xtoprobit — Random-effects ordered probit models

Description

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.

Quick start

Random-effects ordered probit model of y as a function of x using xtset data
   xtoprobit y x

Add indicators for levels of categorical variable a
   xtoprobit y x i.a

With cluster–robust standard errors for panels nested within cvar
   xtoprobit y x i.a, vce(cluster cvar)

Menu

Statistics > Longitudinal/panel data > Ordinal outcomes > Probit regression (RE)
xtoprobit — Random-effects ordered probit models

Syntax

```
xtoprobit depvar [ indepvars ] [ if ] [ in ] [ weight ] [ , options ]
```

Options

```
Model
offset(varname)
constraints(constraints)
SE/Robust
vce(vcetype)
Reporting
level(#)
lnmodel
nocnsreport
display_options
Integration
intmethod(intmethod)
intpoints(#)
Maximization
maximize_options
startgrid(numlist)
nodeplay
collinear
coeflegend
```

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

**Options**

- **Model**
  - `offset(varname)`: include `varname` in model with coefficient constrained to 1
  - `constraints(constraints)`: apply specified linear constraints
  - `vce(vcetype)`: `vcetype` may be `oim`, `robust`, `cluster clustvar`, `bootstrap`, or `jackknife`

- **Reporting**
  - `level(#)`: set confidence level; default is `level(95)`
  - `lnmodel`: perform the likelihood-ratio model test instead of the default Wald test
  - `nocnsreport`: do not display constraints
  - `display_options`: control columns and 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
  - `nodeplay`: suppress display of header and coefficients
  - `collinear`: keep collinear variables
  - `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.

bayes, by, collect, fp, and statsby are allowed; see [U] 11.1.10 Prefix commands. For more details, see [BAYES] bayes: xtoprobit.

fweights, iweights, and pweights are allowed; see [U] 11.1.6 weight.

startgrid(), nodeplay, 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**
  - `offset(varname)`, `constraints(constraints)`; see [R] Estimation options.
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 and the robust VCE estimator in Methods and formulas.

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

lrmodel, nocnsreport; see [R] Estimation options.

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

intmethod(intmethod), intpoints(#); 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 xtprobit 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, xtprobit 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 xtprobit displaying Grid node 1, Grid node 2, ... following Grid node 0 in the iteration log, that is xtprobit 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.

collinear, coeflegend; see [R] Estimation options.

Remarks and examples

Remarks are presented under the following headings:

   Overview
   Video example

Overview

xtprobit 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] meoprobit 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 https://www.stata-press.com/data/r17/tvsfpors
(Television, School, and Family Project)
. xtset school
Panel variable: school (unbalanced)
. xtprobit thk prethk cc##tv
Fitting comparison model:
Iteration 0: log likelihood = -2212.7755
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 = 1,600
Group variable: school Number of groups = 28
Random effects u_i ~ GaussianObs per group:
min = 18
avg = 57.1
max = 137
Integration method: mvaghermiteIntegration pts. = 12
Wald chi2(4) = 128.05
Log likelihood = -2121.7715 Prob > chi2 = 0.0000

| thk          | Coefficient | Std. err. | z     | P>|z| | [95% conf. interval] |
|--------------|-------------|-----------|-------|-----|----------------------|
| prethk       | .2369804    | .0227739  | 10.41 | 0.000 | .1923444 .2816164 |
| 1.cc         | .5490957    | .1255108  | 4.37  | 0.000 | .303099 .7950923   |
| 1.tv         | .1695405    | .1215889  | 1.39  | 0.163 | -.0687693 .4078504 |
| cc##tv 1 1   | -.2951837   | .1751969  | -1.68 | 0.092 | -.6385634 .0481959 |
| /cut1        | -.0682011   | .1003374  |       |      | -.2648587 .1284565 |
| /cut2        | .67681      | .1008836  |       |      | .4790817 .8745382 |
| /cut3        | 1.390649    | .1037494  |       |      | 1.187304 1.593995  |
| /sigma2_u    | .0288527    | .0146201  |       |      | .0106874 .0778937 |

LR test vs. oprobit model: 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 \(\sigma_u^2\). The parameter estimates can be interpreted just as the output from a standard ordered probit regression would be interpreted; see \texttt{[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 \(\sigma_u^2\) 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.

\section*{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 \texttt{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 \texttt{intpoints()} option and run \texttt{quadchk} again. Do not attempt to interpret the results of estimates when the coefficients reported by \texttt{quadchk} differ substantially. See \texttt{[XT] quadchk} for details and \texttt{[XT] xtprobit} for an example.

Because the \texttt{xtprobit} 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.

\section*{Video example}

Ordered logistic and probit for panel data

\section*{Stored results}

\texttt{xtprobit} stores the following in \texttt{e()}:

**Scalars**

- \texttt{e(N)}: number of observations
- \texttt{e(N_g)}: number of groups
- \texttt{e(k)}: number of parameters
- \texttt{e(k_aux)}: number of auxiliary parameters
- \texttt{e(k_eq)}: number of equations in \texttt{e(b)}
- \texttt{e(k_eq_model)}: number of equations in overall model test
- \texttt{e(k_dv)}: number of dependent variables
- \texttt{e(k_cat)}: number of categories
- \texttt{e(df_m)}: model degrees of freedom
- \texttt{e(ll)}: log likelihood
- \texttt{e(ll_0)}: log likelihood, constant-only model
- \texttt{e(ll_c)}: log likelihood, comparison model
- \texttt{e(chi2)}: \(\chi^2\)
- \texttt{e(chi2_c)}: \(\chi^2\) for comparison test
- \texttt{e(N_clust)}: number of clusters
- \texttt{e(sigma_u)}: panel-level standard deviation
- \texttt{e(n_quad)}: number of quadrature points
- \texttt{e(g_min)}: smallest group size
Macros
\[
e(g_{\text{avg}}) \quad \text{average group size}
\]
\[
e(g_{\text{max}}) \quad \text{largest group size}
\]
\[
e(p) \quad p\text{-value for model test}
\]
\[
e(\text{rank}) \quad \text{rank of } e(V)
\]
\[
e(\text{rank0}) \quad \text{rank of } e(V) \text{ for constant-only model}
\]
\[
e(ic) \quad \text{number of iterations}
\]
\[
e(rc) \quad \text{return code}
\]
\[
e(\text{converged}) \quad 1 \text{ if converged, } 0 \text{ otherwise}
\]

