xtheckman — Random-effects regression with sample selection

Description

xtheckman fits a random-effects linear regression model with endogenous sample selection.

Quick start

Random-effects model of $y$ on $x_1$ using xtset data, with selection indicated by binary variable
selected and predicted by $v_1$ and $x_1$

xtheckman $y$ $x_1$, select(selected = $v_1$ $x_1$)

As above, but constraining random effects to be independent

xtheckman $y$ $x_1$, select(selected = $v_1$ $x_1$) norecorrelation

As above, but omit random effects from selection model

xtheckman $y$ $x_1$, select(selected = $v_1$ $x_1$, nore)

Menu

Statistics > Longitudinal/panel data > Sample-selection models > Linear regression with sample selection (RE)
Syntax

```
xtheckman  depvar [ indepvars ] [ if ] [ in ],
          select ( depvar_s = varlist_s [, sel_options ] ) [ options ]
```

**options**

<table>
<thead>
<tr>
<th>Description</th>
<th>Model Options</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>select()</em></td>
<td>specify selection equation: dependent and independent variables; whether to have constant term and offset variable or include random effect</td>
</tr>
<tr>
<td>noconstant</td>
<td>suppress constant term</td>
</tr>
<tr>
<td>norecorrelation</td>
<td>constrain the random effects to be independent</td>
</tr>
<tr>
<td>offset(varname_o)</td>
<td>include varname_o in model with coefficient constrained to 1</td>
</tr>
<tr>
<td>constraints(numlist)</td>
<td>apply specified linear constraints</td>
</tr>
<tr>
<td>SE/Robust</td>
<td>vcetype may be oim, robust, cluster clustvar, opg, bootstrap, or jackknife</td>
</tr>
<tr>
<td>Reporting</td>
<td>level(#) set confidence level; default is level(95)</td>
</tr>
<tr>
<td>nocnsreport do not display constraints</td>
<td></td>
</tr>
<tr>
<td>display_options control columns and column formats, row spacing, line width, display of omitted variables and base and empty cells, and factor-variable labeling</td>
<td></td>
</tr>
<tr>
<td>Integration</td>
<td>intmethod(intmethod) integration method for random effects; intmethod may be mvaghermite (the default) or ghermite</td>
</tr>
<tr>
<td>intpoints(#) set the number of integration (quadrature) points for random-effects integration; default is intpoints(7)</td>
<td></td>
</tr>
<tr>
<td>Maximization</td>
<td>maximize_options control the maximization process; seldom used</td>
</tr>
<tr>
<td>collinear keep collinear variables</td>
<td></td>
</tr>
<tr>
<td>coeflegend display legend instead of statistics</td>
<td></td>
</tr>
</tbody>
</table>

**sel_options**

<table>
<thead>
<tr>
<th>Description</th>
<th>Model Options</th>
</tr>
</thead>
<tbody>
<tr>
<td>noconstant</td>
<td>suppress constant term</td>
</tr>
<tr>
<td>nore</td>
<td>do not include random effects in selection model</td>
</tr>
<tr>
<td>offset(varname_o)</td>
<td>include varname_o in model with coefficient constrained to 1</td>
</tr>
</tbody>
</table>

*select()* is required.

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

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

bootstrap, by, jackknife, and statsby are allowed. See [U] 11.1.10 Prefix commands.

collinear and coeflegend do not appear in the dialog box.

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

Model

```
select(depvar_s = varlist_s [, sel_options]) specifies a random-effects probit model for sample selection with varlist_s as the covariates for the selection model. When depvar_s = 1, the model’s dependent variable is treated as observed (selected); when depvar_s = 0, it is treated as unobserved (not selected). select() is required.
```

```
sel_options are the following:

noconstant suppresses the constant term (intercept) in the selection model.

nore specifies that a random effect not be included in the selection equation.

offset(varname_o) specifies that varname_o be included in the selection model with the coefficient constrained to 1.
```

```
nonconstant; see [R] Estimation options.
```

```
norecorrelation constrains the random effects in the outcome and selection equations to be independent.
```

```
offset(varname_o), constraints(numlist); 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, opg), 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.
```

