is recorded, if timevar is specified.

timevar will often simply be a variable that counts 1, 2, ..., and is to be interpreted as first year of survey, second year, ..., or first month of treatment, second month, .... In these cases, you do not need to specify a unitoption.

In other cases, timevar will be a year variable or the like such as 2001, 2002, ..., and is to be interpreted as year of survey or the like. In those cases, you do not need to specify a unitoption.

In still other, more complicated cases, timevar will be a full-blown %t variable; see [D] datetime. If timevar already has a %t display format assigned to it, you do not need to specify a unitoption; xtset will obtain the units from the format. If you have not yet bothered to assign the appropriate %t format to the %t variable, however, you can use the unitoptions to tell xtset the units. Then xtset will set timevar's display format for you. Thus, the unitoptions are convenience options; they allow you to skip formatting the time variable. The following all have the same net result:

```
xtset pid t xtset pid t, daily xtset pid t, format(%td)
```

Understand that timevar is not required to be a %t variable; it can be any variable of your own concocting so long as it takes on integer values. When you xtset a time variable that is not %t, the display format does not change unless you specify the unitoption generic or use the format() option.

delta() specifies the periodicity of timevar and is commonly used when timevar is %tc. delta() is only sometimes used with the other %t formats or with generic time variables.

If delta() is not specified, delta(1) is assumed. This means that at timevar = 5, the previous time is timevar = 5 - 1 = 4 and the next time would be timevar = 5 + 1 = 6. Lag and lead operators, for instance, would work this way. This would be assumed regardless of the units of timevar.

If you specified delta(2), then at timevar = 5, the previous time would be timevar = 5 - 2 = 3 and the next time would be timevar = 5 + 2 = 7. Lag and lead operators would work this way. In the observation with timevar = 5, L.income would be the value of income in the observation for which timevar = 3 and F.income would be the value of income in the observation for which timevar = 7. If you then add an observation with timevar = 4, the operators will still work appropriately; that is, at timevar = 5, L.income will still have the value of income at timevar = 3.

There are two aspects of timevar: its units and its periodicity. The unitoptions set the units. delta() sets the periodicity. You are not required to specify one to specify the other. You might have a generic timevar but it counts in 12: 0, 12, 24, .... You would skip specifying unitoptions but would specify delta(12).

We mentioned that delta() is commonly used with %tc timevars because Stata’s %tc variables have units of milliseconds. If delta() is not specified and in some model you refer to L.bp, you will be referring to the value of bp 1 ms ago. Few people have data with periodicity of a millisecond. Perhaps your data are hourly. You could specify delta(3600000). Or you could specify delta((60*60*1000)), because delta() will allow expressions if you include an extra pair of parentheses. Or you could specify delta(1 hour). They all mean the same thing: timevar has periodicity of 3,600,000 ms. In an observation for which timevar = 1,489,572,000,000 (corresponding to 15mar2007 10:00:00), L.bp would be the observation for which timevar = 1,489,572,000,000 - 3,600,000 = 1,489,568,400,000 (corresponding to 15mar2007 9:00:00).
When you `xtset` the data and specify `delta()`, `xtset` verifies that all the observations follow the specified periodicity. For instance, if you specified `delta(2)`, then `timevar` could contain any subset of `{..., −4, −2, 0, 2, 4, ...}` or it could contain any subset of `{..., −3, −1, 1, 3, ...}`. If `timevar` contained a mix of values, `xtset` would issue an error message. The check is made on each panel independently, so one panel might contain `timevar` values from one set and the next, another, and that would be fine.

`clear`—used in `xtset`, `clear`—makes Stata forget that the data ever were `xtset`. This is a rarely used programmer’s option.

The following option is available with `xtset` but is not shown in the dialog box:

`noquery` prevents `xtset` from performing most of its summary calculations and suppresses output. With this option, only the following results are posted:

- `r(tdelta)`
- `r(panelvar)`
- `r(timevar)`
- `r(tsfmt)`
- `r(unit)`
- `r(unit1)`

### Remarks and examples

`xtset` declares the dataset in memory to be panel data. You need to do this before you can use the other `xt` commands. The storage types of both `panelvar` and `timevar` must be numeric, and both variables must contain integers only.

#### Technical note

In previous versions of Stata there was no `xtset` command. The other `xt` commands instead had the `i(panelvar)` and `t(timevar)` options. Older commands still have those options, but they are no longer documented and, if you specify them, they just perform the `xtset` for you. Thus, do-files that you previously wrote will continue to work. Modern usage, however, is to `xtset` the data first.

#### Technical note

`xtset` is related to the `tsset` command, which declares data to be time series. One of the syntaxes of `tsset` is `tsset panelvar timevar`, which is identical to one of `xtset`’s syntaxes, namely, `xtset panelvar timevar`. Here they are in fact the same command, meaning that `xtsetting` your data is sufficient to allow you to use the `ts` commands and `tssetting` your data is sufficient to allow you to use the `xt` commands. You do not need to set both, but it will not matter if you do.

`xtset` and `tsset` are different, however, when you set just a `panelvar`—you type `xtset panelvar`—or when you set just a `timevar`—you type `tsset timevar`.

#### Example 1: Panel data without a time variable

Many panel datasets contain a variable identifying panels but do not contain a time variable. For example, you may have a dataset where each panel is a family, and the observations within panel are family members, or you may have a dataset in which each person made a decision multiple times but the ordering of those decisions is unimportant and perhaps unknown. In this latter case, if the time
of the decision were known, we would advise you to `xtset` it. The other `xt` statistical commands
do not do something different because `timevar` has been set—they will ignore `timevar` if `timevar` is
irrelevant to the statistical method that you are using. You should always set everything that is true
about the data.

In any case, let’s consider the case where there is no `timevar`. We have data on U.S. states and
cities within states:

```
.list state city in 1/10, sepby(state)
```

<table>
<thead>
<tr>
<th>state</th>
<th>city</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alabama</td>
<td>Birmingham</td>
</tr>
<tr>
<td>Alabama</td>
<td>Mobile</td>
</tr>
<tr>
<td>Alabama</td>
<td>Montgomery</td>
</tr>
<tr>
<td>Alabama</td>
<td>Huntsville</td>
</tr>
<tr>
<td>Alaska</td>
<td>Anchorage</td>
</tr>
<tr>
<td>Alaska</td>
<td>Fairbanks</td>
</tr>
<tr>
<td>Arizona</td>
<td>Phoenix</td>
</tr>
<tr>
<td>Arizona</td>
<td>Tucson</td>
</tr>
<tr>
<td>Arkansas</td>
<td>Fayetteville</td>
</tr>
<tr>
<td>Arkansas</td>
<td>Fort Smith</td>
</tr>
</tbody>
</table>

Here we do not type `xtset state city` because `city` is not a time variable. Instead, we type `xtset state`:

```
.xtset state
   varlist:  state:  string variable not allowed
           r(109);
```

You cannot `xtset` a string variable. We must make a numeric variable from our string variable and
`xtset` that. One alternative is

```
. egen statenum = group(state)
. list state statenum in 1/10, sepby(state)
```

<table>
<thead>
<tr>
<th>state</th>
<th>statenum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alabama</td>
<td>1</td>
</tr>
<tr>
<td>Alabama</td>
<td>1</td>
</tr>
<tr>
<td>Alabama</td>
<td>1</td>
</tr>
<tr>
<td>Alabama</td>
<td>1</td>
</tr>
<tr>
<td>Alaska</td>
<td>2</td>
</tr>
<tr>
<td>Alaska</td>
<td>2</td>
</tr>
<tr>
<td>Arizona</td>
<td>3</td>
</tr>
<tr>
<td>Arizona</td>
<td>3</td>
</tr>
<tr>
<td>Arkansas</td>
<td>4</td>
</tr>
<tr>
<td>Arkansas</td>
<td>4</td>
</tr>
</tbody>
</table>

```
.xtset statenum
   panel variable:  statenum (unbalanced)
```
Perhaps a better alternative is

```stata
   . encode state, gen(st)
   . list state st in 1/10, sepby(state)
```

<table>
<thead>
<tr>
<th>state</th>
<th>st</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>1.</td>
<td></td>
</tr>
<tr>
<td>2.</td>
<td></td>
</tr>
<tr>
<td>3.</td>
<td></td>
</tr>
<tr>
<td>4.</td>
<td></td>
</tr>
<tr>
<td>5.</td>
<td></td>
</tr>
<tr>
<td>6.</td>
<td></td>
</tr>
<tr>
<td>7.</td>
<td></td>
</tr>
<tr>
<td>8.</td>
<td></td>
</tr>
<tr>
<td>9.</td>
<td></td>
</tr>
<tr>
<td>10.</td>
<td></td>
</tr>
</tbody>
</table>

```stata
state st
1. Alabama  Alabama
2. Alabama  Alabama
3. Alabama  Alabama
4. Alabama  Alabama
5. Alaska   Alaska
6. Alaska   Alaska
7. Arizona  Arizona
8. Arizona  Arizona
9. Arkansas Arkansas
10. Arkansas Arkansas
```

`encode` (see [D] `encode`) produces a numeric variable with a value label, so when we list the result, new variable `st` looks just like our original. It is, however, numeric:

```stata
   . list state st in 1/10, nolabel sepby(state)
```

<table>
<thead>
<tr>
<th>state</th>
<th>st</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>1.</td>
<td>1</td>
</tr>
<tr>
<td>2.</td>
<td>1</td>
</tr>
<tr>
<td>3.</td>
<td>1</td>
</tr>
<tr>
<td>4.</td>
<td>1</td>
</tr>
<tr>
<td>5.</td>
<td>2</td>
</tr>
<tr>
<td>6.</td>
<td>2</td>
</tr>
<tr>
<td>7.</td>
<td>3</td>
</tr>
<tr>
<td>8.</td>
<td>3</td>
</tr>
<tr>
<td>9.</td>
<td>4</td>
</tr>
<tr>
<td>10.</td>
<td>4</td>
</tr>
</tbody>
</table>

We can `xtset` new variable `st`:

```stata
   . xtset st
   panel variable: st (unbalanced)
```

> **Example 2: Panel data with a time variable**

Some panel datasets do contain a time variable. Dataset `abdata.dta` contains labor demand data from a panel of firms in the United Kingdom. Here are wage data for the first two firms in the dataset:
```
. use http://www.stata-press.com/data/r13/abdata, clear
. list id year wage if id==1 | id==2, sepby(id)

<table>
<thead>
<tr>
<th>id</th>
<th>year</th>
<th>wage</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1977</td>
<td>13.1516</td>
</tr>
<tr>
<td>2</td>
<td>1978</td>
<td>12.3018</td>
</tr>
<tr>
<td>3</td>
<td>1979</td>
<td>12.8395</td>
</tr>
<tr>
<td>4</td>
<td>1980</td>
<td>13.8039</td>
</tr>
<tr>
<td>5</td>
<td>1981</td>
<td>14.2897</td>
</tr>
<tr>
<td>6</td>
<td>1982</td>
<td>14.8681</td>
</tr>
<tr>
<td>7</td>
<td>1983</td>
<td>13.7784</td>
</tr>
<tr>
<td>8</td>
<td>1977</td>
<td>14.7909</td>
</tr>
<tr>
<td>9</td>
<td>1978</td>
<td>14.1036</td>
</tr>
<tr>
<td>10</td>
<td>1979</td>
<td>14.9534</td>
</tr>
<tr>
<td>11</td>
<td>1980</td>
<td>15.491</td>
</tr>
<tr>
<td>13</td>
<td>1982</td>
<td>16.1314</td>
</tr>
<tr>
<td>14</td>
<td>1983</td>
<td>16.3051</td>
</tr>
</tbody>
</table>
```

To declare this dataset as a panel dataset, you type

```
. xtset id year, yearly
panel variable: id (unbalanced)
time variable: year, 1976 to 1984
delta: 1 year
```

The output from `list` shows that the last observations for these two firms are for 1983, but `xtset` shows that for some firms data are available for 1984 as well. If one or more panels contain data for nonconsecutive periods, `xtset` will report that gaps exist in the time variable. For example, if we did not have data for firm 1 for 1980 but did have data for 1979 and 1981, `xtset` would indicate that our data have a gap.

For yearly data, we could omit the `yearly` option and just type `xtset id year` because years are stored and listed just like regular integers.

Having declared our data to be a panel dataset, we can use time-series operators to obtain lags:

```
. list id year wage L.wage if id==1 | id==2, sepby(id)
```

```
<table>
<thead>
<tr>
<th>id</th>
<th>year</th>
<th>wage</th>
<th>L.wage</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1977</td>
<td>13.1516</td>
<td>.</td>
</tr>
<tr>
<td>2</td>
<td>1978</td>
<td>12.3018</td>
<td>13.1516</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(output omitted)</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>1982</td>
<td>14.8681</td>
<td>14.2897</td>
</tr>
<tr>
<td>7</td>
<td>1983</td>
<td>13.7784</td>
<td>14.8681</td>
</tr>
<tr>
<td>8</td>
<td>1977</td>
<td>14.7909</td>
<td>.</td>
</tr>
<tr>
<td>9</td>
<td>1978</td>
<td>14.1036</td>
<td>14.7909</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(output omitted)</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>1983</td>
<td>16.3051</td>
<td>16.1314</td>
</tr>
</tbody>
</table>
```

`L.wage` is missing for 1977 in both panels because we have no wage data for 1976. In observation 8, the lag operator did not incorrectly reach back into the previous panel.
The terms **balanced** and **unbalanced** are often used to describe whether a panel dataset is missing some observations. If a dataset does not contain a time variable, then panels are considered **balanced** if each panel contains the same number of observations; otherwise, the panels are **unbalanced**.

When the dataset contains a time variable, panels are said to be **strongly balanced** if each panel contains the same time points, **weakly balanced** if each panel contains the same number of observations but not the same time points, and **unbalanced** otherwise.

#### Example 3: Applying time-series formats to the time variable

If our data are observed more than once per year, applying time-series formats to the time variable can improve readability.

We have a dataset consisting of individuals who joined a gym’s weight-loss program that began in January 2005 and ended in December 2005. Each participant’s weight was recorded once per month. Some participants did not show up for all the monthly weigh-ins, so we do not have all 12 months’ records for each person. The first two people’s data are:

```
use http://www.stata-press.com/data/r13/gymdata
list id month wt if id==1 | id==2, sepby(id)
```

<table>
<thead>
<tr>
<th>id</th>
<th>month</th>
<th>wt</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>145</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>144</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>144</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>143</td>
</tr>
<tr>
<td>11</td>
<td>1</td>
<td>124</td>
</tr>
<tr>
<td>12</td>
<td>1</td>
<td>120</td>
</tr>
<tr>
<td>13</td>
<td>1</td>
<td>123</td>
</tr>
<tr>
<td>14</td>
<td>2</td>
<td>143</td>
</tr>
<tr>
<td>23</td>
<td>1</td>
<td>122</td>
</tr>
<tr>
<td>24</td>
<td>2</td>
<td>118</td>
</tr>
</tbody>
</table>

To set these data, we can type

```
xtset id month
(panel variable: id (unbalanced)
  time variable: month, 1 to 12, but with gaps
delta: 1 unit)
```

The note “but with gaps” above is no cause for concern. It merely warns us that, within some panels, some time values are missing. We already knew that about our data—some participants did not show up for the monthly weigh-ins.

The rest of this example concerns making output more readable. Month numbers such as 1, 2, ..., 12 are perfectly readable here. In another dataset, where month numbers went to, say 127, they would not be so readable. In such cases, we can make a more readable date—2005m1, 2005m2, ...—by using Stata’s %t variables. For a discussion, see [D] datetime. We will go quickly here. One of the %t formats is %tm—monthly—and it says that 1 means 1960m1. Thus, we need to recode our month variable so that, rather than taking on values from 1 to 12, it takes on values from 540 to 551. Then we can put a %tm format on that variable. Working out 540–551 is subject to mistakes. Stata function tm(2005m1) tells us the %tm month corresponding to January of 2005, so we can type
. generate month2 = month + tm(2005m1) - 1
. format month2 %tm

New variable month2 will work just as well as the original month in an xtset, and even a little better, because output will be a little more readable:

. xtset id month2
    panel variable: id (unbalanced)
    time variable: month2, 2005m1 to 2005m12, but with gaps
    delta: 1 month

By the way, we could have omitted typing format month2 %tm and then, rather than typing xtset id month2, we would have typed xtset id month2, monthly. The monthly option specifies that the time variable is %tm. When we did not specify the option, xtset determined that it was monthly from the display format we had set.

Example 4: Clock times

We have data from a large hotel in Las Vegas that changes the reservation prices for its room reservations hourly. A piece of the data looks like

. list in 1/5

<table>
<thead>
<tr>
<th>roomtype</th>
<th>time</th>
<th>price</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>02.13.2007 08:00</td>
<td>140</td>
</tr>
<tr>
<td>2</td>
<td>02.13.2007 09:00</td>
<td>155</td>
</tr>
<tr>
<td>3</td>
<td>02.13.2007 10:00</td>
<td>160</td>
</tr>
<tr>
<td>4</td>
<td>02.13.2007 11:00</td>
<td>155</td>
</tr>
<tr>
<td>5</td>
<td>02.13.2007 12:00</td>
<td>160</td>
</tr>
</tbody>
</table>

The panel variable is roomtype and, although you cannot see it from the output above, it takes on 1, 2, ..., 20. Variable time is a string variable. The first step in making this dataset xt is to translate the string to a numeric variable:

. generate double t = clock(time, "MDY hm")
. list in 1/5

<table>
<thead>
<tr>
<th>roomtype</th>
<th>time</th>
<th>price</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>02.13.2007 08:00</td>
<td>140</td>
<td>1.487e+12</td>
</tr>
<tr>
<td>2</td>
<td>02.13.2007 09:00</td>
<td>155</td>
<td>1.487e+12</td>
</tr>
<tr>
<td>3</td>
<td>02.13.2007 10:00</td>
<td>160</td>
<td>1.487e+12</td>
</tr>
<tr>
<td>4</td>
<td>02.13.2007 11:00</td>
<td>155</td>
<td>1.487e+12</td>
</tr>
<tr>
<td>5</td>
<td>02.13.2007 12:00</td>
<td>160</td>
<td>1.487e+12</td>
</tr>
</tbody>
</table>

See [D] datetime translation for an explanation of what is going on here. clock() is the function that converts strings to datetime (%tc) values. We typed clock(time, "MDY hm") to convert string variable time, and we told clock() that the values in time were in the order month, day, year, hour, and minute. We stored new variable t as a double because time values are large and that is required to prevent rounding. Even so, the resulting values 1.487e+12 look rounded, but that is only because of the default display format for new variables. We can see the values better if we change the format:
. format t %20.0gc
. list in 1/5

<table>
<thead>
<tr>
<th>roomtype</th>
<th>time</th>
<th>price</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>02.13.2007 08:00</td>
<td>140</td>
<td>1,486,972,800,000</td>
</tr>
<tr>
<td>2</td>
<td>02.13.2007 09:00</td>
<td>155</td>
<td>1,486,976,400,000</td>
</tr>
<tr>
<td>3</td>
<td>02.13.2007 10:00</td>
<td>160</td>
<td>1,486,980,000,000</td>
</tr>
<tr>
<td>4</td>
<td>02.13.2007 11:00</td>
<td>155</td>
<td>1,486,983,600,000</td>
</tr>
<tr>
<td>5</td>
<td>02.13.2007 12:00</td>
<td>160</td>
<td>1,486,987,200,000</td>
</tr>
</tbody>
</table>

Even better, however, would be to change the format to %tc—Stata’s clock-time format:

. format t %tc
. list in 1/5

<table>
<thead>
<tr>
<th>roomtype</th>
<th>time</th>
<th>price</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>02.13.2007 08:00</td>
<td>140</td>
<td>13feb2007 08:00:00</td>
</tr>
<tr>
<td>2</td>
<td>02.13.2007 09:00</td>
<td>155</td>
<td>13feb2007 09:00:00</td>
</tr>
<tr>
<td>3</td>
<td>02.13.2007 10:00</td>
<td>160</td>
<td>13feb2007 10:00:00</td>
</tr>
<tr>
<td>4</td>
<td>02.13.2007 11:00</td>
<td>155</td>
<td>13feb2007 11:00:00</td>
</tr>
<tr>
<td>5</td>
<td>02.13.2007 12:00</td>
<td>160</td>
<td>13feb2007 12:00:00</td>
</tr>
</tbody>
</table>

We could now drop variable time. New variable t contains the same information as time and t is better because it is a Stata time variable, the most important property of which being that it is numeric rather than string. We can xtset it. Here, however, we also need to specify the periodicity with xtset’s delta() option. Stata’s time variables are numeric, but they record milliseconds since 01jan1960 00:00:00. By default, xtset uses delta(1), and that means the time-series operators would not work as we want them to work. For instance, L.price would look back only 1 ms (and find nothing). We want L.price to look back 1 hour (3,600,000 ms):

. xtset roomtype t, delta(1 hour)
   _panel variable: roomtype (strongly balanced)
   _time variable: t,
   13feb2007 08:00:00 to 31mar2007 18:00:00,
   but with gaps
   _delta: 1 hour

. list price l.price in 1/5

<table>
<thead>
<tr>
<th>t</th>
<th>price</th>
<th>L.price</th>
</tr>
</thead>
<tbody>
<tr>
<td>13feb2007 08:00:00</td>
<td>140</td>
<td>140</td>
</tr>
<tr>
<td>13feb2007 09:00:00</td>
<td>155</td>
<td>140</td>
</tr>
<tr>
<td>13feb2007 10:00:00</td>
<td>160</td>
<td>155</td>
</tr>
<tr>
<td>13feb2007 11:00:00</td>
<td>155</td>
<td>160</td>
</tr>
<tr>
<td>13feb2007 12:00:00</td>
<td>160</td>
<td>155</td>
</tr>
</tbody>
</table>
Example 5: Clock times must be double

In the previous example, it was of vital importance that when we generated the \%tc variable \textit{t},
\begin{verbatim}
  . generate double t = clock(time, "MDY hm")
\end{verbatim}
we generated it as a double. Let’s see what would have happened had we forgotten and just typed
\begin{verbatim}
  generate t = clock(time, "MDY hm")
\end{verbatim}. Let’s go back and start with the same original data:
\begin{verbatim}
  . list in 1/5
  +-----------------+----------+------+
  | roomtype | time       | price |
  +-----------------+----------+------+
  1.     1 02.13.2007 08:00 140
  2.     1 02.13.2007 09:00 155
  3.     1 02.13.2007 10:00 160
  4.     1 02.13.2007 11:00 155
  5.     1 02.13.2007 12:00 160
  +-----------------+----------+------+
\end{verbatim}

Remember, variable \textit{time} is a string variable, and we need to translate it to numeric. So we translate,
but this time we forget to make the new variable a double:
\begin{verbatim}
  . generate t = clock(time, "MDY hm")
  . list in 1/5
  +-----------------+----------+------+----------+
  | roomtype | time       | price | t        |
  +-----------------+----------+------+----------+
  1.     1 02.13.2007 08:00 140 1.49e+12
  2.     1 02.13.2007 09:00 155 1.49e+12
  3.     1 02.13.2007 10:00 160 1.49e+12
  4.     1 02.13.2007 11:00 155 1.49e+12
  5.     1 02.13.2007 12:00 160 1.49e+12
  +-----------------+----------+------+----------+
\end{verbatim}

We see the first difference—\textit{t} now lists as 1.49e+12 rather than 1.487e+12 as it did previously—but this is nothing that would catch our attention. We would not even know that the value is different. Let’s continue.

