

**eprobit** — Extended probit regression

<a href="#">Description</a>	<a href="#">Quick start</a>	<a href="#">Menu</a>	<a href="#">Syntax</a>
<a href="#">Options</a>	<a href="#">Remarks and examples</a>	<a href="#">Stored results</a>	<a href="#">Methods and formulas</a>
<a href="#">References</a>	<a href="#">Also see</a>		

## Description

`eprobit` fits a probit regression model that accommodates any combination of endogenous covariates, nonrandom treatment assignment, and endogenous sample selection. Continuous, binary, and ordinal endogenous covariates are allowed. Treatment assignment may be endogenous or exogenous. A probit or tobit model may be used to account for endogenous sample selection.

`xteprobit` fits a random-effects probit regression model that accommodates endogenous covariates, treatment, and sample selection in the same way as `eprobit` and also accounts for correlation of observations within panels or within groups.

## Quick start

Probit regression of  $y$  on  $x$  with continuous endogenous covariate  $y_2$  modeled by  $x$  and  $z$

```
eprobit y x, endogenous(y2 = x z)
```

Same as above, but adding continuous endogenous covariate  $y_3$  modeled by  $x$  and  $z_2$

```
eprobit y x, endogenous(y2 = x z) endogenous(y3 = x z2)
```

Probit regression of  $y$  on  $x$  with binary endogenous covariate  $d$  modeled by  $x$  and  $z$

```
eprobit y x, endogenous(d = x z, probit)
```

Probit regression of  $y$  on  $x$  with endogenous treatment recorded in `trtvar` and modeled by  $x$  and  $z$

```
eprobit y x, entreat(trtvar = x z)
```

Probit regression of  $y$  on  $x$  with exogenous treatment recorded in `trtvar`

```
eprobit y x, extreat(trtvar)
```

Random-effects probit regression of  $y$  on  $x$  using `xtset` data

```
xteprobit y x
```

Probit regression of  $y$  on  $x$  with endogenous sample-selection indicator `selvar` modeled by  $x$  and  $z$

```
eprobit y x, select(selvar = x z)
```

Same as above, but adding endogenous covariate  $y_2$  modeled by  $x$  and  $z_2$

```
eprobit y x, select(selvar = x z) endogenous(y2 = x z2)
```

Same as above, but adding endogenous treatment recorded in `trtvar` and modeled by  $x$  and  $z_3$

```
eprobit y x, select(selvar = x z) endogenous(y2 = x z2) ///
    entreat(trtvar = x z3)
```

Same as above, but with random effects and without endogenous treatment

```
xteprobit y x, select(selvar = x z) endogenous(y2 = x z2)
```

## Menu

### **eprobit**

Statistics > Endogenous covariates > Models adding selection and treatment > Probit regression

### **xteprobit**

Statistics > Longitudinal/panel data > Endogenous covariates > Models adding selection and treatment > Probit regression (RE)

## Syntax

*Basic probit regression with endogenous covariates*

```
eprobit depvar [indepvars] , endogenous(depvarsen = varlisten) [options]
```

*Basic probit regression with endogenous treatment assignment*

```
eprobit depvar [indepvars] , entreat(depvartr [= varlisttr]) [options]
```

*Basic probit regression with exogenous treatment assignment*

```
eprobit depvar [indepvars] , extreat(tvar) [options]
```

*Basic probit regression with sample selection*

```
eprobit depvar [indepvars] , select(depvars = varlists) [options]
```

*Basic probit regression with tobit sample selection*

```
eprobit depvar [indepvars] , tobitselect(depvars = varlists) [options]
```

*Basic probit regression with random effects*

```
xteprobit depvar [indepvars] [, options]
```

*Probit regression combining endogenous covariates, treatment, and selection*

```
eprobit depvar [indepvars] [if] [in] [weight] [, extensions options]
```

*Probit regression combining random effects, endogenous covariates, treatment, and selection*

```
xteprobit depvar [indepvars] [if] [in] [, extensions options]
```