Reporting

```
level(#), nocnsreport; see [R] Estimation options.
```

```
display_options: noci, nopvalues, noomitted, vsquish, noemptycells, baselevels, allbaselevels, nofvlabel, fwrap(#), fvwrapon(style), cformat(%fmt), pformat(%fmt), sformat(%fmt), and nolstretch; see [R] Estimation options.
```

Integration

```
intmethod(intmethod) and intpoints(#) control how the integration of random effects is numerically calculated.
```

```
intmethod() specifies the integration method. The default method is mean-variance adaptive Gauss–Hermite quadrature, intmethod(mvaghermite). We recommend this method. intmethod(ghermite) specifies that nonadaptive Gauss–Hermite quadrature be used. This method is less computationally intensive and less accurate. It is sometimes useful to try intmethod(ghermite) to get the model to converge and then perhaps use the results as initial values specified in option from when fitting the model using the more accurate intmethod(mvaghermite). See Methods and formulas for more details.
```

```
intpoints() sets the number of integration (quadrature) points used for integration of the random effects. The default is intpoints(7). Increasing the number increases accuracy but also increases computational time. Computational time is roughly proportional to the number specified. See Methods and formulas for more details.
```
**Random-effects regression with sample selection**

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

The default technique is `technique(bhhh 10 nr 2)`.

Setting the optimization type to `technique(bhhh)` resets the default `vcetype` to `vce(opg)`.

The following options are available with `xtheckman` but are not shown in the dialog box: `collinear, coeflegend`; see [R] `Estimation options`.

**Remarks and examples**

`xtheckman` fits a panel-data model with endogenous sample selection. Endogenous sample selection is sometimes called nonignorability of selection, missing not at random, or selection bias. Within-panel correlation is accounted for by using panel-level random effects.

The outcome of interest $y_{it}$ is modeled as

$$y_{it} = x_{it} \beta + \nu_{1i} + \epsilon_{1it}$$

where $x_{it}$ are the covariates modeling the outcome, $\nu_{1i}$ is the panel-level random effect, and $\epsilon_{1it}$ is the observation-level error.

We model the selection process for the outcome by

$$s_{it} = 1 (z_{it} \alpha + \nu_{2i} + \epsilon_{2it} > 0)$$

where $s_{it} = 1$ if we observe $y_{it}$ and 0 otherwise, $z_{it}$ are the covariates modeling selection, $\nu_{2i}$ is the panel-level random effect for selection, and $\epsilon_{2it}$ is the observation-level selection error.

The random effects $\nu_{1i}$ and $\nu_{2i}$ are bivariate normal with mean 0 and variance

$$\begin{bmatrix} \sigma^2_{1\nu} & \rho \sigma_{1\nu} \sigma_{2\nu} \\ \rho \sigma_{1\nu} \sigma_{2\nu} & \sigma^2_{2\nu} \end{bmatrix}$$

The observation-level errors $\epsilon_{1it}$ and $\epsilon_{2it}$ are bivariate normal, with mean 0 and variance

$$\begin{bmatrix} \sigma^2_{1} & \rho \sigma_{1} \\ \rho \sigma_{1} & \sigma^2_{1} \end{bmatrix}$$

These observation-level errors are independent of the random effects.

Using the Heckman estimator (`heckman`) for this model will provide inefficient estimates because it ignores the within-panel correlation. Instead, we use maximum likelihood to model both the selection and outcome equations and account for the panel structure of the data. This random-effects estimator is used by `xteregress` and was discussed in Rabe-Hesketh, Skrondal, and Pickles (2002). There is no parametric fixed-effects estimator for panel data from an endogenously selected sample. See Honoré, Kyriazidou, and Powell (2000) and Kyriazidou (1997) for semiparametric fixed-effects estimators of panel-data endogenous sample-selection models.
Example 1

Suppose that we wish to study the relationship between wage, job tenure, and age for college-educated adults. We have fictional data on 600 adults observed from 2013 to 2016. We use these data to model hourly wage as a function of age, age squared, and job tenure. However, an individual’s wage is observed only if he or she works, and not everyone was employed on the dates the data were collected. We are not interested in modeling only the subpopulation of individuals who were employed at the time. We are also interested in the relationship of job tenure and age with the wage an individual would have received if he or she had been employed.