Next we put a \%20.0gc format on \textit{t} to better see the numerical values. In fact, that is not something we would usually do in an analysis. We did that in the example to emphasize to you that the \textit{t} values were really big numbers. We will repeat the exercise just to be complete, but in real analysis, we would not bother.
\begin{verbatim}
  . format t %20.0gc
  . list in 1/5
  +-----------------+----------+------+----------+
  | roomtype | time       | price | t        |
  +-----------------+----------+------+----------+
  1.     1 02.13.2007 08:00 140 1,486,972,780,544
  2.     1 02.13.2007 09:00 155 1,486,976,450,560
  3.     1 02.13.2007 10:00 160 1,486,979,989,504
  4.     1 02.13.2007 11:00 155 1,486,983,659,520
  5.     1 02.13.2007 12:00 160 1,486,987,198,464
  +-----------------+----------+------+----------+
\end{verbatim}

Okay, we see big numbers in \textit{t}. Let’s continue.

Next we put a \%tc format on \textit{t}, and that is something we would usually do, and you should always do. You should also list a bit of the data, as we did:
. format t %tc
. list in 1/5

<table>
<thead>
<tr>
<th>roomtype</th>
<th>time</th>
<th>price</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>02.13.2007 08:00</td>
<td>140</td>
<td>13feb2007 07:59:40</td>
</tr>
<tr>
<td>2</td>
<td>02.13.2007 09:00</td>
<td>155</td>
<td>13feb2007 09:00:50</td>
</tr>
<tr>
<td>3</td>
<td>02.13.2007 10:00</td>
<td>160</td>
<td>13feb2007 09:59:49</td>
</tr>
<tr>
<td>4</td>
<td>02.13.2007 11:00</td>
<td>155</td>
<td>13feb2007 11:00:59</td>
</tr>
<tr>
<td>5</td>
<td>02.13.2007 12:00</td>
<td>160</td>
<td>13feb2007 11:59:58</td>
</tr>
</tbody>
</table>

By now, you should see a problem: the translated datetime values are off by a second or two. That was caused by rounding. Dates and times should be the same, not approximately the same, and when you see a difference like this, you should say to yourself, “The translation is off a little. Why is that?” and then you should think, “Of course, rounding. I bet that I did not create t as a double.”

Let’s assume, however, that you do not do this. You instead plow ahead:

. xtset roomtype t, delta(1 hour)

| time values with periodicity less than delta() found |
| r(451); |

And that is what will happen when you forget to create t as a double. The rounding will cause uneven periodicity, and xtset will complain.

By the way, it is important only that clock times (%tc and %tC variables) be stored as doubles. The other date values %td, %tw, %tm, %tq, %th, and %ty are small enough that they can safely be stored as floats, although forgetting and storing them as doubles does no harm.

Technical note

Stata provides two clock-time formats, %tc and %tC. %tC provides a clock with leap seconds. Leap seconds are occasionally inserted to account for randomness of the earth’s rotation, which gradually slows. Unlike the extra day inserted in leap years, the timing of when leap seconds will be inserted cannot be foretold. The authorities in charge of such matters announce a leap second approximately 6 months before insertion. Leap seconds are inserted at the end of the day, and the leap second is called 23:59:60 (that is, 11:59:60 pm), which is then followed by the usual 00:00:00 (12:00:00 am). Most nonastronomers find these leap seconds vexing. The added seconds cause problems because of their lack of predictability—knowing how many seconds there will be between 01jan2012 and 01jan2013 is not possible—and because there are not necessarily 24 hours in a day. If you use a leap second–adjusted clock, most days have 24 hours, but a few have 24 hours and 1 second. You must look at a table to find out.

From a time-series analysis point of view, the nonconstant day causes the most problems. Let’s say that you have data on blood pressure for a set of patients, taken hourly at 1:00, 2:00, …, and that you have xtset your data with delta(1 hour). On most days, L24.bp would be blood pressure at the same time yesterday. If the previous day had a leap second, however, and your data were recorded using a leap second–adjusted clock, there would be no observation L24.bp because 86,400 seconds before the current reading does not correspond to an on-the-hour time; 86,401 seconds before the current reading corresponds to yesterday’s time. Thus, whenever possible, using Stata’s %tc encoding rather than %tC is better.

When times are recorded by computers using leap second–adjusted clocks, however, avoiding %tC is not possible. For performing most time-series analysis, the recommended procedure is to map the
\%tC values to \%tc and then \texttt{xtset} those. You must ask yourself whether the process you are studying is based on the clock—the nurse does something at 2 o’clock every day—or the true passage of time—the emitter spits out an electron every 86,400,000 ms.

When dealing with computer-recorded times, first find out whether the computer (and its time-recording software) use a leap second–adjusted clock. If it does, translate that to a \%tC value. Then use function \texttt{cofC()} to convert to a \%tc value and \texttt{xtset} that. If variable \texttt{T} contains the \%tC value,

\begin{verbatim}
. generate double t = cofC(T)
. format t %tc
. xtset panelvar t, delta(...)
\end{verbatim}

Function \texttt{cofC()} moves leap seconds forward: 23:59:60 becomes 00:00:00 of the next day.

\textbf{Stored results}

\texttt{xtset} stores the following in \texttt{r()}:

- **Scalars**
  - \texttt{r(imin)}: minimum panel ID
  - \texttt{r(imax)}: maximum panel ID
  - \texttt{r(tmin)}: minimum time
  - \texttt{r(tmax)}: maximum time
  - \texttt{r(tdelta)}: delta

- **Macros**
  - \texttt{r(panelvar)}: name of panel variable
  - \texttt{r(timevar)}: name of time variable
  - \texttt{r(tdeltas)}: formatted delta
  - \texttt{r(tmins)}: formatted minimum time
  - \texttt{r(tmaxs)}: formatted maximum time
  - \texttt{r(tsfmt)}: \%fmt of time variable
  - \texttt{r(unit)}: units of time variable: Clock, clock, daily, weekly, monthly, quarterly, halfyearly, yearly, or generic
  - \texttt{r(unit1)}: units of time variable: C, c, d, w, m, q, h, y, or ""
  - \texttt{r(balanced)}: unbalanced, weakly balanced, or strongly balanced; a set of panels are strongly balanced if they all have the same time values, otherwise balanced if same number of time values, otherwise unbalanced

\textbf{Also see}

- [XT] \texttt{xtdescribe} — Describe pattern of xt data
- [XT] \texttt{xtsum} — Summarize xt data
- [XT] \texttt{tsset} — Declare data to be time-series data
- [TS] \texttt{tsfill} — Fill in gaps in time variable
xtsum — Summarize xt data

Syntax Menu Description Remarks and examples
Stored results Also see

Syntax

xtsum [ varlist ] [ if ]

A panel variable must be specified; use xtset; see [XT] xtset.
varlist may contain time-series operators; see [U] 11.4.4 Time-series varlists.
by is allowed; see [D] by.

Menu

Statistics > Longitudinal/panel data > Setup and utilities > Summarize xt data

Description

xtsum, a generalization of summarize (see [R] summarize), reports means and standard deviations for panel data; it differs from summarize in that it decomposes the standard deviation into between and within components.

Remarks and examples

If you have not read [XT] xt, please do so.

xtsum provides an alternative to summarize. For instance, in the nlswork dataset described in [XT] xt, hours contains the usual hours worked:

```
. use http://www.stata-press.com/data/r13/nlswork
(National Longitudinal Survey, Young Women 14-26 years of age in 1968)
. summarize hours
```

```
Variable       Obs  Mean    Std. Dev.     Min    Max
-------------  ----  ------  -----------  ------  ------
hours         28467  36.56   9.87       1.00    168.00
```

```
. xtsum hours
```

```
Variable       Mean    Std. Dev.     Min    Max    Observations
-------------  ------  -----------  ------  ------  ---------------
hours overall  36.56   9.87       1.00    168.00    N =  28467
between        7.85    83.5      1.00    168.00    n =  4710
within         7.53   -2.15     130.06    2.61     T-bar =  6.04
```

xtsum provides the same information as summarize and more. It decomposes the variable $x_{it}$ into a between ($\bar{x}_i$) and within ($x_{it} - \bar{x}_i + \bar{x}$, the global mean $\bar{x}$ being added back in make results comparable). The overall and within are calculated over 28,467 person-years of data. The between is calculated over 4,710 persons, and the average number of years a person was observed in the hours data is 6.
xtsum also reports minimums and maximums. Hours worked last week varied between 1 and (unbelievably) 168. Average hours worked last week for each woman varied between 1 and 83.5. “Hours worked within” varied between $-2.15$ and 130.1, which is not to say that any woman actually worked negative hours. The within number refers to the deviation from each individual’s average, and naturally, some of those deviations must be negative. Then the negative value is not disturbing but the positive value is. Did some woman really deviate from her average by $+130.1$ hours? No. In our definition of within, we add back in the global average of 36.6 hours. Some woman did deviate from her average by $130.1 - 36.6 = 93.5$ hours, which is still large.

The reported standard deviations tell us something that may surprise you. They say that the variation in hours worked last week across women is nearly equal to that observed within a woman over time. That is, if you were to draw two women randomly from our data, the difference in hours worked is expected to be nearly equal to the difference for the same woman in two randomly selected years.

If a variable does not vary over time, its within standard deviation will be zero:

```
. xtsum birth_yr
```

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
<th>Observations</th>
</tr>
</thead>
<tbody>
<tr>
<td>birth_yr overall</td>
<td>48.08509</td>
<td>3.012837</td>
<td>41</td>
<td>54</td>
<td>N = 28534</td>
</tr>
<tr>
<td>between</td>
<td>3.051795</td>
<td>41</td>
<td>54</td>
<td></td>
<td>n = 4711</td>
</tr>
<tr>
<td>within</td>
<td>0</td>
<td>48.08509</td>
<td>48.08509</td>
<td>T-bar = 6.05689</td>
<td></td>
</tr>
</tbody>
</table>

Stored results

_xtsum_ stores the following in _r()_:

Scalars

- _r(N)_ number of observations
- _r(n)_ number of panels
- _r(Tbar)_ average number of years under observation
- _r(mean)_ mean
- _r(sd)_ overall standard deviation
- _r(min)_ overall minimum
- _r(max)_ overall maximum
- _r(sd_b)_ between standard deviation
- _r(min_b)_ between minimum
- _r(max_b)_ between maximum
- _r(sd_w)_ within standard deviation
- _r(min_w)_ within minimum
- _r(max_w)_ within maximum

Also see

[XT] _xtdescribe_ — Describe pattern of xt data

[XT] _xttab_ — Tabulate xt data
xttab — Tabulate xt data

Syntax

xttab varname [if]  
xttrans varname [if] [, freq]

A panel variable must be specified; use xtset; see [XT] xtset.
by is allowed with xttab and xttrans; see [D] by.

Menu

xttab
Statistics > Longitudinal/panel data > Setup and utilities > Tabulate xt data

xttrans
Statistics > Longitudinal/panel data > Setup and utilities > Report transition probabilities

Description

xttab, a generalization of tabulate (see [R] tabulate oneway), performs one-way tabulations and decomposes counts into between and within components in panel data.

xttrans, another generalization of tabulate (see [R] tabulate oneway), reports transition probabilities (the change in one categorical variable over time).

Option

freq, allowed with xttrans only, specifies that frequencies as well as transition probabilities be displayed.

Remarks and examples

If you have not read [XT] xt, please do so.

> Example 1: xttab

Using the nlswork dataset described in [XT] xt, variable msp is 1 if a woman is married and her spouse resides with her, and 0 otherwise:
. use http://www.stata-press.com/data/r13/nlswork
(National Longitudinal Survey. Young Women 14-26 years of age in 1968)

. xttab msp

<table>
<thead>
<tr>
<th></th>
<th>Overall</th>
<th></th>
<th>Between</th>
<th></th>
<th>Within</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Freq.</td>
<td>Percent</td>
<td>Freq.</td>
<td>Percent</td>
<td>Percent</td>
</tr>
<tr>
<td>0</td>
<td>11324</td>
<td>39.71</td>
<td>3113</td>
<td>66.08</td>
<td>62.69</td>
</tr>
<tr>
<td>1</td>
<td>17194</td>
<td>60.29</td>
<td>3643</td>
<td>77.33</td>
<td>75.75</td>
</tr>
<tr>
<td>Total</td>
<td>28518</td>
<td>100.00</td>
<td>6756</td>
<td>143.41</td>
<td>69.73</td>
</tr>
</tbody>
</table>

(n = 4711)

The overall part of the table summarizes results in terms of person-years. We have 11,324 person-years of data in which msp is 0 and 17,194 in which it is 1—in 60.3% of our data, the woman is married with her spouse present. Between repeats the breakdown, but this time in terms of women rather than person-years; 3,113 of our women ever had msp 0 and 3,643 ever had msp 1, for a grand total of 6,756 ever having either. We have in our data, however, only 4,711 women. This means that there are women who sometimes have msp 0 and at other times have msp 1.

The within percent tells us the fraction of the time a woman has the specified value of msp. If we take the first line, conditional on a woman ever having msp 0, 62.7% of her observations have msp 0. Similarly, conditional on a woman ever having msp 1, 75.8% of her observations have msp 1. These two numbers are a measure of the stability of the msp values, and, in fact, msp 1 is more stable among these younger women than msp 0, meaning that they tend to marry more than they divorce. The total within of 69.75% is the normalized between weighted average of the within percents, that is, $(3113 \times 62.69 + 3643 \times 75.75)/6756$. It is a measure of the overall stability of the msp variable.

A time-invariant variable will have a tabulation with within percents of 100:

. xttab race

<table>
<thead>
<tr>
<th></th>
<th>Overall</th>
<th></th>
<th>Between</th>
<th></th>
<th>Within</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Freq.</td>
<td>Percent</td>
<td>Freq.</td>
<td>Percent</td>
<td>Percent</td>
</tr>
<tr>
<td>white</td>
<td>20180</td>
<td>70.72</td>
<td>3329</td>
<td>70.66</td>
<td>100.00</td>
</tr>
<tr>
<td>black</td>
<td>8051</td>
<td>28.22</td>
<td>1325</td>
<td>28.13</td>
<td>100.00</td>
</tr>
<tr>
<td>other</td>
<td>303</td>
<td>1.06</td>
<td>57</td>
<td>1.21</td>
<td>100.00</td>
</tr>
<tr>
<td>Total</td>
<td>28534</td>
<td>100.00</td>
<td>4711</td>
<td>100.00</td>
<td>100.00</td>
</tr>
</tbody>
</table>

(n = 4711)

Example 2: xtttrans

xttrans shows the transition probabilities. In cross-sectional time-series data, we can estimate the probability that $x_{i,t+1} = v_2$ given that $x_{it} = v_1$ by counting transitions. For instance

. xtttrans msp

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 if</td>
<td>1 if</td>
<td></td>
<td></td>
</tr>
<tr>
<td>married,</td>
<td>married,</td>
<td></td>
<td></td>
</tr>
<tr>
<td>spouse</td>
<td>spouse</td>
<td></td>
<td></td>
</tr>
<tr>
<td>present</td>
<td>present</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>80.49</td>
<td>19.51</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>7.96</td>
<td>92.04</td>
</tr>
<tr>
<td>Total</td>
<td>37.11</td>
<td>62.89</td>
<td>100.00</td>
</tr>
</tbody>
</table>
The rows reflect the initial values, and the columns reflect the final values. Each year, some 80% of the msp 0 persons in the data remained msp 0 in the next year; the remaining 20% became msp 1. Although msp 0 had a 20% chance of becoming msp 1 in each year, the msp 1 had only an 8% chance of becoming (or returning to) msp 0. The freq option displays the frequencies that go into the calculation:

```
. xttrans msp, freq
1 if married, spouse present
1 if married, spouse present
0 1 Total
```

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>7,697</td>
<td>1,866</td>
<td>9,563</td>
</tr>
<tr>
<td></td>
<td>80.49</td>
<td>19.51</td>
<td>100.00</td>
</tr>
<tr>
<td>1</td>
<td>1,133</td>
<td>13,100</td>
<td>14,233</td>
</tr>
<tr>
<td></td>
<td>7.96</td>
<td>92.04</td>
<td>100.00</td>
</tr>
<tr>
<td>Total</td>
<td>8,830</td>
<td>14,966</td>
<td>23,796</td>
</tr>
<tr>
<td></td>
<td>37.11</td>
<td>62.89</td>
<td>100.00</td>
</tr>
</tbody>
</table>

Technical note

The transition probabilities reported by xttrans are not necessarily the transition probabilities in a Markov sense. xttrans counts transitions from each observation to the next once the observations have been put in t order within i. It does not normalize for missing periods. xttrans does pay attention to missing values of the variable being tabulated, however, and does not count transitions from nonmissing to missing or from missing to nonmissing. Thus if the data are fully rectangularized, xttrans produces (inefficient) estimates of the Markov transition matrix. fillin will rectangularize datasets; see [D] fillin. Thus the Markov transition matrix could be estimated by typing

```
. fillin idcode year
. xttrans msp
(output omitted)
```

Stored results

xttab stores the following in r():

Scalars

- r(n) number of panels

Matrices

- r(results) results matrix

Also see

- [XT] xtdescribe — Describe pattern of xt data
- [XT] xtsum — Summarize xt data
### Syntax

```
xttobit depvar [indepvars] [if] [in] [weight] [, options]
```

#### options

<table>
<thead>
<tr>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>suppress constant term</td>
</tr>
<tr>
<td>`ll(varname</td>
</tr>
<tr>
<td>`ul(varname</td>
</tr>
<tr>
<td>include <code>varname</code> in model with coefficient constrained to 1</td>
</tr>
<tr>
<td>apply specified linear constraints</td>
</tr>
<tr>
<td>keep collinear variables</td>
</tr>
<tr>
<td><code>vcetype</code> may be <code>oim</code>, <code>bootstrap</code>, or <code>jackknife</code></td>
</tr>
<tr>
<td>set confidence level; default is <code>level(95)</code></td>
</tr>
<tr>
<td>perform likelihood-ratio test comparing against pooled tobit model</td>
</tr>
<tr>
<td>perform overall model test as a likelihood-ratio test</td>
</tr>
<tr>
<td>do not display constraints</td>
</tr>
<tr>
<td>control column formats, row spacing, line width, display of omitted variables and base and empty cells, and factor-variable labeling</td>
</tr>
<tr>
<td><code>intmethod(intmethod)</code> integration method; <code>intmethod</code> may be <code>mvaghermite</code> (the default) or <code>ghermite</code></td>
</tr>
<tr>
<td>use <code>#</code> quadrature points; default is <code>intpoints(12)</code></td>
</tr>
<tr>
<td>control the maximization process; seldom used</td>
</tr>
<tr>
<td>display legend instead of statistics</td>
</tr>
</tbody>
</table>

A panel variable must be specified; use `xtset`; see [XT] `xtset`. `indepvars` may contain factor variables; see [U] 11.4.3 Factor variables. `depvar` and `indepvars` may contain time-series operators; see [U] 11.4.4 Time-series varlists. `by`, `fp`, and `statsby` are allowed; see [U] 11.1.10 Prefix commands. `iweights` are allowed; see [U] 11.1.6 weight. Weights must be constant within panel. `coeflegend` does not appear in the dialog box. See [U] 20 Estimation and postestimation commands for more capabilities of estimation commands.
xttobit — Random-effects tobit models

Menu

Statistics > Longitudinal/panel data > Censored outcomes > Tobit regression (RE)

Description

xttobit fits random-effects tobit models. There is no command for a parametric conditional fixed-effects model, as there does not exist a sufficient statistic allowing the fixed effects to be conditioned out of the likelihood. Honoré (1992) has developed a semiparametric estimator for fixed-effect tobit models. Unconditional fixed-effects tobit models may be fit with the tobit command with indicator variables for the panels; the indicators can be created with the factor-variable syntax described in [U] 11.4.3 Factor variables. However, unconditional fixed-effects estimates are biased.

Options

Model

- noconstant; see [R] estimation options.

- ll(varname|#) and ul(varname|#) indicate the censoring points. You may specify one or both. ll() indicates the lower limit for left-censoring. Observations with depvar ≤ ll() are left-censored, observations with depvar ≥ ul() are right-censored, and remaining observations are not censored.

- offset(varname), constraints(constraints), collinear; see [R] estimation options.

SE

- vce(vcetype) specifies the type of standard error reported, which includes types that are derived from asymptotic theory (oim) and that use bootstrap or jackknife methods (bootstrap, jackknife); see [XT] vce_options.

Reporting

- level(#); see [R] estimation options.

- tobit specifies that a likelihood-ratio test comparing the random-effects model with the pooled (tobit) model be included in the output.

- noskip; see [R] estimation options.

- nocnsreport; see [R] estimation options.

- display_options: noomitted, vsquish, noemptycells, baselevels, allbaselevels, nofvlabel, fvwrap(#), fvwrapon(style), cformat(%fmt), pformat(%fmt), sformat(%fmt), and nolstretch; see [R] estimation options.

Integration

- intmethod(intmethod), intpoints(#); see [R] estimation options.

Maximization

- maximize_options: difficult, technique(algorithm_spec), iterate(#), [no]log, trace, gradient, showstep, hessian, showtolerance, tolerance(#), ltolerance(#), nrtolerance(#), nonrtolerance, and from(init_specs); see [R] maximize. These options are seldom used.

The following option is available with xttobit but is not shown in the dialog box:

- coeflegend; see [R] estimation options.
Remarks and examples

Consider the linear regression model with panel-level random effects

\[ y_{it} = x_{it} \beta + \nu_i + \epsilon_{it} \]

for \( i = 1, \ldots, n \) panels, where \( t = 1, \ldots, n_i \). The random effects, \( \nu_i \), are i.i.d., \( N(0, \sigma^2_\nu) \), and \( \epsilon_{it} \) are i.i.d. \( N(0, \sigma^2_\epsilon) \) independently of \( \nu_i \).