<i>extensions</i>	Description
Model	
<u>endogenous</u> ( <i>enspec</i> )	model for endogenous covariates; may be repeated
<u>entreat</u> ( <i>entrspec</i> )	model for endogenous treatment assignment
<u>extreat</u> ( <i>extrspec</i> )	exogenous treatment
<u>select</u> ( <i>selspec</i> )	probit model for selection
<u>tobitselect</u> ( <i>tselspec</i> )	tobit model for selection
<i>options</i>	Description
Model	
<u>noconstant</u>	suppress constant term
<u>offset</u> ( <i>varname<sub>o</sub></i> )	include <i>varname<sub>o</sub></i> in model with coefficient constrained to 1
<u>constraints</u> ( <i>numlist</i> )	apply specified linear constraints
SE/Robust	
<u>vce</u> ( <i>vcetype</i> )	<i>vcetype</i> may be <u>oim</u> , <u>robust</u> , <u>cluster</u> <i>clustvar</i> , <u>opg</u> , <u>bootstrap</u> , or <u>jackknife</u>
Reporting	
<u>level</u> (#)	set confidence level; default is <u>level</u> (95)
<u>nocnsreport</u>	do not display constraints
<u>display_options</u>	control columns and column formats, row spacing, line width, display of omitted variables and base and empty cells, and factor-variable labeling
Integration	
<u>intpoints</u> (#)	set the number of integration (quadrature) points for integration over four or more dimensions; default is <u>intpoints</u> (128)
<u>triintpoints</u> (#)	set the number of integration (quadrature) points for integration over three dimensions; default is <u>triintpoints</u> (10)
<u>reintpoints</u> (#)	set the number of integration (quadrature) points for random-effects integration; default is <u>reintpoints</u> (7)
<u>reintmethod</u> ( <i>intmethod</i> )	integration method for random effects; <i>intmethod</i> may be <u>mvaghermite</u> (the default) or <u>ghermite</u>
Maximization	
<u>maximize_options</u>	control the maximization process; seldom used
<u>collinear</u>	keep collinear variables
<u>coeflegend</u>	display legend instead of statistics

*enspec* is *depvars<sub>en</sub>* = *varlist<sub>en</sub>* [ , *enopts* ]

where *depvars<sub>en</sub>* is a list of endogenous covariates. Each variable in *depvars<sub>en</sub>* specifies an endogenous covariate model using the common *varlist<sub>en</sub>* and options.

*entrspec* is *depvar<sub>tr</sub>* [= *varlist<sub>tr</sub>*] [ , *entropts* ]

where *depvar<sub>tr</sub>* is a variable indicating treatment assignment. *varlist<sub>tr</sub>* is a list of covariates predicting treatment assignment.

## 4 eprobit — Extended probit regression

---

*extrspec* is *tvar* [ , *extropts* ]

where *tvar* is a variable indicating treatment assignment.

*selspec* is *depvar<sub>s</sub>* = *varlist<sub>s</sub>* [ , *selopts* ]

where *depvar<sub>s</sub>* is a variable indicating selection status. *depvar<sub>s</sub>* must be coded as 0, indicating that the observation was not selected, or 1, indicating that the observation was selected. *varlist<sub>s</sub>* is a list of covariates predicting selection.

*tselspec* is *depvar<sub>s</sub>* = *varlist<sub>s</sub>* [ , *tselopts* ]

where *depvar<sub>s</sub>* is a continuous variable. *varlist<sub>s</sub>* is a list of covariates predicting *depvar<sub>s</sub>*. The censoring status of *depvar<sub>s</sub>* indicates selection, where a censored *depvar<sub>s</sub>* indicates that the observation was not selected and a noncensored *depvar<sub>s</sub>* indicates that the observation was selected.

<i>enopts</i>	Description
<hr/>	
Model	
<u>probit</u>	treat endogenous covariate as binary
<u>oprobit</u>	treat endogenous covariate as ordinal
<u>pocorrelation</u>	estimate different correlations for each level of a binary or an ordinal endogenous covariate
<u>nomain</u>	do not add endogenous covariate to main equation
<u>nore</u>	do not include random effects in model for endogenous covariate
<u>noconstant</u>	suppress constant term

---

nore is available only with `xteprobit`.

<i>entroppts</i>	Description
<hr/>	
Model	
<u>pocorrelation</u>	estimate different correlations for each potential outcome
<u>nomain</u>	do not add treatment indicator to main equation
<u>nointeract</u>	do not interact treatment with covariates in main equation
<u>nore</u>	do not include random effects in model for endogenous treatment
<u>noconstant</u>	suppress constant term
<u>offset</u> ( <i>varname<sub>o</sub></i> )	include <i>varname<sub>o</sub></i> in model with coefficient constrained to 1

---

nore is available only with `xteprobit`.

<i>extropts</i>	Description
<hr/>	
Model	
<u>pocorrelation</u>	estimate different correlations for each potential outcome
<u>nomain</u>	do not add treatment indicator to main equation
<u>nointeract</u>	do not interact treatment with covariates in main equation

---

<i>selopts</i>	Description
Model	
<code>nore</code>	do not include random effects in selection model
<code>noconstant</code>	suppress constant term
<code>offset(varname<sub>o</sub>)</code>	include <i>varname<sub>o</sub></i> in model with coefficient constrained to 1

`nore` is available only with `xteprobit`.

<i>tselopts</i>	Description
Model	
<code>*ll(varname   #)</code>	left-censoring variable or limit
<code>*ul(varname   #)</code>	right-censoring variable or limit
<code>main</code>	add censored selection variable to main equation
<code>nore</code>	do not include random effects in tobit selection model
<code>noconstant</code>	suppress constant term
<code>offset(varname<sub>o</sub>)</code>	include <i>varname<sub>o</sub></i> in model with coefficient constrained to 1

\* You must specify either `ll()` or `ul()`.