We suspect that the unobserved factors that affect an individual’s wage are related to the unobserved factors that affect employment status. These unobserved factors could include person-level characteristics like ability and time-varying factors like an individual’s family situation. We suspect that we have an endogenously selected sample. We have data on the local job market conditions (market). This variable is used with age and tenure to model the employment status of an individual.

Before we can fit a random-effects model to our data, we need to specify the panel structure of the data using *xtset*. Our panel variable is *personid*, the identification code for the individual. The time variable is *year*, and it ranges from 2013 to 2016.

```
. use https://www.stata-press.com/data/r16/wagework
   (Wages for 20 to 77 year olds, 2013-2016)
. xtset personid year
    panel variable:  personid (strongly balanced)
    time variable:   year, 2013 to 2016
    delta: 1 unit
```
We are now ready to fit our model.

```
.xcheckman wage c.age#c.age tenure, select(working = c.age##c.age market)
(setting technique to bhhh)
Iteration 0:  log likelihood = -5384.5076
Iteration 1:  log likelihood = -5377.4625
Iteration 2:  log likelihood = -5376.4805
Iteration 3:  log likelihood = -5376.4505
Iteration 4:  log likelihood = -5376.4464
Iteration 5:  log likelihood = -5376.4454
Iteration 6:  log likelihood = -5376.4451
Iteration 7:  log likelihood = -5376.4451
Iteration 8:  log likelihood = -5376.4451
Iteration 9:  log likelihood = -5376.445
(switching technique to nr)
Iteration 10: log likelihood = -5376.445
```

Random-effects regression with selection

<table>
<thead>
<tr>
<th>Number of obs</th>
<th>Selected</th>
<th>Nonselected</th>
</tr>
</thead>
<tbody>
<tr>
<td>2,400</td>
<td>1,928</td>
<td>472</td>
</tr>
</tbody>
</table>

Group variable: personid

<table>
<thead>
<tr>
<th>Number of groups</th>
</tr>
</thead>
<tbody>
<tr>
<td>600</td>
</tr>
</tbody>
</table>

Obs. per group:

<table>
<thead>
<tr>
<th>min</th>
<th>avg</th>
<th>max</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>4.0</td>
<td>4</td>
</tr>
</tbody>
</table>

Integration method: mvaghermite

<table>
<thead>
<tr>
<th>Integration pts.</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
</tr>
</tbody>
</table>

Log likelihood = -5376.445

| Coef. | Std. Err. | z     | P>|z| | [95% Conf. Interval] |
|-------|-----------|-------|------|-----------------------|
| wage  | age       | .5722234 | .0477613 | 11.98 | 0.000 | .4786129 | .6658339 |
|       | c.age#c.age | -.0042448 | .0005329 | -7.97 | 0.000 | -.0052893 | -.0032003 |
| tenure| _cons     | .5927719 | .0169866 | 34.90 | 0.000 | .5594787 | .626065  |
| working | age      | .2305309 | .0207988 | 11.08 | 0.000 | .1897661 | .2712958 |
|        | c.age#c.age | -.0026832 | .0002241 | -11.97 | 0.000 | -.0031225 | -.0022439 |
|        | market    | .1894934 | .019038  | 9.95  | 0.000 | .1521796 | .2268072 |
|        | _cons     | -3.276904 | .4352836 | -7.53 | 0.000 | -4.130045 | -2.423764 |
| var(e.wage) |        | 4.458219 | .2235342 | 4.040939 | 4.918588 |
| corr(e.working, e.wage) | | .4091115 | .1391856 | 2.94  | 0.003 | .1065022 | .642359  |
| var( wage[person]) |        | 2.493737 | .2547628 | 2.041226 | 3.046564 |
| var( working[person]) | | .3831411 | .0830963 | 2.50466 | 5.860961 |
| corr( working[person], wage[person]) | | .6021096 | .0845675 | 7.12  | 0.000 | .4106863 | .7426953 |
The first two sections of the output provide the estimated coefficients for the wage equation and the selection (working) equation. We can interpret the coefficients in the wage equation using the standard linear regression interpretation. For example, we expect an increase of $0.59 per hour for an additional year of job tenure.