The observed data, \( y^o_{it} \), represent possibly censored versions of \( y_{it} \). If they are left-censored, all that is known is that \( y_{it} \leq y^o_{it} \). If they are right-censored, all that is known is that \( y_{it} \geq y^o_{it} \). If they are uncensored, \( y_{it} = y^o_{it} \). If they are right-censored, \( y^o_{it} \) is determined by \( \text{ul()} \). If they are uncensored, \( y^o_{it} \) is determined by \( \text{depvar} \).

Example 1

Using the \texttt{nlswork} data described in \cite{xt} \texttt{xt}, we fit a random-effects tobit model of adjusted (log) wages. We use the \texttt{ul()} option to impose an upper limit on the recorded log of wages. We use the \texttt{intpoints(25)} option to increase the number of integration points to 25 from 12, which aids convergence of this model.

```
. use http://www.stata-press.com/data/r13/nlswork3
(National Longitudinal Survey. Young Women 14-26 years of age in 1968)
. xttobit ln_wage union age grade not_smsa south##c.year, ul(1.9)
> intpoints(25) tobit
(output omitted)
```

Random-effects tobit regression Number of obs = 19224
Group variable: idcode Number of groups = 4148
Random effects \( u_i \sim \text{Gaussian} \) Obs per group: min = 1
Integration method: mvaghermite avg = 4.6
Integration points = 25
max = 12
Log likelihood = -6814.4638 Wald chi2(7) = 2924.91
Prob > chi2 = 0.0000

```

| ln_wage | Coef. | Std. Err. | z     | P>|z| | [95% Conf. Interval] |
|---------|-------|-----------|-------|------|----------------------|
| union   | .1430525 | .0069719 | 20.52 | 0.000 | .1293878              | .1567172 |
| age     | .009913  | .0017517 | 5.66  | 0.000 | .0064797              | .0133463 |
| grade   | .0784843 | .0022767 | 34.47 | 0.000 | .074022               | .0829466 |
| not_smsa| -.1339973| .0092061 | -14.56| 0.000 | -.1520409             | -.1159536 |
| 1.south | -.3507181| .0695557 | -5.04 | 0.000 | -.4870447             | -.2143915 |
| year    | -.0008283| .0018372 | -0.45 | 0.652 | -.0044291             | .0027725 |
| south#c.year | .0031938 | .0008606 | 3.71  | 0.000 | .0015071              | .0048805 |
| _cons   | .5101968 | .1006681 | 5.07  | 0.000 | .312891               | .7075025 |
| /sigma_u| .3045995 | .0048346 | 63.00 | 0.000 | .2951239              | .3140755 |
| /sigma_e| .2488682 | .0018254 | 136.34| 0.000 | .2452904              | .2524459 |
| rho     | .599684  | .0084097 | .5831174 | 0.6160733 |

Likelihood-ratio test of sigma_u=0: chibar2(01) = 6650.63 Prob>=chibar2 = 0.000
Observation summary: 0 left-censored observations
12334 uncensored observations
6890 right-censored observations
The output includes the overall and panel-level variance components (labeled $\text{sigma}_e$ and $\text{sigma}_u$, respectively) together with $\rho$ (labeled rho)

$$
\rho = \frac{\sigma^2_u}{\sigma^2_e + \sigma^2_u}
$$

which is the percent contribution to the total variance of the panel-level variance component.

When $\rho$ is zero, the panel-level variance component is unimportant, and the panel estimator is not different from the pooled estimator. A likelihood-ratio test of this is included at the bottom of the output. This test formally compares the pooled estimator (tobit) with the panel estimator.

Technical note

The random-effects model is calculated using quadrature, which is an approximation whose accuracy depends partially on the number of integration points used. We can use the quadchk command to see if changing the number of integration points affects the results. If the results change, the quadrature approximation is not accurate given the number of integration points. Try increasing the number of integration points using the intpoints() option and run quadchk again. Do not attempt to interpret the results of estimates when the coefficients reported by quadchk differ substantially. See [XT] quadchk for details and [XT] xtprobit for an example.

Because the xttobit likelihood function is calculated by Gauss–Hermite quadrature, on large problems the computations can be slow. Computation time is roughly proportional to the number of points used for the quadrature.
xttobit stores the following in **e()**: 

**Scalars**

- `e(N)` number of observations
- `e(N_g)` number of groups
- `e(N_unc)` number of uncensored observations
- `e(N_l)` number of left-censored observations
- `e(N_rc)` number of right-censored observations
- `e(N_cd)` number of completely determined observations
- `e(k)` number of parameters
- `e(k_eq)` number of equations in `e(b)`
- `e(k_eq_model)` number of equations in overall model test
- `e(k_dv)` number of dependent variables
- `e(ll)` log likelihood
- `e(ll_0)` log likelihood, constant-only model
- `e(chi2)` $\chi^2$
- `e(chi2_c)` $\chi^2$ for comparison test
- `e(rho)` $\rho$
- `e(sigma_u)` panel-level standard deviation
- `e(sigma_e)` standard deviation of $\epsilon_{it}$
- `e(n_quad)` number of quadrature points
- `e(g_min)` smallest group size
- `e(g_avg)` average group size
- `e(g_max)` largest group size
- `e(p)` significance
- `e(rank)` rank of `e(V)`
- `e(rank0)` rank of `e(V)` for constant-only model
- `e(ic)` number of iterations
- `e(rc)` return code
- `e(converged)` 1 if converged, 0 otherwise

**Macros**

- `e(cmd)` xttobit
- `e(cmdline)` command as typed
- `e(depvar)` names of dependent variables
- `e(ivar)` variable denoting groups
- `e(llopt)` contents of `ll()`, if specified
- `e(ulopt)` contents of `ul()`, if specified
- `e(k_aux)` number of auxiliary parameters
- `e(wtype)` weight type
- `e(wexp)` weight expression
- `e(title)` title in estimation output
- `e(offset)` offset
- `e(chi2type)` Wald or LR; type of model $\chi^2$ test
- `e(chi2_ct)` Wald or LR; type of model $\chi^2$ test corresponding to `e(chi2_c)
- `e(vce)` vcetype specified in `vce()`
- `e(vcetype)` title used to label Std. Err.
- `e(intmethod)` maximization technique
- `e(distrib)` Gaussian; the distribution of the random effect
- `e(opt)` type of optimization
- `e(which)` `max` or `min`; whether optimizer is to perform maximization or minimization
- `e(ml_method)` type of `ml` method
- `e(user)` name of likelihood-evaluator program
- `e(technique)` maximization technique
- `e(properties)` `b V`
- `e(predict)` program used to implement `predict`
- `e(asbalanced)` factor variables `fvset` as `asbalanced`
- `e(asobserved)` factor variables `fvset` as `asobserved`
Methods and formulas

Assuming a normal distribution, $N(0, \sigma^2_i)$, for the random effects $\nu_i$, we have the joint (unconditional of $\nu_i$) density of the observed data from the $i$th panel

$$f(y_{i1}, \ldots, y_{in_i} | x_{i1}, \ldots, x_{in_i}) = \int_{-\infty}^{\infty} \frac{e^{-\nu_i^2 / 2\sigma^2_i}}{\sqrt{2\pi\sigma_i}} \prod_{t=1}^{n_i} F(y_{it}, x_{it}\beta + \nu_i) \, d\nu_i$$

where

$$F(y_{it}, \Delta_{it}) = \begin{cases} (\sqrt{2\pi\sigma_i})^{-1} e^{-(y_{it} - \Delta_{it})^2 / (2\sigma^2_i)} & \text{if } y_{it}^0 \in C \\ \Phi\left(\frac{y_{it}^0 - \Delta_{it}}{\sigma_i}\right) & \text{if } y_{it}^0 \in L \\ 1 - \Phi\left(\frac{y_{it}^0 - \Delta_{it}}{\sigma_i}\right) & \text{if } y_{it}^0 \in R \end{cases}$$

where $C$ is the set of noncensored observations, $L$ is the set of left-censored observations, $R$ is the set of right-censored observations, and $\Phi()$ is the cumulative normal distribution.

The panel level likelihood $l_i$ is given by

$$l_i = \int_{-\infty}^{\infty} \frac{e^{-\nu_i^2 / 2\sigma^2_i}}{\sqrt{2\pi\sigma_i}} \prod_{t=1}^{n_i} F(y_{it}^0, x_{it}\beta + \nu_i) \, d\nu_i$$

$$\equiv \int_{-\infty}^{\infty} g(y_{it}^0, x_{it}, \nu_i) d\nu_i$$

This integral can be approximated with $M$-point Gauss–Hermite quadrature

$$\int_{-\infty}^{\infty} e^{-x^2} h(x) dx \approx \sum_{m=1}^{M} w_m^* h(a_m^*)$$

This is equivalent to

$$\int_{-\infty}^{\infty} f(x) dx \approx \sum_{m=1}^{M} w_m^* \exp\left\{ (a_m^*)^2 \right\} f(a_m^*)$$

where the $w_m^*$ denote the quadrature weights and the $a_m^*$ denote the quadrature abscissas. The log likelihood, $L$, is the sum of the logs of the panel level likelihoods $l_i$. 

Matrices
- e(b) coefficient vector
- e(Cns) constraints matrix
- e(ilog) iteration log
- e(gradient) gradient vector
- e(V) variance–covariance matrix of the estimator

Functions
- e(sample) marks estimation sample
The default approximation of the log likelihood is by adaptive Gauss–Hermite quadrature, which approximates the panel level likelihood with

\[ l_i \approx \sqrt{2} \tilde{\sigma}_i \sum_{m=1}^{M} w_m^* \exp \{ (a_m^*)^2 \} g(y_{it}^o, x_{it}, \sqrt{2} \tilde{\sigma}_i a_m^* + \tilde{\mu}_i) \]

where \( \tilde{\sigma}_i \) and \( \tilde{\mu}_i \) are the adaptive parameters for panel \( i \). Therefore, with the definition of \( g(y_{it}^o, x_{it}, \nu_i) \), the total log likelihood is approximated by

\[ L \approx \sum_{i=1}^{n} w_i \log \left[ \sqrt{2} \tilde{\sigma}_i \sum_{m=1}^{M} w_m^* \exp \{ (a_m^*)^2 \} \exp \left\{ -\left( \sqrt{2} \tilde{\sigma}_i a_m^* + \tilde{\mu}_i \right)^2 / 2\sigma^2 \right\} \right] \]

Letting \( \tau_{i,m,k-1} = \sqrt{2} \tilde{\sigma}_{i,k-1} a_m^* + \tilde{\mu}_{i,k-1} \)

\[ \tilde{\mu}_{i,k} = \sum_{m=1}^{M} \frac{\tau_{i,m,k-1}}{l_{i,k}} \sum_{m=1}^{M} w_m^* \exp \{ (a_m^*)^2 \} g(y_{it}^o, x_{it}, \tau_{i,m,k-1}) \]

and

\[ \tilde{\sigma}_{i,k} = \sum_{m=1}^{M} \left( \frac{\tau_{i,m,k-1}}{l_{i,k}} \right)^2 \sum_{m=1}^{M} w_m^* \exp \{ (a_m^*)^2 \} g(y_{it}^o, x_{it}, \tau_{i,m,k-1}) \]

and this is repeated until \( \tilde{\mu}_{i,k} \) and \( \tilde{\sigma}_{i,k} \) have converged for this iteration of the maximization algorithm. This adaptation is applied on every iteration until the log-likelihood change from the preceding iteration is less than a relative difference of 1e–6; after this, the quadrature parameters are fixed.
The log likelihood can also be calculated by nonadaptive Gauss–Hermite quadrature, the \texttt{int-}
\texttt{method(ghermite)} option:
\begin{align*}
L &= \sum_{i=1}^{n} w_i \log \left\{ \Pr(y_{i1}, \ldots, y_{in_i} | x_{i1}, \ldots, x_{in_i}) \right\} \\
&\approx \sum_{i=1}^{n} w_i \log \left[ \frac{1}{\sqrt{\pi}} \sum_{m=1}^{M} w_m^* \prod_{t=1}^{n_i} F \left( y_{it}^o, x_{it} \beta + \sqrt{2} \sigma \nu^* a_m^* \right) \right]
\end{align*}

Both quadrature formulas require that the integrated function be well approximated by a polynomial
of degree equal to the number of quadrature points. The number of periods (panel size) can affect
whether
\[
\prod_{t=1}^{n_i} F(y_{it}^o, x_{it} \beta + \nu_i)
\]
is well approximated by a polynomial. As panel size and \( \rho \) increase, the quadrature approximation can
become less accurate. For large \( \rho \), the random-effects model can also become unidentified. Adaptive
quadrature gives better results for correlated data and large panels than nonadaptive quadrature; however, we recommend that you use the \texttt{quadchk} command (see \cite{XTquadchk}) to verify the
quadrature approximation used in this command, whichever approximation you choose.

References

Honoré, B. E. 1992. Trimmed LAD and least squares estimation of truncated and censored regression models with
fixed effects. \textit{Econometrica} 60: 533–565.

Naylor, J. C., and A. F. M. Smith. 1982. Applications of a method for the efficient computation of posterior


Structural Equation Models}. Boca Raton, FL: Chapman & Hall/CRC.

Also see

\cite{XTxttobit} \texttt{xttobit postestimation} — Postestimation tools for \texttt{xttobit}

\cite{XTquadchk} \texttt{quadchk} — Check sensitivity of quadrature approximation

\cite{XTxtintreg} \texttt{xtintreg} — Random-effects interval-data regression models

\cite{XTxtreg} \texttt{xtreg} — Fixed-, between-, and random-effects and population-averaged linear models

\cite{Rtobit} \texttt{tobit} — Tobit regression

\cite{U20} \textit{20 Estimation and postestimation commands}
### Description

The following postestimation commands are available after `xttobit`:

<table>
<thead>
<tr>
<th>Command</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>contrast</code></td>
<td>contrasts and ANOVA-style joint tests of estimates</td>
</tr>
<tr>
<td><code>estat ic</code></td>
<td>Akaike’s and Schwarz’s Bayesian information criteria (AIC and BIC)</td>
</tr>
<tr>
<td><code>estat summarize</code></td>
<td>summary statistics for the estimation sample</td>
</tr>
<tr>
<td><code>estat vce</code></td>
<td>variance–covariance matrix of the estimators (VCE)</td>
</tr>
<tr>
<td><code>estimates</code></td>
<td>cataloging estimation results</td>
</tr>
<tr>
<td><code>forecast</code></td>
<td>dynamic forecasts and simulations</td>
</tr>
<tr>
<td><code>lincom</code></td>
<td>point estimates, standard errors, testing, and inference for linear combinations of coefficients</td>
</tr>
<tr>
<td><code>lrtest</code></td>
<td>likelihood-ratio test</td>
</tr>
<tr>
<td><code>margins</code></td>
<td>marginal means, predictive margins, marginal effects, and average marginal effects</td>
</tr>
<tr>
<td><code>marginsplot</code></td>
<td>graph the results from margins (profile plots, interaction plots, etc.)</td>
</tr>
<tr>
<td><code>nlcom</code></td>
<td>point estimates, standard errors, testing, and inference for nonlinear combinations of coefficients</td>
</tr>
<tr>
<td><code>predict</code></td>
<td>predictions, residuals, influence statistics, and other diagnostic measures</td>
</tr>
<tr>
<td><code>predictnl</code></td>
<td>point estimates, standard errors, testing, and inference for generalized predictions</td>
</tr>
<tr>
<td><code>pwcompare</code></td>
<td>pairwise comparisons of estimates</td>
</tr>
<tr>
<td><code>test</code></td>
<td>Wald tests of simple and composite linear hypotheses</td>
</tr>
<tr>
<td><code>testnl</code></td>
<td>Wald tests of nonlinear hypotheses</td>
</tr>
</tbody>
</table>

### Syntax for predict

```
predict [type] newvar [if] [in] [, statistic nooffset]
```

<table>
<thead>
<tr>
<th>statistic</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>xb</code></td>
<td>linear prediction assuming $\nu_i = 0$, the default</td>
</tr>
<tr>
<td><code>stpd</code></td>
<td>standard error of the linear prediction</td>
</tr>
<tr>
<td><code>stdf</code></td>
<td>standard error of the linear forecast</td>
</tr>
<tr>
<td><code>pr0(a,b)</code></td>
<td>$\Pr(a &lt; y &lt; b)$ assuming $\nu_i = 0$</td>
</tr>
<tr>
<td><code>e0(a,b)</code></td>
<td>$E(y \mid a &lt; y &lt; b)$ assuming $\nu_i = 0$</td>
</tr>
<tr>
<td><code>ystar0(a,b)</code></td>
<td>$E(y^<em>)$, $y^</em> = \max{a, \min(y,b)}$ assuming $\nu_i = 0$</td>
</tr>
</tbody>
</table>

These statistics are available both in and out of sample; type `predict ... if e(sample) ...` if wanted only for the estimation sample.
where \( a \) and \( b \) may be numbers or variables; \( a \) missing (\( a \geq . \)) means \(-\infty\), and \( b \) missing (\( b \geq . \)) means \(+\infty\); see [U] 12.2.1 Missing values.

### Menu for predict

Statistics > Postestimation > Predictions, residuals, etc.

### Options for predict

- **xb**, the default, calculates the linear prediction.

- **stdp** calculates the standard error of the prediction. It can be thought of as the standard error of the predicted expected value or mean for the observation’s covariate pattern. The standard error of the prediction is also referred to as the standard error of the fitted value.

- **stdf** calculates the standard error of the forecast. This is the standard error of the point prediction for 1 observation. It is commonly referred to as the standard error of the future or forecast value. By construction, the standard errors produced by **stdf** are always larger than those produced by **stdp**; see Methods and formulas in [R] regress.

- **pr0(\(a, b\))** calculates estimates of \( \Pr(a < y < b \mid x = x_{it}, \nu_i = 0) \), which is the probability that \( y \) would be observed in the interval \((a, b)\), given the current values of the predictors, \( x_{it} \), and given a zero random effect. In the discussion that follows, these two conditions are implied.

- \( a \) and \( b \) may be specified as numbers or variable names; \( lb \) and \( ub \) are variable names; \( \text{pr0}(20,30) \) calculates \( \Pr(20 < y < 30) \);

- \( \text{pr0}(lb,ub) \) calculates \( \Pr(lb < y < ub) \); and

- \( \text{pr0}(20,ub) \) calculates \( \Pr(20 < y < ub) \).

- \( a \) missing (\( a \geq . \)) means \(-\infty\); \( \text{pr0}(.,30) \) calculates \( \Pr(-\infty < y < 30) \);

- \( \text{pr0}(lb,30) \) calculates \( \Pr(-\infty < y < 30) \) in observations for which \( lb \geq . \)

- (and calculates \( \Pr(lb < y < 30) \) elsewhere).

- \( b \) missing (\( b \geq . \)) means \(+\infty\); \( \text{pr0}(20,.) \) calculates \( \Pr(+\infty > y > 20) \);

- \( \text{pr0}(20,ub) \) calculates \( \Pr(+\infty > y > 20) \) in observations for which \( ub \geq . \)

- (and calculates \( \Pr(20 < y < ub) \) elsewhere).

- **e0(\(a, b\))** calculates estimates of \( E(y \mid a < y < b, x = x_{it}, \nu_i = 0) \), which is the expected value of \( y \) conditional on \( y \) being in the interval \((a, b)\), meaning that \( y \) is truncated. \( a \) and \( b \) are specified as they are for **pr0()**.

- **ystar0(\(a, b\))** calculates estimates of \( E(y^* \mid x = x_{it}, \nu_i = 0) \), where \( y^* = a \) if \( y \leq a \), \( y^* = b \) if \( y \geq b \), and \( y^* = y \) otherwise, meaning that \( y^* \) is the censored version of \( y \). \( a \) and \( b \) are specified as they are for **pr0()**.

- **nooffset** is relevant only if you specify offset(varname) for xttobit. It modifies the calculations made by predict so that they ignore the offset variable; the linear prediction is treated as \( x_{it}/\beta \) rather than \( x_{it}/\beta + \text{offset}_{it} \).

### Also see

- [XT] xttobit — Random-effects tobit models

- [U] 20 Estimation and postestimation commands
Syntax

Levin–Lin–Chu test

```
xtrunitroot llc varname [if] [in] [, LLC_options]
```

Harris–Tzavalis test

```
xtrunitroot ht varname [if] [in] [, HT_options]
```

Breitung test

```
xtrunitroot breitung varname [if] [in] [, Breitung_options]
```

Im–Pesaran–Shin test

```
xtrunitroot ips varname [if] [in] [, IPS_options]
```

Fisher-type tests (combining p-values)

```
xtrunitroot fisher varname [if] [in] , {dfuller|pperron} lags(#) [Fisher_options]
```

Hadri Lagrange multiplier stationarity test