Macros
\[
e(cmd) \quad \text{meglm}
\]
\[
e(cmd2) \quad \text{xtoprobit}
\]
\[
e(cmdline) \quad \text{command as typed}
\]
\[
e(depvar) \quad \text{name of dependent variable}
\]
\[
e(covariates) \quad \text{list of covariates}
\]
\[
e(ivar) \quad \text{variable denoting groups}
\]
\[
e(wtype) \quad \text{weight type}
\]
\[
e(wexp) \quad \text{weight expression}
\]
\[
e(title) \quad \text{title in estimation output}
\]
\[
e(clustvar) \quad \text{name of cluster variable}
\]
\[
e(offset) \quad \text{linear offset variable}
\]
\[
e(chi2type) \quad \text{Wald or LR; type of model } \chi^2 \text{ test}
\]
\[
e(vcetype) \quad \text{vcetype specified in } \text{vce()}
\]
\[
e(intmethod) \quad \text{integration method}
\]
\[
e(distrib) \quad \text{Gaussian; the distribution of the random effect}
\]
\[
e(opt) \quad \text{type of optimization}
\]
\[
e(which) \quad \text{max or min; whether optimizer is to perform maximization or minimization}
\]
\[
e(ml\_method) \quad \text{type of } ml \text{ method}
\]
\[
e(user) \quad \text{name of likelihood-evaluator program}
\]
\[
e(technique) \quad \text{maximization technique}
\]
\[
e(properties) \quad b \ V
\]
\[
e(predict) \quad \text{program used to implement predict}
\]
\[
e(marginsok) \quad \text{predictions allowed by margins}
\]
\[
e(marginswtype) \quad \text{weight type for margins}
\]
\[
e(marginswexp) \quad \text{weight expression for margins}
\]
\[
e(marginsdefault) \quad \text{default predict() specification for margins}
\]
\[
e(asbalanced) \quad \text{factor variables fvset as asbalanced}
\]
\[
e(asobserved) \quad \text{factor variables fvset as asobserved}
\]

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

Functions
\[
e(sample) \quad \text{marks estimation sample}
\]

In addition to the above, the following is stored in \textit{r()}: 

Matrices
\[
r(table) \quad \text{matrix containing the coefficients with their standard errors, test statistics, } p\text{-values, and confidence intervals}
\]

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

\[ p_{itk} \equiv \Pr(y_{it} = k | \kappa, x_{it}, \nu_i) = \Pr(\kappa_{k-1} < x_{it}\beta + \nu_i + \epsilon_{it} \leq \kappa_k) = \Phi(\kappa_k - x_{it}\beta - \nu_i) - \Phi(\kappa_{k-1} - x_{it}\beta - \nu_i) \]

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

\[ f(y_{it}, \kappa, x_{it}\beta + \nu_i) = \prod_{k=1}^{K} p_{itk} \]

\[ = \text{exp} \sum_{k=1}^{K}\left\{ I_k(y_{it}) \log(p_{itk}) \right\} \]

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

\[ l_i(\beta, \kappa, \sigma^2_\nu) = \int_{-\infty}^{\infty} e^{-\nu_i^2/2\sigma^2_\nu} \left\{ \prod_{t=1}^{n_i} f(y_{it}, \kappa, x_{it}\beta + \nu_i) \right\} d\nu_i \]

\[ \equiv \int_{-\infty}^{\infty} g(y_{it}, \kappa, x_{it}, \nu_i) d\nu_i \]
This integral can be approximated with $M$-point Gauss–Hermite quadrature

$$
\int_{-\infty}^{\infty} e^{-x^2} h(x) \, dx \approx \sum_{m=1}^{M} w_m h(a_m^*)
$$

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 mean–variance adaptive Gauss–Hermite quadrature, which approximates the panel-level likelihood with

$$
L_i \approx \sum_{m=1}^{M} w_m^* \exp \left\{ (a_m^*)^2 \right\} g(y_{it}, \kappa, x_{it}, \sqrt{2} \hat{\sigma}_i a_m^* + \hat{\mu}_i)
$$

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

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

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

and

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

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 $\rho = \sigma_0^2 / (\sigma_0^2 + 1)$:

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

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

### xtoprobit 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 (2020) and Arellano (2003) discuss this application of the Huber/White/sandwich VCE estimator. As discussed by Wooldridge (2020), 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] `xtoprobit postestimation` — Postestimation tools for xtoprobit

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

[XT] `xtoprobit` — Extended random-effects ordered probit regression

[XT] `xtologit` — Random-effects ordered logistic models

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

[BAYES] bayes: `xtoprobit` — Bayesian random-effects ordered probit model

[ME] `meoprobit` — Multilevel mixed-effects ordered probit regression

[R] `probit` — Probit regression

[U] 20 Estimation and postestimation commands