Next we see var(e.wage), an estimate of the variance of the observation-level error for wage; this is followed by corr(e.working,e.wage), an estimate of its correlation with the observation-level error for the selection model. The next section of the output reports estimates of the variances of the random effects, var(wage[personid]) and var(working[personid]). The last section reports an estimate of the correlation of these random effects. If at least one of the correlations is significantly different from zero, we can conclude that we have endogenous sample selection. In our case, the correlation between the observation-level errors is 0.41, and the correlation between the random effects is 0.60. Because both are positive and significantly different from zero, we conclude that we have endogenous selection and that unobserved individual-level factors that increase the chance of being employed tend to increase wage. Additionally, unobserved observation-level (time-varying) factors that increase the chance of being employed tend to increase wage.

We estimated coefficients for age and age squared. We can use margins and marginsplot to gain a clearer understanding of the effect of the individuals’ age on hourly wage. We use margins with at() to profile the expected wages for individuals between ages 30 and 70 and with 0 and 5 years of job tenure. Then, we use marginsplot to graph the estimates.

```
. margins, at(age=(30(5)70) tenure =(0 5))
. marginsplot
```

Based on this model, and assuming the data are from a random or otherwise representative sample, the plotted points represent the expected wage for individuals with the specified job tenure and age. We see that age has an increasing effect on expected wage until the mid-60s and then attenuates. Having 5 years of job tenure instead of none shifts the curve up by about $3.00 per hour.
Stored results

`xtheckman` stores the following in `e()`:

Scalars
- `e(N)` number of observations
- `e(N_g)` number of groups
- `e(N_selected)` number of selected observations
- `e(N_nonselected)` number of nonselected 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_aux)` number of auxiliary parameters
- `e(k_dv)` number of dependent variables
- `e(df_m)` model degrees of freedom
- `e(ll)` log likelihood
- `e(N_clust)` number of clusters
- `e(chi2)` \( \chi^2 \)
- `e(p)` \( p \)-value for model test
- `e(n_requad)` number of integration points for random effects
- `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(ic)` number of iterations
- `e(rc)` return code
- `e(converged)` 1 if converged, 0 otherwise

Macros
- `e(cmd)` `xtheckman`
- `e(cmdline)` command as typed
- `e(depvar)` names of dependent variables
- `e(ivar)` variable denoting groups
- `e(title)` title in estimation output
- `e(clustvar)` name of cluster variable
- `e(offset#)` offset for the \#th `depvar`, where \# is determined by equation order in output
- `e(chi2type)` Wald; type of model \( \chi^2 \) test
- `e(vcetype)` `vcetype` specified in `vce()`
- `e(vce)` title used to label Std. Err.
- `e(reintmethod)` integration method for random effects
- `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(estat_cmd)` program used to implement `estat`
- `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 (up to 20 iterations)
- `e(gradient)` gradient vector
- `e(V)` variance–covariance matrix of the estimators
- `e(V_modelbased)` model-based variance

Functions
- `e(sample)` marks estimation sample
In addition to the above, the following is stored in `r()`:

Matrices

\( r(table) \)  
matrix containing the coefficients with their standard errors, test statistics, \( p \)-values, and confidence intervals

Note that results stored in `r()` are updated when the command is replayed and will be replaced when any r-class command is run after the estimation command.

### Methods and formulas

`xtheckman` fits a random-effects linear regression model with endogenous sample selection via maximum likelihood estimation. For details on maximum likelihood estimators, see the results in Wooldridge (2010, chap. 13) and White (1996).

The log-likelihood function maximized by `xtheckman` is implied by the triangular structure of the model. Specifically, the joint distribution of the outcome and selection variables is a product of conditional and marginal distributions because the model is triangular. For a few of the many relevant applications of this result in literature, see Amemiya (1985, chap. 10); Heckman (1976, 1979); Maddala (1983, chap. 5); Maddala and Lee (1976); Wooldridge (2010, sec. 15.7.2, 15.7.3, 16.3.3, 17.5.2, and 19.7.1; 2014). Roodman (2011) and Bartus and Roodman (2014) used this result to derive the formulas discussed below.