```
xtrunitroot hadri varname [if] [in] [, Hadri_options]
```

**LLC_options**

<table>
<thead>
<tr>
<th>Description</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>trend</strong></td>
<td>include a time trend</td>
</tr>
<tr>
<td><strong>noconstant</strong></td>
<td>suppress panel-specific means</td>
</tr>
<tr>
<td><strong>demean</strong></td>
<td>subtract cross-sectional means</td>
</tr>
<tr>
<td><strong>lags(lag_spec)</strong></td>
<td>specify lag structure for augmented Dickey–Fuller (ADF) regressions</td>
</tr>
<tr>
<td><strong>kernel(kernel_spec)</strong></td>
<td>specify method to estimate long-run variance</td>
</tr>
</tbody>
</table>

*lag_spec* is either a nonnegative integer or one of aic, bic, or hqic followed by a positive integer.

*kernel_spec* takes the form `kernel maxlags`, where `kernel` is one of bartlett, parzen, or quadraticspectral and *maxlags* is either a positive number or one of nwest or llc.
**HT_options**

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>trend</code></td>
<td>include a time trend</td>
</tr>
<tr>
<td><code>noconstant</code></td>
<td>suppress panel-specific means</td>
</tr>
<tr>
<td><code>demean</code></td>
<td>subtract cross-sectional means</td>
</tr>
<tr>
<td><code>altt</code></td>
<td>make small-sample adjustment to T</td>
</tr>
</tbody>
</table>

**Breitung_options**

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>trend</code></td>
<td>include a time trend</td>
</tr>
<tr>
<td><code>noconstant</code></td>
<td>suppress panel-specific means</td>
</tr>
<tr>
<td><code>demean</code></td>
<td>subtract cross-sectional means</td>
</tr>
<tr>
<td><code>robust</code></td>
<td>allow for cross-sectional dependence</td>
</tr>
<tr>
<td><code>lags(#)</code></td>
<td>specify lag structure for prewhitening</td>
</tr>
</tbody>
</table>

**IPS_options**

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>trend</code></td>
<td>include a time trend</td>
</tr>
<tr>
<td><code>demean</code></td>
<td>subtract cross-sectional means</td>
</tr>
<tr>
<td><code>lags(lag_spec)</code></td>
<td>specify lag structure for ADF regressions</td>
</tr>
</tbody>
</table>

`lag_spec` is either a nonnegative integer or one of `aic`, `bic`, or `hqic` followed by a positive integer.

**Fisher_options**

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>dfuller</code></td>
<td>use ADF unit-root tests</td>
</tr>
<tr>
<td><code>pperron</code></td>
<td>use Phillips–Perron unit-root tests</td>
</tr>
<tr>
<td><code>lags(#)</code></td>
<td>specify lag structure for prewhitening</td>
</tr>
<tr>
<td><code>dfuller_opts</code></td>
<td>any options allowed by the <code>dfuller</code> command</td>
</tr>
<tr>
<td><code>pperron_opts</code></td>
<td>any options allowed by the <code>pperron</code> command</td>
</tr>
</tbody>
</table>

*Either `dfuller` or `pperron` is required.

*`lags(#)` is required.

**Hadri_options**

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>trend</code></td>
<td>include a time trend</td>
</tr>
<tr>
<td><code>demean</code></td>
<td>subtract cross-sectional means</td>
</tr>
<tr>
<td><code>robust</code></td>
<td>allow for cross-sectional dependence</td>
</tr>
<tr>
<td><code>kernel(kernel_spec)</code></td>
<td>specify method to estimate long-run variance</td>
</tr>
</tbody>
</table>

`kernel_spec` takes the form `kernel [ # ]`, where `kernel` is one of `bartlett`, `parzen`, or `quadraticspectral` and `#` is a positive number.

`varname` may contain time-series operators; see [U] 11.4.4 Time-series varlists.

**Menu**

Statistics > Longitudinal/panel data > Unit-root tests
Description

xtunitroot performs a variety of tests for unit roots (or stationarity) in panel datasets. The Levin–Lin–Chu (2002), Harris–Tzavalis (1999), Breitung (2000; Breitung and Das 2005), Im–Pesaran–Shin (2003), and Fisher-type (Choi 2001) tests have as the null hypothesis that all the panels contain a unit root. The Hadri (2000) Lagrange multiplier (LM) test has as the null hypothesis that all the panels are (trend) stationary. The top of the output for each test makes explicit the null and alternative hypotheses. Options allow you to include panel-specific means (fixed effects) and time trends in the model of the data-generating process.

Options

LLC_options

trend includes a linear time trend in the model that describes the process by which the series is generated.

noconstant suppresses the panel-specific mean term in the model that describes the process by which the series is generated. Specifying noconstant imposes the assumption that the series has a mean of zero for all panels.

lags(lag_spec) specifies the lag structure to use for the ADF regressions performed in computing the test statistic.

Specifying lags(#) requests that # lags of the series be used in the ADF regressions. The default is lags(1).

Specifying lags(aic #) requests that the number of lags of the series be chosen such that the Akaike information criterion (AIC) for the regression is minimized. xtunitroot llc will fit ADF regressions with 1 to # lags and choose the regression for which the AIC is minimized. This process is done for each panel so that different panels may use ADF regressions with different numbers of lags.

Specifying lags(bic #) is just like specifying lags(aic #), except that the Bayesian information criterion (BIC) is used instead of the AIC.

Specifying lags(hqic #) is just like specifying lags(aic #), except that the Hannan–Quinn information criterion is used instead of the AIC.

kernel(kernel_spec) specifies the method used to estimate the long-run variance of each panel’s series. kernel_spec takes the form kernel maxlags. kernel is one of bartlett, parzen, or quadraticspectral. maxlags is a number, nwest to request the Newey and West (1994) bandwidth selection algorithm, or llc to request the lag truncation algorithm in Levin, Lin, and Chu (2002).

Specifying, for example, kernel(bartlett 3) requests the Bartlett kernel with 3 lags.

Specifying kernel(bartlett nwest) requests the Bartlett kernel with the maximum number of lags determined by the Newey and West bandwidth selection algorithm.

Specifying kernel(bartlett llc) requests the Bartlett kernel with a maximum lag determined by the method proposed in Levin, Lin, and Chu’s (2002) article:

\[ \text{maxlags} = \text{int} \left( 3.21 T^{1/3} \right) \]

where \( T \) is the number of observations per panel. This is the default.
demean requests that xtunitroot first subtract the cross-sectional averages from the series. When specified, for each time period xtunitroot computes the mean of the series across panels and subtracts this mean from the series. Levin, Lin, and Chu suggest this procedure to mitigate the impact of cross-sectional dependence.

**HT_options**

trend includes a linear time trend in the model that describes the process by which the series is generated.

noconstant suppresses the panel-specific mean term in the model that describes the process by which the series is generated. Specifying noconstant imposes the assumption that the series has a mean of zero for all panels.

altt requests that xtunitroot use $T - 1$ instead of $T$ in the formulas for the mean and variance of the test statistic under the null hypothesis. When the number of time periods, $T$, is small (less than 10 or 15), the test suffers from severe size distortions when fixed effects or time trends are included; in these cases, using altt results in much improved size properties at the expense of significantly less power.

demean requests that xtunitroot first subtract the cross-sectional averages from the series. When specified, for each time period xtunitroot computes the mean of the series across panels and subtracts this mean from the series. Levin, Lin, and Chu suggest this procedure to mitigate the impact of cross-sectional dependence.

**Breitung_options**

trend includes a linear time trend in the model that describes the process by which the series is generated.

noconstant suppresses the panel-specific mean term in the model that describes the process by which the series is generated. Specifying noconstant imposes the assumption that the series has a mean of zero for all panels.

lags(#) specifies the number of lags used to remove higher-order autoregressive components of the series. The Breitung test assumes the data are generated by an AR(1) process; for higher-order processes, the first-differenced and lagged-level data are replaced by the residuals from regressions of those two series on the first # lags of the first-differenced data. The default is to not perform this prewhitening step.

robust requests a variant of the test that is robust to cross-sectional dependence.

demean requests that xtunitroot first subtract the cross-sectional averages from the series. When specified, for each time period xtunitroot computes the mean of the series across panels and subtracts this mean from the series. Levin, Lin, and Chu suggest this procedure to mitigate the impact of cross-sectional dependence.

**IPS_options**

trend includes a linear time trend in the model that describes the process by which the series is generated.

lags(lag_spec) specifies the lag structure to use for the ADF regressions performed in computing the test statistic. With this option, xtunitroot reports Im, Pesaran, and Shin’s (2003) $W_{t-bar}$ statistic that is predicated on $T$ going to infinity first, followed by $N$ going to infinity. By default, no lags are included, and xtunitroot instead reports Im, Pesaran, and Shin’s $t-bar$ and $Z_{t-bar}$ statistics that assume $T$ is fixed while $N$ goes to infinity, as well as the $t-bar$ statistic and exact critical values that assume both $N$ and $T$ are fixed.
Specifying `lags(#)` requests that # lags of the series be used in the ADF regressions. By default, no lags are included.

Specifying `lags(aic #)` requests that the number of lags of the series be chosen such that the AIC for the regression is minimized. `xtunitroot llc` will fit ADF regressions with 1 to # lags and choose the regression for which the AIC is minimized. This process is done for each panel so that different panels may use ADF regressions with different numbers of lags.

Specifying `lags(bic #)` is just like specifying `lags(aic #)`, except that the BIC is used instead of the AIC.

Specifying `lags(hqic #)` is just like specifying `lags(aic #)`, except that the Hannan–Quinn information criterion is used instead of the AIC.

If you specify `lags(0)`, then `xtunitroot` reports the $W_{t-bar}$ statistic instead of the $Z_{t-bar}$, $Z_{t-bar}$, and $t-bar$ statistics.

demean requests that `xtunitroot` first subtract the cross-sectional averages from the series. When specified, for each time period `xtunitroot` computes the mean of the series across panels and subtracts this mean from the series. Levin, Lin, and Chu suggest this procedure to mitigate the impact of cross-sectional dependence.

**Fisher_options**

`dfuller` requests that `xtunitroot` conduct ADF unit-root tests on each panel by using the `dfuller` command. You must specify either the `dfuller` or the `pperron` option.

`pperron` requests that `xtunitroot` conduct Phillips–Perron unit-root tests on each panel by using the `pperron` command. You must specify either the `pperron` or the `dfuller` option.

`lags(#)` specifies the number of lags used to remove higher-order autoregressive components of the series. The Fisher test assumes the data are generated by an AR(1) process; for higher-order processes, the first-differenced and lagged-level data are replaced by the residuals from regressions of those two series on the first # lags of the first-differenced data. `lags(#)` is required.

`dfuller_opts` are any options accepted by the `dfuller` command, including `noconstant`, `trend`, `drift`, and `lags()`. Because `xtunitroot` calls `dfuller` quietly, the `dfuller` option `regress` has no effect. See [TS] `dfuller`.

`pperron_opts` are any options accepted by the `pperron` command, including `noconstant`, `trend`, and `lags()`. Because `xtunitroot` calls `pperron` quietly, the `pperron` option `regress` has no effect. See [TS] `pperron`.

demean requests that `xtunitroot` first subtract the cross-sectional averages from the series. When specified, for each time period `xtunitroot` computes the mean of the series across panels and subtracts this mean from the series. Levin, Lin, and Chu suggest this procedure to mitigate the impact of cross-sectional dependence.

**Hadri_options**

trend includes a linear time trend in the model that describes the process by which the series is generated.

robust requests a variant of the test statistic that is robust to heteroskedasticity across panels.

`kernel(kernel_spec)` requests a variant of the test statistic that is robust to serially correlated errors. `kernel_spec` specifies the method used to estimate the long-run variance of each panel’s series. `kernel_spec` takes the form `kernel [ #]`. Three kernels are supported: bartlett, parzen, and quadraticspectral.
Specifying, for example, `kernel(bartlett 3)` requests the Bartlett kernel with 3 lags. If `#` is not specified, then 1 lag is used.

demean requests that `xtunitroot` first subtract the cross-sectional averages from the series. When specified, for each time period `xtunitroot` computes the mean of the series across panels and subtracts this mean from the series. Levin, Lin, and Chu suggest this procedure to mitigate the impact of cross-sectional dependence.

Remarks and examples

Remarks are presented under the following headings:

Overview
Levin–Lin–Chu test
Harris–Tsavalis test
Breitung test
Im–Pesaran–Shin test
Fisher-type tests
Hadri LM test

Overview

We consider a simple panel-data model with a first-order autoregressive component:

\[ y_{it} = \rho_i y_{i,t-1} + z_{it}^\prime \gamma_i + \epsilon_{it} \] (1)

where \( i = 1, \ldots, N \) indexes panels; \( t = 1, \ldots, T_i \) indexes time; \( y_{it} \) is the variable being tested; and \( \epsilon_{it} \) is a stationary error term. The \( z_{it} \) term can represent panel-specific means, panel-specific means and a time trend, or nothing, depending on the options specified to `xtunitroot`. By default, \( z_{it} = 1 \), so that the term \( z_{it}^\prime \gamma_i \) represents panel-specific means (fixed effects). If trend is specified, \( z_{it}^\prime = (1, t) \) so that \( z_{it}^\prime \gamma_i \) represents panel-specific means and linear time trends. For tests that allow it, specifying noconstant omits the \( z_{it}^\prime \gamma_i \) term. The Im–Pesaran–Shin (`xtunitroot ips`), Fisher-type (`xtunitroot fisher`), and Hadri LM (`xtunitroot hadri`) tests allow unbalanced panels, while the remaining tests require balanced panels so that \( T_i = T \) for all \( i \).

Panel unit-root tests are used to test the null hypothesis \( H_0: \rho_i = 1 \) for all \( i \) versus the alternative \( H_a: \rho_i < 1 \). Depending on the test, \( H_a \) may hold, for one \( i \), a fraction of all \( i \), or all \( i \); the output of the respective test precisely states the alternative hypothesis. Equation (1) is often written as

\[ \Delta y_{it} = \phi_i y_{i,t-1} + z_{it}^\prime \gamma_i + \epsilon_{it} \] (1')

so that the null hypothesis is then \( H_0: \phi_i = 0 \) for all \( i \) versus the alternative \( H_a: \phi_i < 0 \).

The Hadri LM test for panel stationarity instead assumes the null hypothesis that all panels are stationary versus the alternative that at least some of the panels contain unit roots. We discuss the Hadri LM test in detail later, though for now our remarks focus on tests whose null hypothesis is that the panels contain unit roots.

The various panel unit-root tests implemented by `xtunitroot` differ in several key aspects. First, the Levin–Lin–Chu (`xtunitroot llc`), Harris–Tsavalis (`xtunitroot ht`), and Breitung (`xtunitroot breitung`) tests make the simplifying assumption that all panels share the same autoregressive parameter so that \( \rho_i = \rho \) for all \( i \). The other tests implemented by `xtunitroot`, however, allow the autoregressive parameter to be panel specific. Maddala and Wu (1999) provide an example of testing whether countries’ economic growth rates converge to a long-run value. Imposing the restriction that \( \rho_i = \rho \) for all \( i \) implies that the rate of convergence would be the same for all countries, an implication that is too restrictive in practice.
Second, the various tests make differing assumptions about the rates at which the number of panels, $N$, and the number of time periods, $T$, tend to infinity or whether $N$ or $T$ is fixed. For microeconomic panels of firms, for example, increasing the sample size would involve gathering data on more firms while holding the number of time periods fixed; here $N$ tends to infinity whereas $T$ is fixed. In a macroeconomic analysis of OECD countries, one would typically assume that $N$ is fixed whereas $T$ tends to infinity.

Related to the previous point, the size of one’s sample will in large part determine which test is most appropriate in a given situation. If a dataset has a small number of panels and a large number of time periods, then a panel unit-root test that assumes that $N$ is fixed or that $N$ tends to infinity at a slower rate than $T$ will likely perform better than one that is designed for cases where $N$ is large.

Hlouskova and Wagner (2006) provide a good overview of the types of panel unit-root tests available with xtunitroot, and they present exhaustive Monte Carlo simulations examining the tests’ performance. Baltagi (2013, chap. 12) also concisely discusses the tests implemented by xtunitroot.

The following table summarizes some of the key differences among the various tests:

<table>
<thead>
<tr>
<th>Test</th>
<th>Options</th>
<th>Asymptotics</th>
<th>$\rho$ under $H_a$</th>
<th>Panels</th>
</tr>
</thead>
<tbody>
<tr>
<td>LLC</td>
<td>noconstant</td>
<td>$\sqrt{N/T} \to 0$</td>
<td>common</td>
<td>balanced</td>
</tr>
<tr>
<td>LLC</td>
<td>$N/T \to 0$</td>
<td>$N/T \to 0$</td>
<td>common</td>
<td>balanced</td>
</tr>
<tr>
<td>LLC</td>
<td>trend</td>
<td>$N/T \to 0$</td>
<td>common</td>
<td>balanced</td>
</tr>
<tr>
<td>HT</td>
<td>noconstant</td>
<td>$N \to \infty$, $T$ fixed</td>
<td>common</td>
<td>balanced</td>
</tr>
<tr>
<td>HT</td>
<td>trend</td>
<td>$N \to \infty$, $T$ fixed</td>
<td>common</td>
<td>balanced</td>
</tr>
<tr>
<td>Breitung</td>
<td>noconstant</td>
<td>$(T,N) \to_{seq} \infty$</td>
<td>common</td>
<td>balanced</td>
</tr>
<tr>
<td>Breitung</td>
<td>trend</td>
<td>$(T,N) \to_{seq} \infty$</td>
<td>common</td>
<td>balanced</td>
</tr>
<tr>
<td>IPS</td>
<td>$N \to \infty$, $T$ fixed</td>
<td>panel-specific</td>
<td>unbalanced</td>
<td></td>
</tr>
<tr>
<td>IPS</td>
<td>trend</td>
<td>$N \to \infty$, $T$ fixed</td>
<td>panel-specific</td>
<td>unbalanced</td>
</tr>
<tr>
<td>IPS</td>
<td>lags()</td>
<td>$(T,N) \to_{seq} \infty$</td>
<td>panel-specific</td>
<td>unbalanced</td>
</tr>
<tr>
<td>IPS</td>
<td>trend lags()</td>
<td>$(T,N) \to_{seq} \infty$</td>
<td>panel-specific</td>
<td>unbalanced</td>
</tr>
<tr>
<td>Fisher-type</td>
<td>$T \to \infty$, $N$ finite</td>
<td>panel-specific</td>
<td>unbalanced</td>
<td></td>
</tr>
<tr>
<td>Hadri LM</td>
<td>(not applicable)</td>
<td>$(T,N) \to_{seq} \infty$</td>
<td>balanced</td>
<td></td>
</tr>
<tr>
<td>Hadri LM</td>
<td>trend</td>
<td>$(T,N) \to_{seq} \infty$</td>
<td>(not applicable)</td>
<td>balanced</td>
</tr>
</tbody>
</table>

The first column identifies the test procedure, where we use LLC to denote the Levin–Lin–Chu test, HT to denote the Harris–Tsavalis test, and IPS to denote the Im–Pesaran–Shin test. The second column indicates the deterministic components included in (1) or (1'). The column labeled “Asymptotics” indicates the behavior of the number of panels, $N$, and time periods, $T$, required for the test statistic to have a well-defined asymptotic distribution. For example, the LLC test without the noconstant
option requires that \( T \) grow at a faster rate than \( N \) so that \( N/T \) approaches zero; with the noconstant option, we need only for \( T \) to grow faster than the square root of \( N \) (so \( T \) could grow more slowly than \( N \)).

The HT tests and the IPS tests without accommodations for serial correlation assume that the number of time periods, \( T \), is fixed, whereas \( N \) tends to infinity; \texttt{xtunitroot} also reports critical values for the IPS tests that are valid in finite samples (where \( N \) and \( T \) are fixed).

Many of the tests are justified using sequential limit theory, which we denote as \((T, N) \to_{\text{seq}} \infty\). First, the time dimension goes to infinity, and then the number of panels goes to infinity. As a practical matter, these tests work best with “large” \( T \) and at least “moderate” \( N \). See Phillips and Moon (2000) for an introduction to asymptotics that depend on both \( N \) and \( T \) and their relation to nonstationary panels. Phillips and Moon (1999) contains a more technical discussion of “multi-indexed” asymptotics.

The fourth column refers to the parameter \( \rho_i \) in (1) and \( \phi_i \) in (1′). As we mentioned previously, some tests assume that all panels have the same autoregressive parameter under the alternative hypothesis of stationarity (denoted “common” in the table), while others allow for panel-specific autoregressive parameters (denoted “panel-specific” in the table). The Hadri LM tests are not framed in terms of an equation like (1) or (1′), so the distinction based on \( \rho \) is not applicable.

The final column indicates whether the panel dataset must be strongly balanced, meaning each panel has the same number of observations covering the same time span. Except for the Fisher tests, all the tests require that there be no gaps in any panel’s series.

We now discuss each test in turn.

**Levin–Lin–Chu test**

The starting point for the Levin–Lin–Chu (LLC) test is (1′) with the restriction that all panels share a common autoregressive parameter. In a regression model like (1), \( e_{it} \) is likely to be plagued by serial correlation, so to mitigate this problem, LLC augment the model with additional lags of the dependent variable:

\[
\Delta y_{it} = \phi y_{i,t-1} + z'_{it} \gamma_i + \sum_{j=1}^{p} \theta_{ij} \Delta y_{i,t-j} + u_{it} \tag{2}
\]

The number of lags, \( p \), can be specified using the \texttt{lags()} option, or you can have \texttt{xtunitroot llc} select the number of lags that minimizes one of several information criteria. The LLC test assumes that \( e_{it} \) is independently distributed across panels and follows a stationary invertible autoregressive moving-average process for each panel. By including sufficient lags of \( \Delta y_{i,t} \) in (2), \( u_{it} \) will be white noise; the test does not require \( u_{it} \) to have the same variance across panels.

Under the null hypothesis of a unit root, \( y_{it} \) is nonstationary, so a standard OLS regression \( t \) statistic for \( \phi \) will have a nonstandard distribution that depends in part on the specification of the \( z_{it} \) term. Moreover, the inclusion of a fixed-effect term in a dynamic model like (2) causes the OLS estimate of \( \phi \) to be biased toward zero; see Nickell (1981). The LLC method produces a bias-adjusted \( t \) statistic, which the authors denote as \( t^{*}_{\delta} \), that has an asymptotically normal distribution.

The LLC test without panel-specific intercepts or time trends, requested by specifying the noconstant option with \texttt{xtunitroot llc}, is justified asymptotically if \( \sqrt{N}/T \to 0 \), allowing the time dimension \( T \) to grow more slowly than the cross-sectional dimension \( N \); LLC (2002) mention that this assumption is particularly relevant for panel datasets typically encountered in microeconomic applications.

If model (2) includes panel-specific means (the default for \texttt{xtunitroot llc}) or time trends (requested with the \texttt{trend} option), then you must assume that \( N/T \to 0 \) for the \( t^{*}_{\delta} \) statistic to have
an asymptotically standard normal distribution. This implies that the time dimension, $T$, must grow faster than the cross-sectional dimension, $N$, a situation more plausible with macroeconomic datasets.

LLC (2002) recommend using their test with panels of “moderate” size, which they describe as having between 10 and 250 panels and 25 to 250 observations per panel. Baltagi (2013, sec. 12.2.3) mentions that the requirement $N/T \to 0$ implies that $N$ should be small relative to $T$.

Technical note

Panel unit-root tests have frequently been used to test the purchasing power parity (PPP) hypothesis. We use a PPP dataset to illustrate the xtunitroot command, but understanding PPP is not required to understand how these tests are applied. Here we outline PPP and explain how to test it using panel unit-root tests; uninterested readers can skip the remainder of this technical note. Our discussion and examples are motivated by those in Oh (1996) and Patterson (2000, chap. 13). Also see Rogoff (1996) for a broader introduction to PPP.

The PPP hypothesis is based on the Law of One Price, which stipulates that the price of a tradeable good will be the same everywhere. Absolute PPP stipulates that the nominal exchange rate, $E$, is

$$E = \frac{P}{P^*}$$

where $P$ is the price of a basket of goods in the home country and $P^*$ is the price of the same basket in the foreign country. The exchange rate, $E$, indicates the price of a foreign currency in terms of our “home” currency or, equivalently, how many units of the home currency are needed to buy one unit of the foreign currency.

Now consider the real exchange rate, $\lambda$, which tells us the prices of goods and services—things we actually consume—in a foreign country relative to their prices at home. We have

$$\lambda = \frac{E P^*}{P} \quad \text{(3)}$$

$\lambda$ in general does not equal unity for many reasons, including the fact that not all goods are tradeable across countries (haircuts being the textbook example), trade barriers such as tariffs and quotas, differences among countries in how price indices are constructed, and the Harrod–Balassa–Samuelson effect, which links productivity and price levels; see Obstfeld and Rogoff (1996, 210–216).

Taking logs of both sides of (3), we have

$$y \equiv \ln \lambda = \ln E + \ln P^* - \ln P$$

PPP holds only if the real exchange rate reverts to its equilibrium value over time. Thus, to test for PPP, we test whether $y$ contains a unit root. If $y$ does contain a unit root, we reject PPP.

The dataset pennxrate.dta contains real exchange-rate data based on the Penn World Table version 6.2 (Heston, Summers, and Aten 2006). The data are a balanced panel consisting of 151 countries observed over 34 years, from 1970 through 2003. The United States was treated as the domestic country and is therefore not included. The variable lnrxrate contains the log of the real exchange rate and is the variable on which we conduct panel unit-root tests in the examples.

Two indicator variables are included in the dataset as well. The variable oecd flags 27 countries aside from the United States that are members of the Organization for Economic Cooperation and Development (OECD). (The Czech Republic and the Slovak Republic are excluded because they did not become independent countries until 1993.) The variable g7 flags the six countries aside from the United States that are members of the Group of Seven (G7) nations.
Example 1

The dataset *pennxrate.dta* contains real exchange-rate data for a panel of countries observed over 34 years. Here we use the LLC test to determine whether the series \( \text{lnrxrate} \), the log of real exchange rates, contains a unit root for six nations that are currently in the G7 group of advanced economies. We do not have any reason to believe \( \text{lnrxrate} \) should exhibit a global trend, so we do not include the `trend` option.

Looking at (2), we have no a priori knowledge of the number of lags, \( p \), needed to ensure that \( u_{it} \) is white noise, so we let `xtunitroot` choose the number of lags for each panel by minimizing the AIC, subject to a maximum of 10 lags.

We type

```
. use http://www.stata-press.com/data/r13/pennxrate
. xtunitroot llc lnrxrate if g7, lags(aic 10)
```

Levin-Lin-Chu unit-root test for lnrxrate

<table>
<thead>
<tr>
<th>Statistic</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unadjusted t</td>
<td>-6.7538</td>
</tr>
<tr>
<td>Adjusted t*</td>
<td>-4.0277</td>
</tr>
</tbody>
</table>

The header of the output summarizes the exact specification of the test and dataset. Because we did not specify the `noconstant` option, the test allowed for panel-specific means. On average, \( p = 1 \) lag of the dependent variable of (2) were included as regressors in the ADF regressions. By default, `xtunitroot` estimated the long-run variance of \( \Delta \text{lnrxrate}_{it} \) by using a Bartlett kernel with an average of 10 lags.

The LLC bias-adjusted test statistic \( t_\delta^* = -4.0277 \) is significantly less than zero (\( p < 0.00005 \)), so we reject the null hypothesis of a unit-root [that is, that \( \phi = 0 \) in (2)] in favor of the alternative that \( \text{lnrxrate} \) is stationary (that is, that \( \phi < 0 \)). This conclusion supports the PPP hypothesis.

Labeled “Unadjusted t” in the output is a conventional \( t \) statistic for testing \( H_0: \phi = 0 \). When the model does not include panel-specific means or trends, this test statistic has a standard normal limiting distribution and its \( p \)-value is shown in the output; the unadjusted statistic, \( t_\delta \), diverges to negative infinity if trends or panel-specific constants are included, so a \( p \)-value is not displayed in those cases.

Because the G7 economies have many similarities, our results could be affected by cross-sectional correlation in real exchange rates; O’Connell’s (1998) results showed that the LLC test exhibits severe size distortions in the presence of cross-sectional correlation. LLC (2002) suggested removing cross-sectional averages from the data to help control for this correlation. We can do this by specifying the `demean` option to `xtunitroot`:
Levin-Lin-Chu unit-root test for lnrxrate

<table>
<thead>
<tr>
<th>Statistic</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unadjusted t</td>
<td>-5.5473</td>
</tr>
<tr>
<td>Adjusted t*</td>
<td>-2.0813</td>
</tr>
</tbody>
</table>

Once we control for cross-sectional correlation by removing cross-sectional means, we can no longer reject the null hypothesis of a unit root at the 1% significance level, though we can reject at the 5% level.

Harris–Tsavalis test

In many datasets, particularly in microeconomics, the time dimension, $T$, is small, so tests whose asymptotic properties are established by assuming that $T$ tends to infinity can lead to incorrect inference. HT (1999) derived a unit-root test that assumes that the time dimension, $T$, is fixed. Their simulation results suggest that the test has favorable size and power properties for $N$ greater than 25, and they report (p. 213) that power improves faster as $T$ increases for a given $N$ than when $N$ increases for a given $T$.

The HT test statistic is based on the OLS estimator, $\hat{\rho}$, in the regression model

$$y_{it} = \rho y_{i,t-1} + z_{it}' \gamma_i + \epsilon_{it}$$

(4)

where the term $z_{it}' \gamma_i$ allows for panel-specific means and trends and was discussed in Overview. Harris and Tsavalis assume that $\epsilon_{it}$ is independent and identically distributed (i.i.d.) normal with constant variance across panels. Because of the bias induced by the inclusion of the panel means and time trends in this model, the expected value of the OLS estimator is not equal to unity under the null hypothesis. Harris and Tsavalis derived the mean and standard error of $\hat{\rho}$ for (4) under the null hypothesis $H_0: \rho = 1$ when neither panel-specific means nor time trends are included (requested with the noconstant option), when only panel-specific means are included (the default), and when both panel-specific means and time trends are included (requested with the trend option). The asymptotic distribution of the test statistic is justified as $N \to \infty$, so you should have a relatively large number of panels when using this test. Notice that, like the LLC test, the HT test assumes that all panels share the same autoregressive parameter.
Example 2

Because the HT test is designed for cases where \( N \) is relatively large, here we test whether the series lnrxrate contains a unit root using all 151 countries in our dataset. We will again remove cross-sectional means to help control for contemporaneous correlation. We type

\[ \texttt{. xtunitroot ht lnrxrate, demean} \]

Harris-Tzavalis unit-root test for lnrxrate

<table>
<thead>
<tr>
<th>Statistic</th>
<th>z</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>rho</td>
<td>0.8184</td>
<td>-13.1239</td>
</tr>
</tbody>
</table>

Here we strongly reject the null hypothesis of a unit root, again finding support for PPP. The point estimate of \( \rho \) in (4) is 0.8184, and the \( z \) statistic is -13.12.

Can we directly compare the results from the LLC and HT tests? We used a subset of the data for the LLC test but used all the data for the HT test. That leads to the obvious answer that no, our results are not entirely comparable. However, a more subtle issue regarding the asymptotic properties of the tests also warrants caution when comparing results.

The LLC test assumes that \( N/T \to 0 \), so \( N \) should be small relative to \( T \). Moreover, with our exchange-rate dataset, we are much more likely to be able to add more years of data rather than add more countries, because the number of countries in the world is for the most part fixed. Hence, assuming \( T \) grows faster than \( N \) is certainly plausible.

On the other hand, the HT test assumes that \( T \) is fixed whereas \( N \) goes to infinity. Is that assumption plausible for our dataset? As we just mentioned, \( T \) likely grows faster than \( N \) here, so using a test that assumes \( T \) is fixed whereas \( N \) grows is hard to justify with our dataset.

In short, when selecting a panel unit-root test, you must consider the relative sizes of \( N \) and \( T \) and the relative speeds at which they tend to infinity or whether either \( N \) or \( T \) is fixed.

Breitung test

Both the LLC and HT tests take the approach of first fitting a regression model and subsequently adjusting the autoregressive parameter or its \( t \) statistic to compensate for the bias induced by having a dynamic regressor and fixed effects in the model. The Breitung (2000; Breitung and Das 2005) test takes a different tact, adjusting the data before fitting a regression model so that bias adjustments are not needed.

In the LLC test, additional lags of the dependent variable could be included in (2) to control for serial correlation. The Breitung procedure instead allows for a prewhitening of the series before computing the test. If the \texttt{trend} option is not specified, we regress \( \Delta y_{it} \) and \( y_{i,t-1} \) on \( \Delta y_{i,t-1}, \ldots, \Delta y_{i,t-p} \) and use the residuals from those regressions in place of \( \Delta y_{i,t} \) and \( y_{i,t-1} \) in computing the test. You specify the number of lags, \( p \), to use by specifying \texttt{lags(#)}. If the \texttt{trend} option is specified, then the Breitung method uses a different prewhitening procedure that involves fitting only one (instead of two) preliminary regressions; see \textit{Methods and formulas} for details.
Monte Carlo simulations by Breitung (2000) show that bias-corrected statistics such as LLC’s \( t^* \) suffer from low power, particularly against alternative hypotheses with autoregressive parameters near one and when panel-specific effects are included. In contrast, the Breitung (2000) test statistic exhibits much higher power in these cases. Moreover, the Breitung test has good power even with small datasets \((N = 25, T = 25)\), though the power of the test appears to deteriorate when \( T \) is fixed and \( N \) is increased.

The Breitung test assumes that the error term \( \epsilon_{it} \) is uncorrelated across both \( i \) and \( t \). \texttt{xtunitroot breitung} optionally also reports a version of the statistic based on Breitung and Das (2005) that is robust to cross-sectional correlation.

### Example 3

Here we test whether \( \lnrxrate \) contains a unit root for the subset of 27 OECD countries in our dataset. We will use the \texttt{robust} option to obtain a test statistic that is robust to cross-sectional correlation, so we will not subtract the cross-sectional means via the \texttt{demean} option. We type

```
.xtunitroot breitung lnrxrate if oecd, robust
```

Breitung unit-root test for lnrxrate

<table>
<thead>
<tr>
<th>Statistic</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>lambda*</td>
<td>-1.6794</td>
</tr>
</tbody>
</table>

We can reject the null of a unit root at the 5% level but not at the 1% level.

### Im–Pesaran–Shin test

All the tests we have discussed thus far assume that all panels share a common autoregressive parameter, \( \rho \). Cultural, institutional, and other factors make such an assumption tenuous for both macro- and microeconometric panel datasets. IPS (2003) developed a set of tests that relax the assumption of a common autoregressive parameter. Moreover, the IPS test does not require balanced datasets, though there cannot be gaps within a panel. The starting point for the IPS test is a set of Dickey–Fuller regressions of the form

\[
\Delta y_{it} = \phi_i y_{i,t-1} + z_{it}' \gamma_i + \epsilon_{it}
\]  

(5)

Notice that here \( \phi \) is panel-specific, indexed by \( i \), whereas in (2), \( \phi \) is constant. Im, Pesaran, and Shin assume that \( \epsilon_{it} \) is independently distributed normal for all \( i \) and \( t \), and they allow \( \epsilon_{it} \) to have heterogeneous variances \( \sigma_i^2 \) across panels.

As described by Maddala and Wu (1999), one way to view the key difference between the IPS and LLC tests is that here we fit (5) to each panel separately and average the resulting \( t \) statistics, whereas in the LLC test we pool the data before fitting an equation such as (2) (thus we impose a common autoregressive parameter) and compute a test statistic based on the pooled regression results.
Under the null hypothesis that all panels contain a unit root, we have $\phi_i = 0$ for all $i$. The alternative is that the fraction of panels that follow stationary processes is nonzero; that is, as $N$ tends to infinity, the fraction $N_1/N$ converges to a nonzero value, where $N_1$ is the number of panels that are stationary.

Whether you allow for serially correlated errors determines the test statistics produced, and because there are substantive differences in the output, we consider the serially uncorrelated and serially correlated cases separately. First, we consider the serially uncorrelated case, which xtunitroot assumes when you do not specify the lags() option.

The IPS test allowing for heterogeneous panels with serially uncorrelated errors assumes that the number of time periods, $T$, is fixed; xtunitroot ips produces statistics both for the case where $N$ is fixed and for the case where $N \to \infty$. Under the null hypothesis of a unit root, the usual $t$ statistic, $t_i$, for testing $H_0: \phi_i = 0$ in (5) does not have a mean of zero. For the case where $N$ is fixed, IPS used simulation to tabulate “exact” critical values for the average of the $t_i$ statistics when the dataset is balanced; these critical values are not available with unbalanced datasets. The critical values are “exact” only when the error term is normally distributed and when $T$ corresponds to one of the sample sizes used in their simulation studies. For other values of $T$, xtunitroot ips linearly interpolates the values in IPS (2003, table 2).

For the case where $N \to \infty$, they used simulation to tabulate the mean and variance of $t_i$ for various values of $T$ under the null hypothesis and showed that a bias-adjusted average of the $t_i$’s has a standard normal limiting distribution. We illustrate the test with an example.

Example 4

Here we test whether lnrxrate contains a unit root for the subset of OECD countries. We type

```
. xtunitroot ips lnrxrate if oecd, demean
Im-Pesaran-Shin unit-root test for lnrxrate

Ho: All panels contain unit roots Number of panels = 27
Ha: Some panels are stationary Number of periods = 34
AR parameter: Panel-specific Asymptotics: T,N -> Infinity
Panel means: Included sequentially
Time trend: Not included Cross-sectional means removed
ADF regressions: No lags included

<table>
<thead>
<tr>
<th>Statistic</th>
<th>p-value</th>
<th>Fixed-N exact critical values</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1%</td>
</tr>
<tr>
<td>t-bar</td>
<td>-3.1327</td>
<td>-1.810</td>
</tr>
<tr>
<td>t-tilde-bar</td>
<td>-2.5771</td>
<td></td>
</tr>
<tr>
<td>Z-t-tilde-bar</td>
<td>-7.3911</td>
<td>0.0000</td>
</tr>
</tbody>
</table>
```

As with the other unit-root tests available with xtunitroot, the header of the output contains a summary of the dataset’s dimensions and the null and alternative hypotheses. First, consider the statistic labeled $t$-bar, which IPS denote as $t-bar_{NT}$. This statistic is appropriate when you assume that both $N$ and $T$ fixed; exact critical values reported in IPS (2003) are reported immediately to its right. Here, because $t-bar_{NT}$ is less than even its 1% critical value, we strongly reject the null hypothesis that all series contain a unit root in favor of the alternative that a nonzero fraction of the panels represent stationary processes.

The statistic labeled $t$-tilde-bar is IPS’s $\tilde{t}$-bar$_{NT}$ statistic and is similar to the $t$-bar$_{NT}$ statistic, except that a different estimator of the Dickey–Fuller regression error variance is used. A standardized version of this statistic, $Z_{t-bar}$, is labeled Z-t-tilde-bar in the output and has an asymptotic standard
normal distribution. Here the $p$-value corresponding to $Z_{-tilde-bar}$ is essentially zero, so we strongly reject the null that all series contain a unit root.

### Technical note

Just as the $Z_{-bar}$ statistic corresponds to $\tilde{t}_{NT}$, IPS present a $Z_{t-bar}$ statistic corresponding to $t_{NT}$. However, the $Z_{t-bar}$ statistic does not have an asymptotic normal distribution, and so it is not presented in the output. $Z_{t-bar}$ is available in the stored results as $r(zt)$.

When serial correlation is present, we augment the Dickey–Fuller regression with further lags of the dependent variable:

$$
\Delta y_{it} = \phi_i y_{i,t-1} + z'_{it} \gamma_i + \sum_{j=1}^{p} \Delta y_{i,t-j} + \epsilon_{it}
$$

where the number of lags, $p$, is specified using the `lags()` option, and if the `trend` option is specified, we also include a time trend with panel-specific slope. You can either specify a number or have `xtunitroot` choose the number of lags for each panel by minimizing an information criterion. Here `xtunitroot` produces the IPS $W_{t-bar}$ statistic, which has an asymptotically standard normal distribution as $T \to \infty$ followed by $N \to \infty$. As a practical matter, this means you should have a reasonably large number of both time periods and panels to use this test.

Part of the computation of the $W_{t-bar}$ statistic involves retrieving expected values and variances of the $t$ statistic for $\beta_i$ in (6) in table 3 of IPS (2003). Because expected values have not been computed beyond $p = 8$ lags in (6), you cannot request more than 8 lags in the `lags()` option.

### Example 5

We again test whether $lnrxrate$ contains a unit root for the subset of OECD countries, except we allow for serially correlated errors. We will choose the number of lags for the ADF regressions by minimizing the AIC criterion, subject to a maximum of 8 lags. We type