We have panels \( i = 1, \ldots, N \) and observations \( t = 1, \ldots, N_i \). We model \( y_{it} \) as

\[
y_{it} = x_{it} \beta + \nu_{1i} + \epsilon_{1it}
\]

where \( x_{it} \) are the outcome covariates, \( \nu_{1i} \) is the panel-level random effect, and \( \epsilon_{1it} \) is the observation-level error.

The selection process for the outcome is modeled by

\[
s_{it} = 1 \left( z_{it} \alpha + \nu_{2i} + \epsilon_{2it} > 0 \right)
\]

where \( s_{it} = 1 \) if we observe \( y_{it} \) and 0 otherwise, \( z_{it} \) are the selection covariates, \( \nu_{2i} \) is the panel-level random effect for selection, and \( \epsilon_{2it} \) is the observation-level selection error.

The random effects \( \nu_{1i} \) and \( \nu_{2i} \) are bivariate normal with mean 0 and variance

\[
\Sigma_{\nu} = \begin{bmatrix}
\sigma_{1\nu}^2 & \rho_{\nu} \sigma_{1\nu} \sigma_{2\nu} \\
\rho_{\nu} \sigma_{1\nu} \sigma_{2\nu} & \sigma_{2\nu}^2
\end{bmatrix}
\]

The observation-level errors \( \epsilon_{1it} \) and \( \epsilon_{2it} \) are independent of the random effects and are also bivariate normal, with mean 0 and variance

\[
\Sigma = \begin{bmatrix}
\sigma_{1}^2 & \rho \sigma_{1} \\
\rho \sigma_{1} & 1
\end{bmatrix}
\]

When we condition on the random effects \( \nu_{1i} \) and \( \nu_{2i} \), we can write the joint density of the \( y_{it} \) and \( s_{it} \) using the conditional density of the selection error \( \epsilon_{2it} \) on the outcome error \( \epsilon_{1it} \).

For the selection indicator \( s_{i} \), we have lower limit \( l_{it} \) and upper limit \( u_{it} \),

\[
l_{it} = \begin{cases}
-\infty & s_{it} = 0 \\
-z_{it} \alpha - \nu_{2i} - \frac{\rho}{\sigma_{1}} (y_{it} - x_{it} \beta - \nu_{1i}) & s_{it} = 1
\end{cases}
\]

\[
u_{it} = \begin{cases}
-\infty & s_{it} = 0 \\
-z_{it} \alpha - \nu_{2i} & s_{it} = 1
\end{cases}
\]

\[
l_{it} = \begin{cases}
-\infty & s_{it} = 0 \\
-z_{it} \alpha - \nu_{2i} & s_{it} = 1
\end{cases}
\]
Then, the joint density of \( y_{it} \) and \( s_{it} \) conditional on the random effects is

\[
f(y_{it}, s_{it} \mid \nu_{1i}, \nu_{2i}) = \begin{cases} 
\Phi \left( \frac{-l_{it}}{\sqrt{1-\rho^2}} \right) + \phi \left( \frac{y_{it} - x_{it} \beta}{\sigma_1} \right) & s_{it} = 1 \\
\Phi \left( \frac{u_{it}}{\sqrt{1-\rho^2}} \right) & s_{it} = 0
\end{cases}
\]

Note that each panel has the same random effects for each observation. So the likelihood for panel \( i \) is

\[
L_i = \int_{\mathbb{R}^2} \prod_{t=1}^{N_i} f(y_{it}, s_{it} \mid \nu_{1i}, \nu_{2i}) \phi_2 \left( \left( \nu_{1i}, \nu_{2i} \right) \right) d\nu_{1i} d\nu_{2i} \tag{1}
\]

This multivariate integral is generally not tractable. We can use a change-of-variables technique to transform it into a set of nested univariate integrals. Let \( L \) be the Cholesky decomposition of \( \Sigma_\nu \); that is, \( \Sigma_\nu = LL' \). It follows that \( (\nu_{1i}, \nu_{2i})' = L \psi_i \), where \( \psi_i \) is a vector of independent standard normal random variables.