```
. xtunitroot ips lnrxrate if oecd, lags(aic 8) demean
```

<table>
<thead>
<tr>
<th>Im-Pesaran-Shin unit-root test for lnrxrate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ho: All panels contain unit roots</td>
</tr>
<tr>
<td>Ha: Some panels are stationary</td>
</tr>
<tr>
<td>Number of panels = 27</td>
</tr>
<tr>
<td>Number of periods = 34</td>
</tr>
<tr>
<td>AR parameter: Panel-specific</td>
</tr>
<tr>
<td>Asymptotics: $T, N \to \infty$</td>
</tr>
<tr>
<td>Panel means: Included</td>
</tr>
<tr>
<td>Cross-sectional means removed</td>
</tr>
<tr>
<td>Time trend: Not included</td>
</tr>
<tr>
<td>ADF regressions: 1.48 lags average (chosen by AIC)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Statistic</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$W_{-t-bar}$</td>
<td>-7.3075 0.0000</td>
</tr>
</tbody>
</table>
Fisher-type tests

In our discussion of the IPS test, we intimated that the test statistics could be viewed as averages of bias-adjusted $t$ statistics for each panel. As Maddala and Wu (1999, 635) describe the IPS test, “... the IPS test is a way of combining the evidence on the unit-root hypothesis from the $N$ unit-root tests performed on the $N$ cross-section units.” Fisher-type panel unit-root tests make this approach explicit.

Meta-analysis, frequently used in biostatistics and medical sciences, is the combination of results from multiple studies designed to test a similar hypothesis in order to yield a more decisive conclusion. One type of meta-analysis, first proposed by R. A. Fisher, combines the $p$-values from independent tests to obtain an overall test statistic and is frequently called a Fisher-type test. See Whitehead (2002, sec. 9.8) for an introduction. In the context of panel data unit-root tests, we perform a unit-root test on each panel’s series separately, then combine the $p$-values to obtain an overall test of whether the panel series contains a unit root.

`xtunitroot fisher` performs either ADF or Phillips–Perron unit-root tests on each panel depending on whether you specify the `dfuller` or `pperron` option. The actual tests are conducted by the `dfuller` and `pperron` commands, and you can specify to `xtunitroot fisher` any options those commands take; see [TS] `dfuller` and [TS] `pperron`.

`xtunitroot fisher` combines the $p$-values from the panel-specific unit-root tests using the four methods proposed by Choi (2001). Three of the methods differ in whether they use the inverse $\chi^2$, inverse-normal, or inverse-logit transformation of $p$-values, and the fourth is a modification of the inverse $\chi^2$ transformation that is suitable for when $N$ tends to infinity. The inverse-normal and inverse-logit transformations can be used whether $N$ is finite or infinite.

The null hypothesis being tested by `xtunitroot fisher` is that all panels contain a unit root. For a finite number of panels, the alternative is that at least one panel is stationary. As $N$ tends to infinity, the number of panels that do not have a unit root should grow at the same rate as $N$ under the alternative hypothesis.

例 6

Here we test for a unit root in `lnrXrate` using all 151 countries in our sample. We will use the ADF test. As before, we do not include a trend in real exchange rates and will therefore not specify the `trend` option. However, because the mean real exchange rate for any country is nonzero, we will specify the `drift` option. We will use two lags in the ADF regressions, and we will remove cross-sectional means by using `demean`. We type
. xtunitroot fisher lnrxrate, dfuller drift lags(2) demean

Fisher-type unit-root test for lnrxrate
Based on augmented Dickey-Fuller tests

<table>
<thead>
<tr>
<th>Ho: All panels contain unit roots</th>
<th>Number of panels = 151</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ha: At least one panel is stationary</td>
<td>Number of periods = 34</td>
</tr>
<tr>
<td>AR parameter: Panel-specific</td>
<td>Asymptotics: T -&gt; Infinity</td>
</tr>
<tr>
<td>Panel means: Included</td>
<td>Cross-sectional means removed</td>
</tr>
<tr>
<td>Time trend: Not included</td>
<td>ADF regressions: 2 lags</td>
</tr>
<tr>
<td>Drift term: Included</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Statistic</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inverse chi-squared(302)</td>
<td>P</td>
</tr>
<tr>
<td>Inverse normal</td>
<td>Z</td>
</tr>
<tr>
<td>Inverse logit t(759)</td>
<td>L*</td>
</tr>
<tr>
<td>Modified inv. chi-squared Pm</td>
<td>27.4211</td>
</tr>
</tbody>
</table>

P statistic requires number of panels to be finite.
Other statistics are suitable for finite or infinite number of panels.

All four of the tests strongly reject the null hypothesis that all the panels contain unit roots. Choi’s (2001) simulation results suggest that the inverse normal $Z$ statistic offers the best trade-off between size and power, and he recommends using it in applications. We have observed that the inverse logit $L^*$ test typically agrees with the $Z$ test. Under the null hypothesis, $Z$ has a standard normal distribution and $L^*$ has a $t$ distribution with $5N + 4$ degrees of freedom. Low values of $Z$ and $L^*$ cast doubt on the null hypothesis.

When the number of panels is finite, the inverse $\chi^2$ $P$ test is applicable; this statistic has a $\chi^2$ distribution with $2N$ degrees of freedom, and large values are cause to reject the null hypothesis. Under the null hypothesis, as $T \to \infty$ followed by $N \to \infty$, $P$ tends to infinity so that $P$ has a degenerate limiting distribution. For large panels, Choi (2001) therefore proposes the modified inverse $\chi^2$ $P_m$ test which converges to a standard normal distribution; a large value of $P_m$ casts doubt on the null hypothesis. Choi’s simulation results do not reveal a specific value of $N$ over which $P_m$ should be preferred to $P$, though he mentions that $N = 100$ is still too small for $P_m$ to have an approximately normal distribution.

Hadri LM test

All the tests we have discussed so far take as the null hypothesis that the series contains a unit root. Classical statistical methods are designed to reject the null hypothesis only when the evidence against the null is sufficiently overwhelming. However, because unit-root tests typically are not very powerful against alternative hypotheses of somewhat persistent but stationary processes, reversing roles and testing the null hypothesis of stationarity against the alternative of a unit root is appealing. For pure time series, the KPSS test of Kwiatkowski et al. (1992) is one such test.

The Hadri (2000) LM test uses panel data to test the null hypothesis that the data are stationary versus the alternative that at least one panel contains a unit root. The test is designed for cases with large $T$ and moderate $N$. The motivation for the test is straightforward. Suppose we include a panel-specific time trend (using the trend option with xtunitroot hadri) and write our series, $y_{it}$, as

$$y_{it} = r_{it} + \beta_i t + \epsilon_{it}$$
where \( r_{it} \) is a random walk,
\[
r_{it} = r_{i,t-1} + u_{it}
\]
and \( \epsilon_{it} \) and \( u_{it} \) are zero-mean i.i.d. normal errors. If the variance of \( u_{it} \) were zero, then \( r_{it} \) would collapse to a constant; \( y_{it} \) would therefore be trend stationary. Using this logic, the Hadri LM test tests the hypothesis
\[
H_0: \lambda = \frac{\sigma^2_u}{\sigma^2_\epsilon} = 0 \quad \text{versus} \quad H_a: \lambda > 0
\]

Two options to \texttt{xtunitroot hadri} allow you to relax the assumption that \( \epsilon_{it} \) is i.i.d., though normality is still required. You can specify the \textit{robust} option to obtain a variant of the test that is robust to heteroskedasticity across panels, or you can specify \texttt{kernel()} to obtain a variant that is robust to serial correlation and heteroskedasticity. Asymptotically, the Hadri LM test is justified as \( T \to \infty \) followed by \( N \to \infty \). As a practical matter, \textit{Hadri (2000)} recommends this test for “large” \( T \) and “moderate” \( N \).

\section{Example 7}

We now test the null hypothesis that \texttt{lnrxrate} is stationary for the subset of OECD countries. To control for serial correlation, we will use a Bartlett kernel with 5 lags. We type

\begin{verbatim}
. xtunitroot hadri lnrxrate if oecd, kernel(bartlett 5) demean
Hadri LM test for lnrxrate
Ho: All panels are stationary Number of panels = 27
Ha: Some panels contain unit roots Number of periods = 34
Time trend: Not included Asymptotics: T, N -> Infinity
Heteroskedasticity: Robust sequentially
LR variance: Bartlett kernel, 5 lags Cross-sectional means removed

<table>
<thead>
<tr>
<th>Statistic</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>z</td>
<td>9.6473</td>
</tr>
<tr>
<td>p-value</td>
<td>0.0000</td>
</tr>
</tbody>
</table>
\end{verbatim}

We strongly reject the null hypothesis that all panels’ series are stationary in favor of the alternative that at least one of them contains a unit root. In contrast, the previous examples generally rejected the null hypothesis that all series contain unit roots in favor of the alternative that at least some are stationary. For cautionary remarks on the use of panel unit-root tests in the examination of PPP, see, for example, \textit{Banerjee, Marcellino, and Osbat (2005)}. In short, our results are qualitatively quite similar to those reported in the literature, though \textit{Banerjee, Marcellino, and Osbat} argue that because of cross-unit cointegration and long-run relationships among countries, panel unit-root tests quite often reject the null hypothesis even when true.
xtunitroot llc stores the following in r():

Scalars
- r(N) number of observations
- r(N_g) number of groups
- r(N_t) number of time periods
- r(sig_adj) standard deviation adjustment
- r(mu_adj) mean adjustment
- r(delta) pooled estimate of $\delta$
- r(se_delta) pooled standard error of $\hat{\delta}$
- r(Var_ep) variance of whitened differenced series
- r(sbar) mean of ratio of long-run to innovation standard deviations
- r(ttilde) observations per panel after lagging and differencing
- r(td) unadjusted $t_\delta$ statistic
- r(p_td) $p$-value for $t_\delta$
- r(tds) adjusted $t^*_\delta$ statistic
- r(p_tds) $p$-value for $t^*_\delta$
- r(hac_lags) lags used in HAC variance estimator
- r(hac_lagm) average lags used in HAC estimator
- r(adf_lags) lags used in ADF regressions
- r(adf_lagm) average lags used in ADF regressions

Macros
- r(test) llc
- r(hac_kernel) kernel used in HAC variance estimator
- r(hac_method) HAC lag-selection algorithm
- r(adf_method) ADF regression lag-selection criterion
- r(demean) demean, if the data were demeaned
- r(deterministics) noconstant, constant, or trend

xtunitroot ht stores the following in r():

Scalars
- r(N) number of observations
- r(N_g) number of groups
- r(N_t) number of time periods
- r(rho) estimated $\rho$
- r(Var_rho) variance of $\rho$ under $H_0$
- r(mean_rho) mean of $\rho$ under $H_0$
- r(z) $z$ statistic
- r(p) $p$-value

Macros
- r(test) ht
- r(demean) demean, if the data were demeaned
- r(deterministics) noconstant, constant, or trend
- r(altt) altt, if altt was specified
xtunitroot breitung stores the following in r():

Scalars
- r(N): number of observations
- r(N_g): number of groups
- r(N_t): number of time periods
- r(lambda): test statistic $\lambda$
- r(lrobust): robust test statistic $\lambda_R$
- r(p): p-value for $\lambda$
- r(p_lrobust): p-value for $\lambda_R$
- r(lags): lags used for prewhitening

Macros
- r(test): breitung
- r(demean): demean, if the data were demeaned
- r(robust): robust, if specified
- r(deterministics): noconstant, constant, or trend

xtunitroot ips stores the following in r():

Scalars
- r(N): number of observations
- r(N_g): number of groups
- r(N_t): number of time periods
- r(tbar): test statistic $t_{\bar{bar}NT}$
- r(cv_10): exact 10% critical value for $t_{\bar{bar}NT}$
- r(cv_5): exact 5% critical value for $t_{\bar{bar}NT}$
- r(cv_1): exact 1% critical value for $t_{\bar{bar}NT}$
- r(zt): test statistic $Z_{t_{\bar{bar}}}$
- r(ttildelbar): test statistic $t_{\tilde{bar}NT}$
- r(zttildelbar): test statistic $Z_{\tilde{t}_{\bar{bar}}}$
- r(p_zttildelbar): p-value for $Z_{\tilde{t}_{\bar{bar}}}$
- r(wtbar): test statistic $W_{t_{\bar{bar}}}$
- r(p_wtbar): p-value for $W_{t_{\bar{bar}}}$
- r(lags): lags used in ADF regressions
- r(lagm): average lags used in ADF regressions

Macros
- r(test): ips
- r(demean): demean, if the data were demeaned
- r(ADF_method): ADF regression lag-selection criterion
- r(deterministics): constant or trend
**xtunitroot fisher** stores the following in \( r() \):

Scalars

- \( r(N) \): number of observations
- \( r(N_g) \): number of groups
- \( r(N_t) \): number of time periods
- \( r(P) \): inverse \( \chi^2 \) \( P \) statistic
- \( r(df_P) \): \( P \) statistic degrees of freedom
- \( r(p_P) \): \( p \)-value for \( P \) statistic
- \( r(L) \): inverse logit \( L \) statistic
- \( r(df_L) \): \( L \) statistic degrees of freedom
- \( r(Z) \): inverse normal \( Z \) statistic
- \( r(p_Z) \): \( p \)-value for \( Z \) statistic
- \( r(Pm) \): modified inverse \( \chi^2 \) \( P_m \) statistic
- \( r(p_Pm) \): \( p \)-value for \( P_m \) statistic

Macros

- \( r(test) \): fisher
- \( r(urtest) \): d Fuller or pperron
- \( r(options) \): options passed to d Fuller or pperron
- \( r(demean) \): demean, if the data were demeaned

**xtunitroot hadri** stores the following in \( r() \):

Scalars

- \( r(N) \): number of observations
- \( r(N_g) \): number of groups
- \( r(N_t) \): number of time periods
- \( r(var) \): variance of \( z \) under \( H_0 \)
- \( r(mu) \): mean of \( z \) under \( H_0 \)
- \( r(z) \): test statistic \( z \)
- \( r(p) \): \( p \)-value for \( z \)
- \( r(lags) \): lags used for HAC variance

Macros

- \( r(test) \): hadri
- \( r(demean) \): demean, if the data were demeaned
- \( r(robust) \): robust, if specified
- \( r(kernel) \): kernel used for HAC variance
- \( r(deterministics) \): constant or trend

### Methods and formulas

Methods and formulas are presented under the following headings:

- Levin–Lin–Chu test
- Harris–Tsavalis test
- Breitung test
  - Breitung test without trend
  - Breitung test with trend
- Im–Pesaran–Shin test
- Fisher-type tests
- Hadri LM test

We consider a simple panel-data model with a first-order autoregressive component:

\[
y_{it} = \rho_{i} y_{i,t-1} + z_{it}' \gamma_i + \epsilon_{it}
\]
where \( i = 1, \ldots, N \) indexes panels and \( t = 1, \ldots, T \) indexes time. For the IPS, Fisher-type, and Hadri LM tests, we instead have \( t = 1, \ldots, T_i \), because they do not require balanced panels. \( \epsilon_{it} \) is a zero-mean error term; we discuss the assumptions about \( \epsilon_{it} \) for each test below. Here we use \( N \) to denote the number of panels, not the total number of observations. By default, \( z_{it} = 1 \), so that the term \( z'_{it} \gamma_i \) represents panel-specific means (fixed effects). If \texttt{noconstant} is specified, \( z'_{it} \gamma_i \) vanishes. If \texttt{trend} is specified, \( z_{it}' = (1, t) \) so that \( z_{it}' \gamma_i \) represents panel-specific means and linear time trends.

**Levin–Lin–Chu test**

The starting point for the LLC test is the regression model

\[
\Delta y_{it} = \phi y_{i,t-1} + \sum_{j=1}^{p_i} \theta_{ij} \Delta y_{ij,t-j} + u_{it} \tag{7}
\]

In (1'), LLC assume \( \epsilon_{it} \) is independently distributed across panels and follows a stationary invertible process so that with sufficient lags of \( \Delta y_{it} \) included in (7), \( u_{it} \) will be white noise with potentially heterogeneous variance across panels. If \texttt{lags (#)} is specified with \texttt{xtunitroot llc}, then we set \( p_i = # \) for all panels \( i = 1, \ldots, N \). Otherwise, we fit (7) for each panel individually for lags \( 1 \ldots p_{max} \) and choose the lag length, \( p_i \), that minimizes the information criterion requested by the user. During this step, we restrict estimation to the subset of observations that are valid when \( p_{max} \) lags are included. Information criteria are defined as follows:

\[
\text{AIC} = (-2 \ln L + 2k)/M \\
\text{BIC} = (-2 \ln L + k \ln M)/M \\
\text{HQIC} = (-2 \ln L + 2k \ln \ln M)/M
\]

where \( \ln L \) is the log likelihood assuming Gaussian errors, \( M = T - p_{max} - 2 \), and \( k \) is the number of parameters in (7).

With the lag orders, \( p_i \), in hand, the test proceeds in three main steps, the first of which is to use panel-by-panel OLS regressions to obtain the orthogonalized residuals

\[
\hat{e}_{it} = \Delta y_{it} - \sum_{j=1}^{p_i} \hat{\theta}_{ij} \Delta y_{ij,t-j} - z'_{it} \hat{\gamma}_i \tag{8}
\]

and

\[
\hat{v}_{i,t-1} = y_{i,t-1} - \sum_{j=1}^{p_i} \hat{\theta}_{ij} \Delta y_{ij} - z_{it} \hat{\gamma}_i \tag{9}
\]

To control for panel-level heterogeneity, compute

\[
\tilde{e}_{it} = \hat{e}_{it}/\hat{\sigma}_{ei} \quad \text{and} \quad \tilde{v}_{i,t-1} = \hat{v}_{i,t-1}/\hat{\sigma}_{ei}
\]

where

\[
\hat{\sigma}_{ei}^2 = \frac{1}{T - p_i - 1} \sum_{t=p_i}^{T} \left( \hat{e}_{it} - \hat{\delta}_i \hat{\gamma}_{i,t-1} \right)^2
\]

and \( \hat{\delta}_i \) is the OLS coefficient from a regression of \( \hat{e}_{it} \) on \( \hat{\gamma}_{i,t-1} \). If time trends are included (by specifying the \texttt{trend} option), then a linear time trend is included in regressions (7), (8), and (9).
In the second step, we estimate the ratio of long-run to short-run variances. Under the null hypothesis of a unit root, the long-run variance of the model without panel-specific intercepts or time trends \((z_{it} = \{0\})\) can be estimated as

\[
\hat{\sigma}^2_{yi} = \frac{1}{T-1} \sum_{t=2}^{T} \Delta y_{it}^2 + \frac{2}{T-1} \sum_{j=1}^{m} K(j, m) \left( \sum_{t=j+2}^{T} \Delta y_{it} \Delta y_{i,t-j} \right)
\]

where \(m\) is the maximum number of lags and \(K(j, m)\) is the kernel weight function. Define \(z = j/(m + 1)\). If \(kernel\) is \texttt{bartlett}, then

\[
K(j, m) = \begin{cases} 
1 - z & 0 \leq z \leq 1 \\
0 & \text{otherwise}
\end{cases}
\]

If \(kernel\) is \texttt{parzen}, then

\[
K(j, m) = \begin{cases} 
1 - 6z^2 + 6z^3 & 0 \leq z \leq 0.5 \\
2(1-z)^3 & 0.5 < z \leq 1 \\
0 & \text{otherwise}
\end{cases}
\]

If \(kernel\) is \texttt{quadraticspectral}, then

\[
K(j, m) = \begin{cases} 
1 & z = 0 \\
\frac{3\{\sin(\theta)/\theta - \cos(\theta)\}}{\theta^2} & \text{otherwise}
\end{cases}
\]

where \(\theta = 6\pi z/5\). If the user requests automatic bandwidth (lag) selection using the Newey–West algorithm, then we use the method documented in \textit{Methods and formulas} of \texttt{R} \texttt{ivregress} with \(z_i = h = 1\). If automatic lag selection with the LLC algorithm is chosen, then \(m = \text{int}(3.21T^{1/3})\).

If panel-specific intercepts are included (by not specifying \texttt{noconstant}), then in the formula for \(\hat{\sigma}^2_{yi}\), we replace \(\Delta y_{it}\) with \(\Delta y_{it} - \bar{\Delta} y_{it}\), where \(\bar{\Delta} y_{it}\) is the panel-level mean of \(\Delta y_{it}\) for panel \(i\). Let \(\hat{s}_i = \hat{\sigma}_{yi}/\hat{\sigma}_{\epsilon i}\), and denote \(\hat{S}_N = N^{-1} \sum_i \hat{s}_i\).

In the third step, we run the OLS regression

\[
\tilde{e}_{it} = \delta \tilde{v}_{i,t-1} + \tilde{e}_{it}
\]

Called the “Basic test statistic” in the output of \texttt{xtunitroot llc} is the standard \(t\) statistic for \(\delta\) computed as

\[
t_\delta = \hat{\delta} / \text{se}(\hat{\delta})
\]
where
\[
\text{se}(\hat{\delta}) = \widehat{\sigma}_\epsilon \left( \sum_{i=1}^{N} \sum_{t=p_i+2}^{T} \tilde{v}_{i,t-1}^2 \right)^{-1/2}
\]
\[
\widehat{\sigma}_\epsilon^2 = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=p_i+2}^{T} (\tilde{e}_{it} - \delta \tilde{v}_{i,t-1})^2
\]
and \(\tilde{T} = T - \bar{p} - 1\) with \(\bar{p}\) the average of \(p_1, \ldots, p_N\).

The adjusted test statistic is then computed as
\[
t^*_\delta = \frac{t_{\delta} - N\tilde{T}\widehat{S}_{N}\text{se}(\hat{\delta})\mu_*}{\sigma^*_{T}}
\]
where \(\mu_*_{T}\) and \(\sigma^*_{T}\) are obtained by linearly interpolating the values in LLC (2002, table 2). \(t^*_\delta\) is asymptotically \(N(0, 1)\), with very negative values casting doubt on \(H_0\). If noconstant is specified, then the asymptotic properties hold as \(\sqrt{N/T} \to \infty\). Otherwise, \(T\) must grow at a faster rate so that \(N/T \to \infty\).

**Harris–Tsavalis test**

The starting point for the HT test is (4), where \(\epsilon_{it}\) is assumed to be i.i.d. normal with constant variance across panels. Denote by \(\hat{\rho}\) the least-squares estimate of \(\rho\).

HT show that \(\sqrt{N}(\hat{\rho} - \mu) \xrightarrow{D} N(0, \sigma^2)\) as \(N \to \infty\) with \(T\) fixed, where \(\mu\) and \(\sigma^2\) depend on the specification of the deterministic component:

<table>
<thead>
<tr>
<th>Option</th>
<th>(\mu)</th>
<th>(\sigma^2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>noconstant</td>
<td>1</td>
<td>(\frac{2}{T(T-1)})</td>
</tr>
<tr>
<td>none</td>
<td>(1 - \frac{3}{T+1})</td>
<td>(\frac{3(17T^2 - 20T + 17)}{5(T-1)(T+1)^3})</td>
</tr>
<tr>
<td>trend</td>
<td>(1 - \frac{15}{2(T+2)})</td>
<td>(\frac{15(193T^2 - 728T + 1147)}{112(T+2)^3(T-2)})</td>
</tr>
</tbody>
</table>

**Breitung test**

Suppose the data are generated by an AR(1) process so that we can express \(y_{it}\) as
\[
y_{it} = z'_{it} \gamma_i + x_{it}
\]
where
\[
x_{it} = \alpha_1 x_{i,t-1} + \alpha_2 x_{i,t-2} + \epsilon_{it}
\]
where \(\epsilon_{it}\) is an error term. A prewhitening step is available to correct for serial correlation. The nonrobust version assumes that \(\epsilon_{it}\) is uncorrelated across panels, whereas the robust version allows for the panels to be contemporaneously correlated with covariance matrix \(\Omega\).