We can rewrite (1) as

\[
L_i = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \prod_{t=1}^{N_i} f(y_{it}, s_{it} \mid (\nu_{1i}, \nu_{2i})' = L \psi_i) \phi(\psi_{1i}) \phi(\psi_{2i}) d\psi_{1i} d\psi_{2i}
\]

Now the univariate integral can be approximated using Gauss–Hermite quadrature (GHQ). For \( q \)-point GHQ, let the abscissa and weight pairs be denoted by \( (a_k^*, w_k^*) \), where \( k = 1, \ldots, q \). The GHQ approximation is then

\[
\int_{-\infty}^{\infty} f(x) \exp(-x^2) \, dx \approx \sum_{k=1}^{q} w_k^* f(a_k^*)
\]

Consider a 2-dimensional quadrature grid containing \( q \) quadrature points in both dimensions. Let the vector of abscissas \( a_k = (a_{k1}, a_{k2})' \) be a point in this grid, and let \( w_k = (w_{k1}, w_{k2})' \) be the vector of corresponding weights.

The GHQ approximation to the likelihood for a given panel is

\[
L_i = \sum_{k_1=1}^{q} \sum_{k_2=1}^{q} \left\{ \prod_{t=1}^{N_i} f(y_{it}, s_{it} \mid (\nu_{1i}, \nu_{2i})' = L a_k) \right\} \left\{ \prod_{s=1}^{2} w_{ks} \right\}
\]

Rather than using regular GHQ, we can use mean–variance adaptive GHQ. Fixing the observed variables and model parameters in the integrand of (1), we see the posterior density for \( \psi_i \) is proportional to

\[
\left\{ \prod_{t=1}^{N_i} f(y_{it}, s_{it} \mid (\nu_{1i}, \nu_{2i})' = L \psi_i) \right\} \phi(\psi_i)
\]
It is reasonable to assume that this posterior density can be approximated by a multivariate normal density with mean vector $\mu_{vi}$ and variance matrix $\tau_{vi}$. Instead of using the prior density of $\psi_i$ as the weighting distribution in the integral, we can use our approximation for the posterior density,

$$L_i = \int_{\mathbb{R}^2} \left\{ \prod_{t=1}^{N_i} f(y_{it}, s_{it} | (\nu_{1i}, \nu_{2i})' = L\psi_i) \right\} \frac{\phi(\psi_i)}{\phi(\psi_i, \mu_{vi}, \tau_{vi})} d\psi_i$$

The likelihood is then approximated by

$$L_i = \sum_{k_1=1}^{q} \cdots \sum_{k_2=1}^{q} \left\{ \prod_{t=1}^{N_i} f(y_{it}, s_{it} | (\nu_{1i}, \nu_{2i})' = L\alpha_k) \right\} \left\{ \prod_{s=1}^{2} \omega_{ks} \right\}$$

where $\alpha_k$ and $\omega_{ks}$ are the adaptive versions of the abscissas and weights after an orthogonalizing transformation, which eliminates posterior covariances between elements of $\psi_i$. The posterior means $\mu_{vi}$ and posterior variances $\tau_{vi}$ are computed iteratively by updating the posterior moments by using the mean–variance adaptive GHQ approximation, starting with a 0 mean vector and identity variance matrix.

The log likelihood for all panels is then

$$\ln L = \sum_{i=1}^{N} \left( \ln \sum_{k_1=1}^{q} \cdots \sum_{k_2=1}^{q} \left\{ \prod_{t=1}^{N_i} f(y_{it}, s_{it} | (\nu_{1i}, \nu_{2i})' = L\alpha_k) \right\} \left\{ \prod_{s=1}^{2} \omega_{ks} \right\} \right)$$

The conditional mean of $y_{it}$ is

$$E(y_{it} | x_{it}) = x_{it} \beta$$

xtheckman results are obtained using xteregress; see Methods and formulas of [ERM] eregress.

References


Heckman, J. 1976. The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. Annals of Economic and Social Measurement 5: 475–492.


Also see

[XT] xheckman postestimation — Postestimation tools for xheckman

[XT] xteregress — Extended random-effects linear regression

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

[XT] xtset — Declare data to be panel data

[ERM] eregress — Extended linear regression

[R] heckman — Heckman selection model

[R] regress — Linear regression

[U] 20 Estimation and postestimation commands