Under the null hypothesis that $y_{it}$ contains a unit root, that is, that $y_{it}$ is difference stationary, $\alpha_1 + \alpha_2 = 1$. Under the alternative that $y_{it}$ is stationary, $\alpha_1 + \alpha_2 < 1$. Some of the time indices and summation limits of the formulas below appear more complex than those in Breitung (2000) and Breitung and Das (2005) because our formulas make explicit the loss of observations because of the prewhitening step.

Breitung test without trend

Let $y_{i,t}^\ell = y_{i,t-1} - y_{i,p+1}$ unless noconstant is specified, in which case let $y_{i,t}^\ell = y_{i,t-1}$. If the lags() option is specified with xtunitroot breitung, then we replace $\Delta y_{it}$ and $y_{i,t}^\ell$ in the following description with the residuals from running regressions of $\Delta y_{it}$ and $y_{i,t}^\ell$ on $\Delta y_{i,t-1}, \ldots, \Delta y_{i,t-p}$, where $p$ is the lag order specified in lags().

Define

$$\sigma_i^2 = \frac{1}{T - p - 2} \sum_{t=p+2}^{T} (\Delta y_{it})^2$$

Then

$$\lambda = \frac{\sum_{i=1}^{N} \sum_{t=p+2}^{T} y_{i,t}^\ell \cdot \Delta y_{it} / \sigma_i^2}{\sqrt{\sum_{i=1}^{N} \sum_{t=p+2}^{T} (y_{i,t}^\ell)^2 / \sigma_i^2}}$$

$\lambda$ is asymptotically distributed $N(0, 1)$ as $T \to \infty$ followed by $N \to \infty$; small values of $\lambda$ cast doubt on $H_0$.

For the robust version of the test statistic, let

$$\phi = \frac{\sum_{i=1}^{N} \sum_{t=p+2}^{T} y_{i,t}^\ell \cdot \Delta y_{it} / \sigma_i^2}{\sum_{i=1}^{N} \sum_{t=p+2}^{T} (y_{i,t}^\ell)^2 / \sigma_i^2}$$

and define $u_{it} = \Delta y_{it} - \phi y_{i,t}^\ell$. Let $u_i = (u_{i,p+2}, \ldots, u_{iT})'$ and let the $N \times N$ matrix $\Omega$ have typical element $u_i' u_j / (T - p - 2)$. Let $\Delta y_t = (\Delta y_{1t}, \ldots, \Delta y_{Nt})'$ and $y_t^\ell = (y_{1,t-1}, \ldots, y_{N,t-1})'$. Then

$$\lambda_{\text{robust}} = \frac{\sum_{t=p+2}^{T} (\Delta y_t)' y_t^\ell}{\sum_{t=p+2}^{T} (y_t^\ell)' \Omega y_t^\ell}$$

For $\Omega$ to be positive definite, we must have $T - p - 1 \geq N$. As a practical matter, for $\Omega$ to have good finite-sample properties, we need $T \gg N$. $\lambda_{\text{robust}}$ is asymptotically distributed $N(0, 1)$ as $T \to \infty$ followed by $N \to \infty$; very negative values of $\lambda_{\text{robust}}$ cast doubt on $H_0$.

Breitung test with trend

Let $p$ denote the number of lags requested in the lags() option. We fit the regression

$$\Delta y_{it} = \alpha_{i0} + \sum_{j=1}^{p} \alpha_{ij} \Delta y_{i,t-j} + \nu_{it}$$

and compute the $1 \times (T - p - 1)$ vectors $\Delta u_i$ and $u_i^\ell$ with typical elements

$$\Delta u_{is} = \Delta y_{is} - \sum_{j=1}^{p} \hat{\alpha}_{ij} \Delta y_{i,s-j}$$
and

\[ u_{is}^\ell = y_{i,s-1} - \sum_{j=1}^{p} \hat{\alpha}_{ij} y_{i,s-j-1} \]

for \( s = 1, \ldots, T - p - 1 \). Let

\[ \sigma_i^2 = \frac{1}{T - p - 2} \sum_{s=1}^{T-p-1} \left( \Delta u_{is} - \bar{\Delta u}_i \right) \Delta u_{is} \]

where \( \bar{\Delta u}_i \) is the mean of \( \Delta u_{is} \) over \( s \). Let \( \Delta v_i \) and \( v_i^\ell \) denote \( 1 \times (T - p - 1) \) vectors with typical elements

\[ \Delta v_{is} = \sqrt{\frac{T - p - s - 1}{T - p - s}} \left( \Delta u_{is} - \frac{1}{T - p - s - 1} \sum_{j=s+1}^{T-p-1} \Delta u_{ij} \right) \]

and

\[ v_{is}^\ell = u_{is}^\ell - u_{i1}^\ell - (T - p - 1) \bar{\Delta u}_i \]

Now

\[ \lambda = \frac{\sum_{i=1}^{N} \sum_{s=1}^{T-p-1} v_{is}^\ell \Delta v_{is} / \sigma_i^2}{\sqrt{\sum_{i=1}^{N} \sum_{s=1}^{T-p-1} (v_{is}^\ell)^2 / \sigma_i^2}} \]

\( \lambda \) is asymptotically distributed \( N(0, 1) \) as \( T \to \infty \) followed by \( N \to \infty \); very negative values of \( \lambda \) cast doubt on \( H_0 \). The computation of the robust form of the statistic proceeds in a fashion entirely analogous to the case without trend.

**Im–Pesaran–Shin test**

Write the model as

\[ \Delta y_{it} = \phi_i y_{i,t-1} + z_{it}' \gamma_i + \epsilon_{it} \]

where \( \epsilon_{it} \) is independently distributed normal for all \( i \) and \( t \) with panel-specific variance \( \sigma_i^2 \). Denote \( \Delta y_i = (\Delta y_{i2}, \ldots, \Delta y_{iT})' \) and \( y_{i,-1} = (y_{i1}, \ldots, y_{iT-1})' \). Note that to be consistent with the notation used in the rest of this documentation, we start the time index at \( t = 1 \) instead of \( t = 0 \) as in IPS (2003). Also let \( \tau_T \) be a conformable vector of ones, \( M_\tau = I - \tau_T (\tau_T' \tau_T)^{-1} \tau_T' \), \( X_i = (\tau_T, y_{i,-1}) \), and \( M_{X_i} = I - X_i (X_i' X_i)^{-1} X_i' \).

First, we consider the case of no serial correlation, where the user does not specify the \texttt{lags()} option. Then

\[ t-bar_{NT} = \frac{1}{N} \sum_{i=1}^{N} \tilde{t}_{iT} \]

where

\[ \tilde{t}_{iT} = \frac{\Delta y_i' M_\tau y_{i,-1}}{\tilde{\sigma}_{iT} (y_{i,-1}' M_\tau y_{i,-1})^{1/2}} \]

and

\[ \tilde{\sigma}_{iT}^2 = \frac{\Delta y_i' M_\tau \Delta y_i}{T - 1} \]
Also

\[ t-bar_{NT} = \frac{1}{N} \sum_{i=1}^{N} t_{iT} \]

where

\[ t_{iT} = \frac{\Delta y'_{i} M_{r} y_{i,-1}}{\hat{\sigma}_{iT} (y'_{i,-1} M_{r} y_{i,-1})^{1/2}} \]

and

\[ \tilde{\sigma}_{iT}^2 = \frac{\Delta y'_{i} M_{X_i} \Delta y_i}{T - 1} \]

Now

\[ Z_{t-bar} = \frac{\sqrt{N} \{ t-bar_{NT} - N^{-1} \sum_{i=1}^{N} E(t_{iT}) \} }{\sqrt{N^{-1} \sum_{i} \text{Var}(t_{iT})}} \]

where \( E(t_{iT}) \) and \( \text{Var}(t_{iT}) \) are obtained by linearly interpolating the values shown in IPS (2003, table 1). \( Z_{t-bar} \) has a standard normal limiting distribution for fixed \( T \) and \( N \to \infty \); very negative values cast doubt on \( H_0 \). Similarly,

\[ Z_{t-bar} = \frac{\sqrt{N} \{ t-bar_{NT} - N^{-1} \sum_{i=1}^{N} E(t_{iT}) \} }{\sqrt{N^{-1} \sum_{i} \text{Var}(t_{iT})}} \]

If the \texttt{lags()} option is specified, then we fit the ADF regressions

\[ \Delta y_{it} = \phi_i y_{i,t-1} + z_{it}' \gamma_i + \sum_{j=1}^{p_i} \rho_{ij} \Delta y_{i,t-j} + \epsilon_{it} \]

In matrix form, we can write this more compactly as

\[ \Delta y_i = \phi_i y_{i,-1} + Q_i \theta_i + \epsilon_i \]

where \( Q_i = (\tau_t, \Delta y_{i,-1}, \ldots, \Delta y_{i,-p_i}) \) and \( \theta_i = (\alpha_i, \rho_{i1}, \ldots, \rho_{ip_i})' \). Then

\[ t-bar_{NT} = \frac{1}{N} \sum_{i=1}^{N} t_{iT}(p_i) \]

where

\[ t_{iT}(p_i) = \frac{\sqrt{T - p_i - 2} (y'_{i,-1} M_{Q_i} \Delta y_i) (\Delta y'_{i,-1} M_{Q_i} \Delta y_{i,-1})^{1/2}}{\Delta y'_{i,-1} M_{Q_i} \Delta y_{i,-1})^{1/2}} \]

where \( M_{Q_i} = I - Q_i (Q_i' Q_i)^{-1} Q_i' \), \( M_{X_i} = I - X_i (X_i' X_i)^{-1} X_i' \), and \( X_i = (y_{i,-1}, Q_i) \). Finally,

\[ W_{t-bar}(p) = \frac{\sqrt{N} \left[ t-bar_{NT} - N^{-1} \sum_{i=1}^{N} E \{ t_{iT}(p_i) \} \right]}{\sqrt{N^{-1} \sum_{i=1}^{N} \text{Var} \{ t_{iT}(p_i) \}}} \]

where \( E \{ t_{iT}(p_i) \} \) and \( \text{Var} \{ t_{iT}(p_i) \} \) are obtained by linearly interpolating the values shown in IPS (2003, table 3). \( W_{t-bar}(p) \) has a standard normal limiting distribution as \( T \to \infty \) followed by \( N \to \infty \); very negative values cast doubt on \( H_0 \).
Fisher-type tests

We use `dfuller` or `pperron` to perform unit-root tests on each panel; denote the \( p \)-value for the respective test on the \( i \)th panel as \( p_i \). All of these tests are predicated on \( T \to \infty \) so that the unit-root test for each panel is consistent. The \( P \) test is for finite \( N \); the other tests are valid whether \( N \) is finite or infinite. Then

\[
P = -2 \sum_{i=1}^{N} \ln(p_i)
\]

\( P \sim \chi^2(2N) \) and large values cast doubt on \( H_0 \).

\[
Z = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \Phi^{-1}(p_i)
\]

where \( \Phi^{-1}(\cdot) \) is the inverse of the standard normal cumulative distribution function. \( Z \sim N(0, 1) \); very negative values of \( Z \) cast doubt on \( H_0 \).

\[
L = \sum_{i=1}^{N} \ln \left( \frac{p_i}{1 - p_i} \right)
\]

\( L^* = \sqrt{k}L \sim t(5N + 4) \) where

\[
k = \frac{3(5N + 4)}{\pi^2N(5N + 2)}
\]

Very negative values of \( L^* \) cast doubt on \( H_0 \). Finally,

\[
P_m = -\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \{ \ln(p_i) + 1 \}
\]

\( P_m \sim N(0, 1) \); very positive values of \( P_m \) cast doubt on \( H_0 \).

Hadri LM test

As discussed in the main text, the Hadri LM test can be viewed as a test of \( H_0 : \sigma_u^2/\sigma^2_\epsilon = 0 \), where both \( u_{it} \) and \( \epsilon_{it} \) are normally distributed random errors.

Let \( \hat{\epsilon}_{it} \) denote the residuals from a regression of \( y_{it} \) on a panel-specific intercept or a panel-specific intercept and time trend if \text{trend} is specified. Then

\[
\hat{L}_M = \frac{1}{N} \sum_i \frac{1}{T^2} \sum_t S_{it}^2
\]

(10)

where

\[
S_{it} = \sum_{j=1}^{t} \hat{\epsilon}_{ij}
\]

and

\[
\hat{\sigma}_\epsilon^2 = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \hat{\epsilon}_{it}^2
\]
where $T' = T - 2$ if `trend` is specified and $T' = T - 1$ otherwise. Then

$$Z = \frac{\sqrt{N} (\hat{L}M - \mu)}{\sigma}$$

where $\mu = 1/15$ and $\sigma = 11/6300$ if `trend` is specified and $\mu = 1/6$ and $\sigma = 1/45$ otherwise. $Z \sim N(0, 1)$ asymptotically as $T \to \infty$ followed by $N \to \infty$. Very positive values of $Z$ cast doubt on $H_0$. If `robust` is specified, then we instead use

$$\hat{L}M = \frac{1}{N} \sum_{i=1}^{N} \left( \frac{\sum_{t=1}^{T} S_{it}^2}{T^2 \hat{\sigma}_{\epsilon,i}^2} \right)$$

where we calculate $\hat{\sigma}_{\epsilon,i}^2$ individually for each panel:

$$\hat{\sigma}_{\epsilon,i}^2 = \sum_{t=1}^{T} \hat{\epsilon}_{it}^2$$

If `kernel()` is specified, then we use (10) with

$$\hat{\sigma}_{\epsilon}^2 = \frac{1}{N} \sum_{i=1}^{N} \left\{ \frac{1}{T} \sum_{t=p+1}^{T} \hat{\epsilon}_{it}^2 + \frac{2}{T} \sum_{j=1}^{m} K(j, m) \sum_{t=j+1}^{T} \hat{\epsilon}_{it} \hat{\epsilon}_{i,t-j} \right\}$$

where $m$ is the maximum number of lags and $K(., .)$ is the kernel function defined previously.

**Acknowledgments**

We gratefully acknowledge users Fabian Bornhorst of the International Monetary Fund and Christopher F. Baum of the Department of Economics at Boston College and author of the Stata Press books *An Introduction to Modern Econometrics Using Stata* and *An Introduction to Stata Programming* for the `ipshin` and `levinlin` commands; Christopher Baum for the `hadrilm` command; and Scott Merryman of the Risk Management Agency for the `xtfisher` command. We also gratefully acknowledge Jörg Breitung of the Department of Economics at the Universität Bonn for useful discussions when implementing his tests. These contributions were beneficial during the development of `xtunitroot`.

**References**


Also see

[TS] *dfgls* — DF-GLS unit-root test

[TS] *dfuller* — Augmented Dickey–Fuller unit-root test

[TS] *pperron* — Phillips–Perron unit-root test
Glossary

Arellano–Bond estimator. The Arellano–Bond estimator is a generalized method of moments (GMM) estimator for linear dynamic panel-data models that uses lagged levels of the endogenous variables as well as first differences of the exogenous variables as instruments. The Arellano–Bond estimator removes the panel-specific heterogeneity by first-differencing the regression equation.

autoregressive process. In autoregressive processes, the current value of a variable is a linear function of its own past values and a white-noise error term. For panel data, a first-order autoregressive process, denoted as an AR(1) process, is $y_{it} = \rho y_{i,t-1} + \epsilon_{it}$, where $i$ denotes panels, $t$ denotes time, and $\epsilon_{it}$ is white noise.

balanced data. A longitudinal or panel dataset is said to be balanced if each panel has the same number of observations. See also weakly balanced and strongly balanced.

between estimator. The between estimator is a panel-data estimator that obtains its estimates by running OLS on the panel-level means of the variables. This estimator uses only the between-panel variation in the data to identify the parameters, ignoring any within-panel variation. For it to be consistent, the between estimator requires that the panel-level means of the regressors be uncorrelated with the panel-specific heterogeneity terms.

canonical link. Corresponding to each family of distributions in a generalized linear model is a canonical link function for which there is a sufficient statistic with the same dimension as the number of parameters in the linear predictor. The use of canonical link functions provides the GLM with desirable statistical properties, especially when the sample size is small.

conditional fixed-effects model. In general, including panel-specific dummies to control for fixed effects in nonlinear models results in inconsistent estimates. For some nonlinear models, the fixed-effect term can be removed from the likelihood function by conditioning on a sufficient statistic. For example, the conditional fixed-effect logit model conditions on the number of positive outcomes within each panel.

correlation structure. A correlation structure is a set of assumptions imposed on the within-panel variance–covariance matrix of the errors in a panel-data model. See [XT] xtgee for examples of different correlation structures.

cross-sectional data. Cross-sectional data refers to data collected over a set of individuals, such as households, firms, or countries sampled from a population at a given point in time.

cross-sectional time-series data. Cross-sectional time-series data is another name for panel data. The term cross-sectional time-series data is sometimes reserved for datasets in which a relatively small number of panels were observed over many periods. See also panel data.

disturbance term. The disturbance term encompasses any shocks that occur to the dependent variable that cannot be explained by the conditional (or deterministic) portion of the model.

dynamic model. A dynamic model is one in which prior values of the dependent variable or disturbance term affect the current value of the dependent variable.

endogenous variable. An endogenous variable is a regressor that is correlated with the unobservable error term. Equivalently, an endogenous variable is one whose values are determined by the equilibrium or outcome of a structural model.

error-components model. The error-components model is another name for the random-effects model. See also random-effects model.
exogenous variable. An exogenous variable is a regressor that is not correlated with any of the unobservable error terms in the model. Equivalently, an exogenous variable is one whose values change independently of the other variables in a structural model.

fixed-effects model. The fixed-effects model is a model for panel data in which the panel-specific errors are treated as fixed parameters. These parameters are panel-specific intercepts and therefore allow the conditional mean of the dependent variable to vary across panels. The linear fixed-effects estimator is consistent, even if the regressors are correlated with the fixed effects. See also random-effects model.

generalized estimating equations (GEE). The method of generalized estimating equations is used to fit population-averaged panel-data models. GEE extends the GLM method by allowing the user to specify a variety of different within-panel correlation structures.

generalized linear model. The generalized linear model is an estimation framework in which the user specifies a distributional family for the dependent variable and a link function that relates the dependent variable to a linear combination of the regressors. The distribution must be a member of the exponential family of distributions. The generalized linear model encompasses many common models, including linear, probit, and Poisson regression.

idiosyncratic error term. In longitudinal or panel-data models, the idiosyncratic error term refers to the observation-specific zero-mean random-error term. It is analogous to the random-error term of cross-sectional regression analysis.

instrumental variables. Instrumental variables are exogenous variables that are correlated with one or more of the endogenous variables in a structural model. The term instrumental variable is often reserved for those exogenous variables that are not included as regressors in the model.

instrumental-variables (IV) estimator. An instrumental variables estimator uses instrumental variables to produce consistent parameter estimates in models that contain endogenous variables. IV estimators can also be used to control for measurement error.

interval data. Interval data are data in which the true value of the dependent variable is not observed. Instead, all that is known is that the value lies within a given interval.

link function. In a generalized linear model, the link function relates a linear combination of predictors to the expected value of the dependent variable. In a linear regression model, the link function is simply the identity function.

longitudinal data. Longitudinal data is another term for panel data. See also panel data.

negative binomial regression model. The negative binomial regression model is for applications in which the dependent variable represents the number of times an event occurs. The negative binomial regression model is an alternative to the Poisson model for use when the dependent variable is overdispersed, meaning that the variance of the dependent variable is greater than its mean.

overidentifying restrictions. The order condition for model identification requires that the number of exogenous variables excluded from the model be at least as great as the number of endogenous regressors. When the number of excluded exogenous variables exceeds the number of endogenous regressors, the model is overidentified, and the validity of the instruments can then be checked via a test of overidentifying restrictions.

panel data. Panel data are data in which the same units were observed over multiple periods. The units, called panels, are often firms, households, or patients who were observed at several points in time. In a typical panel dataset, the number of panels is large, and the number of observations per panel is relatively small.
**Glossary 473**

**panel-corrected standard errors (PCSEs).** The term *panel-corrected standard errors* refers to a class of estimators for the variance–covariance matrix of the OLS estimator when there are relatively few panels with many observations per panel. PCSEs account for heteroskedasticity, autocorrelation, or cross-sectional correlation.

**Poisson regression model.** The Poisson regression model is used when the dependent variable represents the number of times an event occurs. In the Poisson model, the variance of the dependent variable is equal to the conditional mean.

**pooled estimator.** A pooled estimator ignores the longitudinal or panel aspect of a dataset and treats the observations as if they were cross-sectional.

**population-averaged model.** A population-averaged model is used for panel data in which the parameters measure the effects of the regressors on the outcome for the average individual in the population. The panel-specific errors are treated as uncorrelated random variables drawn from a population with zero mean and constant variance, and the parameters measure the effects of the regressors on the dependent variable after integrating over the distribution of the random effects.

**predetermined variable.** A predetermined variable is a regressor in which its contemporaneous and future values are not correlated with the unobservable error term but past values are correlated with the error term.

**prewhiten.** To prewhiten is to apply a transformation to a time series so that it becomes white noise.

**production function.** A production function describes the maximum amount of a good that can be produced, given specified levels of the inputs.

**quadrature.** Quadrature is a set of numerical methods to evaluate a definite integral.

**random-coefficients model.** A random-coefficients model is a panel-data model in which group-specific heterogeneity is introduced by assuming that each group has its own parameter vector, which is drawn from a population common to all panels.

**random-effects model.** A random-effects model for panel data treats the panel-specific errors as uncorrelated random variables drawn from a population with zero mean and constant variance. The regressors must be uncorrelated with the random effects for the estimates to be consistent. See also *fixed-effects model*.

**robust standard errors.** Robust standard errors, also known as Huber/White or Taylor linearization standard errors, are based on the sandwich estimator of variance. Robust standard errors can be interpreted as representing the sample-to-sample variability of the parameter estimates, even when the model is misspecified. See also *semirobust standard errors*.

**semirobust standard errors.** Semirobust standard errors are closely related to robust standard errors and can be interpreted as representing the sample-to-sample variability of the parameter estimates, even when the model is misspecified, as long as the mean structure of the model is specified correctly. See also *robust standard errors*.

**sequential limit theory.** The sequential limit theory is a method of determining asymptotic properties of a panel-data statistic in which one index, say, $N$, the number of panels, is held fixed, while $T$, the number of time periods, goes to infinity, providing an intermediate limit. Then one obtains a final limit by studying the behavior of this intermediate limit as the other index ($N$ here) goes to infinity.

**strongly balanced.** A longitudinal or panel dataset is said to be strongly balanced if each panel has the same number of observations, and the observations for different panels were all made at the same times.
**unbalanced data.** A longitudinal or panel dataset is said to be unbalanced if each panel does not have the same number of observations. See also *weakly balanced* and *strongly balanced.*

**weakly balanced.** A longitudinal or panel dataset is said to be weakly balanced if each panel has the same number of observations but the observations for different panels were not all made at the same times.

**white noise.** A variable $u_t$ represents a white-noise process if the mean of $u_t$ is zero, the variance of $u_t$ is $\sigma^2$, and the covariance between $u_t$ and $u_s$ is zero for all $s \neq t$.

**within estimator.** The within estimator is a panel-data estimator that removes the panel-specific heterogeneity by subtracting the panel-level means from each variable and then performing ordinary least squares on the demeaned data. The within estimator is used in fitting the linear fixed-effects model.
Subject and author index

This is the subject and author index for the *Longitudinal-Data/Panel-Data Reference Manual*. Readers interested in topics other than cross-sectional time-series should see the combined subject index (and the combined author index) in the *Glossary and Index*.

A

abond, estat subcommand, [XT] xtabond postestimation, [XT] xtreg postestimation, [XT] xtdpdsys postestimation
Aigner, D. J., [XT] xfronter
Albert, P. S., [XT] xtgee
Alvarez, J., [XT] xtabond
Amemiya, T., [XT] xthtaylor, [XT] xtivreg
Andrews, M. J., [XT] xtreg
Arelano–Bond estimator, [XT] xtreg
Arelano–Bover estimator, [XT] xtreg
Arellano–Bover estimator, [XT] xtreg
Aten, B., [XT] xtunitroot
Atella, V., [XT] xtfrontier
B
balanced data, [XT] xtivreg
Balestra, P., [XT] xthtaylor
Balotti, F., [XT] xthtaylor
Banerjee, A., [XT] xtunitroot
Battese–Coelli parameterization, [XT] xtfrontier
Battese, G. E., [XT] xtfrontier
Beck, N. L., [XT] xtivreg, [XT] xtpcse
Bellocco, R., [XT] xtgee
Belotti, F., [XT] xtfrontier
between estimators, [XT] xtivreg, [XT] xtreg, [XT] xtreg
between-cell means and variances, [XT] xtdescribe, [XT] xtsum
Bhargava, A., [XT] xtreg
binary outcome model, see outcomes, binary
Blundell–Bond estimator, [XT] xtdp, [XT] xtdpdsys
Blundell, R., [XT] xtdp, [XT] xtdpdsys
Bootstrap standard errors, [XT] vce options
Bornhorst, F., [XT] xtunitroot
Bottai, M., [XT] xtreg
Breitung, J., [XT] xtunitroot
Breitung test, [XT] xtunitroot
Breusch–Pagan Lagrange multiplier test, [XT] xtreg postestimation
Breusch, T. S., [XT] xtreg postestimation
Bruno, G. S. F., [XT] xtreg
B
Cameron, A. C., [XT] xt, [XT] xtnbreg, [XT] xtpoisson
canonical link, [XT] xtivreg
Caria, M. P., [XT] xtlogit
Carter, S. L., [XT] xtpoisson
categorical outcomes, see outcomes, binary, see outcomes, ordinal
Caudill, S. B., [XT] xtfrontier
Center for Human Resource Research, [XT] xt
Chang, Y.-J., [XT] xtivreg, [XT] xtreg
Choi, I., [XT] xtunitroot
Chu, C.-S. J., [XT] xtunitroot
cluster estimator of variance, [XT] vce options
fixed-effects models, linear, [XT] xtreg
Poisson, [XT] xtpoisson
population-averaged models, [XT] xtgee
complementary log-log, [XT] xtcloglog
Gottlieb, A. S., [XT] xthtaylor
Gottlieb, A. S., [XT] xthtaylor
Gottlieb, A. S., [XT] xthtaylor
Gottlieb, A. S., [XT] xthtaylor
Gottlieb, A. S., [XT] xthtaylor
Gottlieb, A. S., [XT] xthtaylor
Gottlieb, A. S., [XT] xthtaylor

cluster estimator of variance, population-averaged models, continued
logit, [XT] xtlogit
negative binomial, [XT] xtnbreg
Poisson, [XT] xtpoisson
probit, [XT] xtprobit
random-effects models
complementary log-log, [XT] xtcloglog
linear, [XT] xtreg
logistic, [XT] xtlogit, [XT] xtologit
Poisson, [XT] xtpoisson
probit, [XT] xtprobit, [XT] xtprobit
Coelli, T. J., [XT] xtfrontier
complementary log-log regression, [XT] xtcloglog,
[XT] xtgee
conditional
fixed-effects model, [XT] Glossary
logistic regression, [XT] xtlogit, [XT] xtologit
confidence interval
for incidence-rate ratios, [XT] xtgee, [XT] xtnbreg,
[XT] xtpoisson
for odds ratios, [XT] xtcloglog, [XT] xtgee,
[XT] xtlogit, [XT] xtologit
constrained estimation,
fixed-effects models
logit, [XT] xtlogit
negative binomial, [XT] xtnbreg
Poisson, [XT] xtpoisson
random-effects models
complementary log-log, [XT] xtcloglog
interval-data regression, [XT] xtintreg
logit, [XT] xtlogit, [XT] xtologit
negative binomial, [XT] xtnbreg
Poisson, [XT] xtpoisson
probit, [XT] xtprobit, [XT] xtprobit
tobit, [XT] xtobit
stochastic frontier models for panel data,
[XT] xtfrontier
Conway, M. R., [XT] xtlogit, [XT] xtologit,
[XT] xtprobit, [XT] xtprobit
correlated error, see robust, Huber/White/sandwich estimator of variance
correlation structure, [XT] xtcloglog, [XT] xtgee,
[XT] xtgls, [XT] xtlogit, [XT] xtnbreg,
[XT] xtpcse, [XT] xtpoisson, [XT] xtprobit,
[XT] xtreg, [XT] Glossary
cost frontier model, [XT] xtfrontier
Cox, N. J., [XT] xtdescribe
cross-sectional
data, [XT] Glossary
time-series data, [XT] Glossary
Cui, J., [XT] xtgee
Cummings, P., [XT] xtpoisson

D
Daidone, S., [XT] xtfrontier
Das, S., [XT] xtunitroot
data,
autocorrelated, see autocorrelation
range of, see range of data
summarizing, see summarizing data
data manipulation, [XT] xtset
Davidson, R., [XT] xtgls, [XT] xtpcse
De Hoyos, R. E., [XT] xtreg
descriptive statistics, displaying, [XT] xtsum,
[XT] xttab
dichotomous outcome model, see outcomes, binary
DiNardo, J., [XT] xtec
dispersion, measures of, [XT] xtsum
disturbance term, [XT] Glossary
Drukker, D. M., [XT] xt
Dwyer, J. H., [XT] xtreg
dynamic model, [XT] Glossary
panel-data regression, [XT] xtabond, [XT] xtddp,
[XT] xtdpdsys

E
Eberhardt, M., [XT] xtec
EGLS, see estimated generalized least squares
endogenous
covariates, [XT] xtdp, [XT] xtdpdsys,
[XT] xthtaylor, [XT] xtivreg
variable, [XT] Glossary
error-components model, [XT] xthtaylor,
[XT] Glossary
estat
abond command, [XT] xtabond postestimation,
[XT] xtdpdsys postestimation
sargan command, [XT] xtabond postestimation,
[XT] xtdpdsys postestimation
wcorrelation command, [XT] xtgee postestimation
estimated generalized least squares, [XT] xtgls,
[XT] xtivreg, [XT] xtreg
exogenous variable, [XT] Glossary

F
feasible generalized least squares, [XT] xtgls,
[XT] xtivreg, [XT] xtreg
Feinleib, M., [XT] xtreg
FGLS, see feasible generalized least squares
first-differenced estimator, [XT] xtabond, [XT] xtdp,
[XT] xtdpdsys, [XT] xtivreg
Fisher, M. R., [XT] xtcloglog, [XT] xtgee,
[XT] xtintreg, [XT] xtlogit, [XT] xtologit,
[XT] xtprobit, [XT] xtprobit, [XT] xtobit
fisher, xtunitroot subcommand, [XT] xtunitroot
Fisher-type test, [XT] xtunitroot
fixed-effects model, [XT] xtabond, [XT] xtdpd,
[XT] xtdpdys, [XT] xtivreg, [XT] xtlogit,
[XT] xtnbreg, [XT] xtprobit, [XT] xtpoisson,
Flay, B. R., [XT] xtlogit, [XT] xtprobit
Ford, J. M., [XT] xfrontier
Frank, M. W., [XT] xtabond, [XT] xtdpd,
[XT] xtdpdys
Franzese, R. J., Jr., [XT] xtpcse
Franzini, L., [XT] xtrc
Frechette, G. R., [XT] xtprobit
frequency table, [XT] xtab
frontier model, see stochastic frontier model

G

Galanti, M. R., [XT] xtgee
Gange, S. J., [XT] xtclolog, [XT] xtgee,
[XT] xtintreg, [XT] xtlogit, [XT] xtlogit,
[XT] xtprobit, [XT] xtprobit, [XT] xtobit
Gavin, M. D., [XT] xtlogit, [XT] xtprobit
GEE, see generalized estimating equations
generalized
least squares,
estimated, see estimated generalized least squares
feasible, see feasible generalized least squares
linear models, [XT] xtgee, [XT] Glossary
method of moments, [XT] xtabond, [XT] xtdpd,
[XT] xtdpdys
Gleason, L. R., [XT] xtlogit, [XT] xtprobit
GLM, see generalized linear models
Glowacz, K. M., [XT] xtlogit, [XT] xtprobit
GMM, see generalized method of moments
Goldstein, R., [XT] xtrc
Gould, W. W., [XT] xfrontier
graphs, cross-sectional time-series data, [XT] xtdta,
[XT] xline
postestimation, [XT] xtpcse, [XT] xtpoiisson,
[XT] xtrc, [XT] xtreg
Griffiths, W. E., [XT] xtgls, [XT] xtpcse, [XT] xtrc,
[XT] xtreg
Griliches, Z., [XT] xtgls, [XT] xtnbreg, [XT] xtpcse,
[XT] xtpoiisson, [XT] xtrc
Gropper, D. M., [XT] xfrontier
Guilkey, D. K., [XT] xtprobit
Guimarães, P., [XT] xtnbreg
Gutiérrez, R. G., [XT] xt

H

Hadri, K., [XT] xtunitroot
Hadri Lagrange multiplier stationarity test,
[XT] xtunitroot
hadri, xtunitroot subcommand, [XT] xtunitroot
Hall, B. H., [XT] xtnbreg, [XT] xtpoisson
Hansen, L. P., [XT] xtabond, [XT] xtdpd,
[XT] xtdpdys
Hansen, W. B., [XT] xtlogit, [XT] xtprobit
Hardin, J. W., [XT] xtgee, [XT] xtpoisson
Harris, R. D. F., [XT] xtunitroot
Harris–Tzavalis test, [XT] xtunitroot
Hauck, W. W., [XT] xtclolog, [XT] xtlogit,
[XT] xtlogit, [XT] xtprobit, [XT] xtprobit,
[XT] xtreg
Hausman, J. A., [XT] xthtaylor, [XT] xtnbreg,
[XT] xtpoisson, [XT] xtreg postestimation
Hausman specification test, [XT] xtreg postestimation
Hausman–Taylor estimator, [XT] xthtaylor
Heston, A., [XT] xtunitroot
heteroskedasticity, [XT] xtgls
Hilbe, J. M., [XT] xtgee, [XT] xtpoisson
Hill, R. C., [XT] xtgls, [XT] xtpcse, [XT] xtrc,
[XT] xtreg
Hlouskova, J., [XT] xtunitroot
Hoechle, D., [XT] xtgls, [XT] xtpcse, [XT] xtreg,
[XT] xtrc
Holtz-Eakin, D., [XT] xtabond, [XT] xtdpd,
[XT] xtdpdys
Honoré, B. E., [XT] xtobit
Horton, N. J., [XT] xtgee
Hosmer, D. W., Jr., [XT] xtgee
Hsiao, C., [XT] xt, [XT] xtabond, [XT] xtdpd,
[XT] xtdpdys, [XT] xtivreg, [XT] xtreg,
ht, xtunitroot subcommand, [XT] xtunitroot
Huber/White/sandwich estimator of variance, see robust,
Huber/White/sandwich estimator of variance
idsyncratic error term, [XT] Glossary
Ilardi, G., [XT] xfrontier
Im, K. S., [XT] xtunitroot
Im–Pesaran–Shin test, [XT] xtunitroot
incidence-rate ratio, [XT] xtgee, [XT] xtnbreg,
[XT] xtpoisson
instrumental-variables, [XT] Glossary
instrumental-variables estimator, [XT] Glossary
regression, [XT] xtabond, [XT] xtdpd,
[XT] xtdpdys, [XT] xthtaylor, [XT] xtivreg,
[XT] xtivreg postestimation
interval data, [XT] xtintreg, [XT] Glossary
interval regression, random-effects, [XT] xtintreg
ipw, xtunitroot subcommand, [XT] xtunitroot
IRR, see incidence-rate ratio

J

jackknife standard errors, [XT] vce_options
Johnson, C. A., [XT] xtlogit, [XT] xtprobit
Johnston, J., [XT] xtrc
Judge, G. G., [XT] xtgls, [XT] xtpcse, [XT] xtrc,
[XT] xtreg
K


Khanti-Akom, S., [XT] xtph

Kiviet, J. F., [XT] xtabond

Klein, M., [XT] xtgee

Kleinbaum, D. G., [XT] xtreg

Kiviet, J. F., [XT] xtabond

Klein, M., [XT] xtgee

Kmenta, J., [XT] xtgee

Kleinbaum, D. G., [XT] xtunitroot

Kiviet, J. F., [XT] xtgee

Khanti-Akom, S., [XT] xtgls

Katz, J. N., [XT] xtpcse


Khanti-Akom, S., [XT] xtph

Kiviet, J. F., [XT] xtabond

Klein, M., [XT] xtgee

Kmenta, J., [XT] xtgee

Kleinbaum, D. G., [XT] xtunitroot

Kiviet, J. F., [XT] xtgee

Khanti-Akom, S., [XT] xtgls

Katz, J. N., [XT] xtpcse


Khanti-Akom, S., [XT] xtph

Kiviet, J. F., [XT] xtabond

Klein, M., [XT] xtgee

Kmenta, J., [XT] xtpcse

Kumbhakar, S. C., [XT] xtunitroot

Kwiatkowski, D., [XT] xtunitroot

L

Lahiri, K., [XT] xtgls


Lee, L. F., [XT] xtreg


Lemeshow, S. A., [XT] xtgee

Leonard, M., [XT] xtgee

Levin, A., [XT] xtunitroot

Levin–Lin–Chu test, [XT] xtunitroot

Li, Q., [XT] xtivreg, [XT] xtrg postestimation, [XT] xtrg


Libois, F., [XT] xtreg


Lin, C.-F., [XT] xtunitroot


link function, [XT] xtgee, [XT] xtpoisson

x1, xtunitroot subcommand, [XT] xtunitroot

Lockwood, J. R., [XT] xtreg


Lovell, C. A. K., [XT] xtfrontier


M


MaCurdy, T. E., [XT] xtph

Maddala, G. S., [XT] xtgls, [XT] xtunitroot

Marcellino, M., [XT] xtunitroot

McCaffrey, D. F., [XT] xtrc


means, displaying, [XT] xtsum

Meen, W., [XT] xtfrontier

Merryman, S., [XT] xtunitroot

Mihaly, K., [XT] xtreg

model specification test, see specification test

Molenberghs, G., [XT] xtrc postestimation

Moon, H. R., [XT] xtunitroot

Mundlak, Y., [XT] xtrc, [XT] xtrg, [XT] xtnbreg

Murphy, J. L., [XT] xtpoisson

N

Narendranathan, W., [XT] xtrc


negative binomial regression, [XT] xtpoisson


Newton, H. J., [XT] xtgee


Nichols, A., [XT] xtrc, [XT] xtreg


O

O’Connell, P. G. J., [XT] xtunitroot

O’Donnell, C. J., [XT] xtfrontier

observed information matrix, [XT] vce_options

Obstfeld, M., [XT] xtunitroot


Oh, K.-Y., [XT] xtunitroot

OIM, see observed information matrix

OPG, see outer product of the gradient

ordinal outcome model, see outcomes, ordinal

Orsini, N., [XT] xtreg

Osbat, C., [XT] xtunitroot
outcomes,
  binary,
    complementary log-log, [XT] xtcloglog
  generalized estimating equations, [XT] xtgee
    logistic, [XT] xtlogit
    probit, [XT] xtprobit
  count,
    generalized estimating equations, [XT] xtgee
    negative binomial, [XT] xtnbreg
    Poisson, [XT] xtpoisson
  ordinal,
    logistic, [XT] xtologit
    probit, [XT] xtprobit
outer product of the gradient, [XT] vce_options

Over, M., [XT] xtdpdsys postestimation
overidentifying restrictions, [XT] Glossary
tests of, [XT] xtabond, [XT] xtabond
  postestimation, [XT] xtintreg, [XT] xtdp, [XT] xtdpdsys,
  [XT] xtdpdsys postestimation

P
  Pagan, A. R., [XT] xtreg postestimation
  Palta, M., [XT] xtcloglog, [XT] xtgee, [XT] xtintreg,
  [XT] xtlogit, [XT] xtologit, [XT] xtologit,
  [XT] xtprobit, [XT] xtprobit
  panel data, [XT] Glossary
  panel-corrected standard error, [XT] xtpcse,
    [XT] Glossary
  panels, variable identifying, [XT] xtset
  Patterson, K., [XT] xtnitroot
  PCSE, see panel-corrected standard error
  Pendergast, J. F., [XT] xtcloglog, [XT] xtgee,
  [XT] xtintreg, [XT] xtlogit, [XT] xtologit,
  [XT] xtprobit, [XT] xtprobit, [XT] xtrobust
  Pesaran, M. H., [XT] xtnitroot
  Phillips, P. C. B., [XT] xtnitroot
  Pickles, A., [XT] xtgee, [XT] xtreg
  Piblado, J. S., [XT] xtpcse
  Poi, B. P., [XT] xtpcse
  xtpcse
  Poisson regression
    fixed-effects, [XT] xtpoisson
    model, [XT] Glossary
    population-averaged, [XT] xtgee, [XT] xtpoisson
    random-effects, [XT] xtpoisson
  pooled estimator, [XT] Glossary
  population-averaged model, [XT] xtcloglog, [XT] xtgee,
    [XT] xtlogit, [XT] xtnbreg, [XT] xtpoisson,
  Prais–Winston regression, [XT] xtpcse
  predetermined variable, [XT] Glossary
  Prentice, R. L., [XT] xtgee
  prewhiten, [XT] Glossary
  probit regression,
    generalized estimating equations, [XT] xtgee
    population-averaged, [XT] xtgee, [XT] xtprobit
    random-effects, [XT] xtprobit, [XT] xtprobit

production
  frontier model, [XT] xtpcse
  function, [XT] Glossary

Q
  Qaqish, B., [XT] xtgee
  quadchk command, [XT] quadchk
  quadrature, [XT] Glossary
  Gauss–Hermite, [XT] quadchk
  qualitative dependent variables, [XT] xtcloglog,
    [XT] xtgee, [XT] xtlogit, [XT] xtologit,
    [XT] xtprobit, [XT] xtprobit
  random-coefficients
    linear regression, [XT] xtrc
    model, [XT] Glossary
  random-effects model, [XT] xtabond, [XT] xtlogit,
    [XT] xtintreg, [XT] xtdp, [XT] xtdpdsys,
    [XT] xtdpdsys postestimation
  range of data, [XT] xtsort
  Rao, D. S. P., [XT] xtreg
  Ratcliffe, S. J., [XT] xtreg
  regression (in generic sense),
    dummy variables, with, [XT] xtreg
    fixed-effects, [XT] xtreg
    instrumental variables, [XT] xtabond, [XT] xtdp,
      [XT] xtdpdsys, [XT] xtdpdsys postestimation
  random-effects, [XT] xtgee, [XT] xtreg
  Revankar, N. S., [XT] xtpcse
  robust standard errors, [XT] Glossary
  robust, Huber/White/sandwich estimator of variance,
    [XT] vce_options
  fixed-effects models,
    linear, [XT] xtreg
    Poisson, [XT] xtpoisson
  linear dynamic panel-data estimation, [XT] xtabond,
    [XT] xtdp, [XT] xtdpdsys
  population-averaged models, [XT] xtlogit
  complementary log-log, [XT] xtcloglog
  logit, [XT] xtlogit
  negative binomial, [XT] xtnbreg
  Poisson, [XT] xtpoisson
  probit, [XT] xtprobit
  random-effects model,
    complementary log-log, [XT] xtcloglog
    linear, [XT] xtreg
    logistic, [XT] xtlogit, [XT] xtologit
robust, Huber/White/sandwich estimator of variance, continued
   Poisson, [XT] xtpoisson
   probit, [XT] xtprobit, [XT] xtprobit
Rogoff, K., [XT] xtunitroot
Roodman, D., [XT] xtdpd, [XT] xtdpdsys
Rosen, H. S., [XT] xtreg
   postestimation
S
sandwich/Huber/White estimator of variance, see robust, Huber/White/sandwich estimator of variance
Sarafidis, V., [XT] xtreg
sargan, estat subcommand, [XT] xtabond
   postestimation, [XT] xtdpd postestimation,
   [XT] xtdpdsys postestimation
Sargan test, [XT] xtabond postestimation, [XT] xtdpd
   postestimation, [XT] xtdpdsys postestimation
Sass, T. R., [XT] xtreg
Schank, T., [XT] xtreg
Schmidt, P., [XT] xtfrontier, [XT] xtunitroot
Schunck, R., [XT] xtreg
semirobust standard errors, [XT] Glossary
   sequential limit theory, [XT] Glossary
Shin, Y., [XT] xtunitroot
Shults, J., [XT] xtgee
Skondal, A., [XT] xtcloglog, [XT] xtgrowth,
   [XT] xtgrowth, [XT] xtintreg, [XT] xtoologit,
   [XT] xtoologit, [XT] xtoologit,
   [XT] xtprobit, [XT] xtpoisson, [XT] xtprobit,
   [XT] xttobit
Smith, A. F. M., [XT] xtcloglog, [XT] xintreg,
   [XT] xtoologit, [XT] xtoologit, [XT] xtoologit,
   [XT] xtpoisson, [XT] xtprobit, [XT] xttobit
Sobol, D. F., [XT] xtoologit, [XT] xtoologit
Sosa-Escudero, W., [XT] xtreg, [XT] xtregr
   postestimation, [XT] xtregr
   specification test, [XT] xtregr postestimation
Spiegel, D. C., [XT] xtoologit, [XT] xtoologit
standard deviations, displaying, [XT] xtsum
standard errors,
   bootstrap, see bootstrap standard errors
   jackknife, see jackknife standard errors
   panel-corrected, see panel-corrected standard error
   robust, see robust, Huber/White/sandwich estimator
     of variance
   semirobust, see semirobust standard errors
Stewart, M. B., [XT] xtprobit
stochastic frontier model, [XT] xtfrontier
Stock, J. H., [XT] xtcloglog, [XT] xttobit,
   [XT] xtoologit, [XT] xtoologit, [XT] xtoologit,
   [XT] xtpoisson, [XT] xtprobit, [XT] xttobit
strongly balanced, [XT] Glossary
Sturdivant, R. X., [XT] xtgee
summarizing data, [XT] xtsum
Summers, R., [XT] xtunitroot
Sussman, S., [XT] xtoologit, [XT] xtoologit
Swamy, P. A. V. B., [XT] xtvreg, [XT] xtregr
T
   table, frequency, see frequency table
   Taub, A. J., [XT] xtreg
   Taylor, C., [XT] xtgee, [XT] xtreg
   Taylor, W. E., [XT] xthtaylor
   test,
     autocorrelation, see autocorrelation test
     Breitung, see Breitung test
     Breusch–Pagan Lagrange multiplier, see Breusch–
     Pagan Lagrange multiplier test
   test, continued
     Fisher-type, see Fisher-type test
     Hadri Lagrange multiplier, see Hadri Lagrange
     multiplier stationarity test
     Harris–Tzavalis, see Harris–Tzavalis test
     Hausman specification, see Hausman specification
     test
     Im–Pesaran–Shin, see Im–Pesaran–Shin test
     Levin–Lin–Chu, see Levin–Lin–Chu test
     model specification, see specification test
     overidentifying restrictions, see overidentifying
     restrictions, tests of
     quadrature, see quadrature
     unit-root, see unit-root test
tobit regression, random-effects, [XT] xttobit
Trivedi, P. K., [XT] xt, [XT] xtnbreg, [XT] xtpoisson
Twisk, J. W. R., [XT] xtgee, [XT] xtoologit,
   [XT] xtoologit, [XT] xtoologit, [XT] xttobit
   two-stage least squares, [XT] xthtaylor, [XT] xtvreg
Tzavalis, E., [XT] xtunitroot
U
Ulene, A. L., [XT] xtoologit, [XT] xtoologit
unbalanced data, [XT] Glossary
unit-root test, [XT] xtunitroot
Upward, R., [XT] xtreg
Ureta, M., [XT] xtreg
V
van den Broeck, J., [XT] xtfrontier
Varadharajan-Krishnakumar, J., [XT] xtvreg
variable identifying panels, [XT] xtsset
variance,
   displaying, [XT] xtsum
   estimators, [XT] vce options
   Huber/White/sandwich estimator, see robust,
   Huber/White/sandwich estimator of variance
vce() option, [XT] vce options
Verardi, V., [XT] xtreg
Verbeke, G., [XT] xtregr postestimation
Wagner, M., [XT] xtunitroot
wcorrelation, estat subcommand, [XT] xtgee postestimation
weakly balanced, [XT] Glossary
Wedderburn, R. W. M., [XT] xtgee
West, K. D., [XT] xtunitroot
White, H. L., Jr., [XT] xtivreg
white noise, [XT] Glossary
White/Huber/sandwich estimator of variance, see robust, Huber/White/sandwich estimator of variance
Whitehead, A., [XT] xtunitroot
Whitney-Saltiel, D. A., [XT] xtologit, [XT] xtoprobit
within-cell means and variances, [XT] xtsum
Wright, J. H., [XT] xthtaylor
Wu, P. X., [XT] xtregar
Wu, S., [XT] xtunitroot
xtabond command, [XT] xtabond, [XT] xtabond postestimation
xtcloglog command, [XT] quadchk, [XT] xtcloglog, [XT] xtcloglog postestimation
xtdata command, [XT] xtdescribe
xtddp command, [XT] xtdpd, [XT] xtdpdsys
xtddpdsys command, [XT] xtdpdsys, [XT] xtdpdsys postestimation
xtfrontier command, [XT] xtfrontier, [XT] xtfrontier postestimation
xtgee command, [XT] xtgee, [XT] xtgee postestimation
xtintreg command, [XT] quadchk, [XT] xintreg, [XT] xintreg postestimation
xtivreg command, [XT] xtivreg, [XT] xtivreg postestimation
xtline command, [XT] xline
xtlogit command, [XT] quadchk, [XT] xtlogit, [XT] xtlogit postestimation
xtnbreg command, [XT] xtnbreg, [XT] xtnbreg postestimation
xtologit command, [XT] xtologit, [XT] xtologit postestimation
xtprobit command, [XT] xtoprobit, [XT] xtoprobit postestimation
xtpce command, [XT] xtpce, [XT] xtpce postestimation
xtpoisson command, [XT] quadchk, [XT] xtpoisson, [XT] xtpoisson postestimation
xtprobit command, [XT] quadchk, [XT] xtprobit, [XT] xtprobit postestimation
xtreg command, [XT] xtreg, [XT] xtreg postestimation
xtregar command, [XT] xtregar, [XT] xtregar postestimation
xtset command, [XT] xtset
xtsum command, [XT] xtsum
xttab command, [XT] xttab
xttest0 command, [XT] xttest0 command, [XT] xtreg postestimation
xttobit command, [XT] quadchk, [XT] xttobit, [XT] xttobit postestimation
xttrans command, [XT] xttrans
xtunitroot command, [XT] xtunitroot
breitung command, [XT] xtunitroot
fisher command, [XT] xtunitroot
hadri command, [XT] xtunitroot
ht command, [XT] xtunitroot
ips command, [XT] xtunitroot
llc command, [XT] xtunitroot
Yogo, M., [XT] xthtaylor
Zellner, A., [XT] xtfrontier
Zhao, L. P., [XT] xtgee