`nore` is available only with `xteprobit`.

*indepvars*, *varlist<sub>en</sub>*, *varlist<sub>tr</sub>*, and *varlist<sub>s</sub>* may contain factor variables; see [U] 11.4.3 Factor variables.

*devar*, *indepvars*, *depvars<sub>en</sub>*, *varlist<sub>en</sub>*, *devar<sub>tr</sub>*, *varlist<sub>tr</sub>*, *tvar*, *devar<sub>s</sub>*, and *varlist<sub>s</sub>* may contain time-series operators; see [U] 11.4.4 Time-series varlists.

`bootstrap`, `by`, `collect`, `jackknife`, and `statsby` are allowed with `eprobit` and `xteprobit`. `rolling` and `svy` are allowed with `eprobit`. See [U] 11.1.10 Prefix commands.

Weights are not allowed with the `bootstrap` prefix; see [R] `bootstrap`.

`vce()` and weights are not allowed with the `svy` prefix; see [SVY] `svy`.

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

`reintpoints()` and `reintmethod()` are available only with `xteprobit`.

`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

`endogenous(enspec)`, `entreat(entspec)`, `extreat(extrspec)`, `select(selspec)`, `tobitselect(tselspec)`; see [ERM] ERM options.

`noconstant`, `offset(varnameo)`, `constraints(numlist)`; see [R] Estimation options.

### SE/Robust

`vce(vcetype)`; see [ERM] ERM options.

### Reporting

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

*display\_options*: `nocl`, `nopvalues`, `noomitted`, `vsquish`, `noemptycells`, `baselevels`, `allbaselevels`, `nofvlabel`, `fvwrap(#)`, `fvwrapon(style)`, `cformat(%fmt)`, `pformat(%fmt)`, `sformat(%fmt)`, and `nolstretch`; see [R] Estimation options.

## Integration

`intpoints(#)`, `triintpoints(#)`, `reintpoints(#)`, `reintmethod(intmethod)`; see [ERM] **ERM 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**.

The default technique for `eprobit` is `technique(nr)`. The default technique for `xteprobit` 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 `eprobit` and `xteprobit` but are not shown in the dialog box:

`collinear`, `coeflegend`; see [R] **Estimation options**.

## Remarks and examples

[stata.com](http://www.stata.com)

`eprobit` and `xteprobit` fit models that we refer to as “extended probit regression models”, meaning that they accommodate endogenous covariates, nonrandom treatment assignment, endogenous sample selection, and panel data or other grouped data.

`eprobit` fits models for cross-sectional data (one-level models). `eprobit` can account for endogenous covariates, treatment, and sample selection, whether these complications arise individually or in combination.

`xteprobit` fits random-effects models (two-level models) for panel data or grouped data. `xteprobit` accounts for endogenous covariates, treatment, and sample selection in the same way as `eprobit` and also accounts for within-panel or within-group correlation among observations.

In this entry, you will find information on the syntax for the `eprobit` and `xteprobit` commands. You can see *Methods and formulas* for a full description of the models that can be fit with `eprobit` and `xteprobit` and details about how those models are fit.

More information on extended probit regression models is found in the separate introductions and example entries. We recommend reading those entries to learn how to use `eprobit` and `xteprobit`. Below, we provide a guide to help you locate the ones that will be helpful to you.

For an introduction to `eprobit` and `xteprobit` and the other extended regression commands for continuous, interval, and ordinal outcomes, see [ERM] **Intro 1**–[ERM] **Intro 9**.

[ERM] **Intro 1** introduces the ERM commands, the problems they address, and their syntax.

[ERM] **Intro 2** provides background on the four types of models—linear regression, interval regression, probit regression, and ordered probit regression—that can be fit using ERM commands.

[ERM] **Intro 3** considers the problem of endogenous covariates and how to solve it using ERM commands.

[ERM] **Intro 4** gives an overview of endogenous sample selection and using ERM commands to account for it.

[ERM] **Intro 5** covers nonrandom treatment assignment and how to account for it using `eprobit` or any of the other ERM commands.

[ERM] **Intro 6** covers random-effects models for panel data and other grouped data. It discusses `xtprobit` and the other ERM commands for panel data.

[ERM] **Intro 7** discusses interpretation of results. You can interpret coefficients from `eprobit` and `xtprobit` in the usual way, but this introduction goes beyond the interpretation of coefficients. We demonstrate how to find answers to interesting questions by using `margins`. If your model includes an endogenous covariate or an endogenous treatment, the use of `margins` differs from its use after other estimation commands, so we strongly recommend reading this intro if you are fitting these types of models.

[ERM] **Intro 8** will be particularly helpful if you are familiar with `ivprobit`, `heckprobit`, `xtprobit`, and other commands that address endogenous covariates, sample selection, nonrandom treatment assignment, or random effects. This introduction is a Rosetta stone that maps the syntax of those commands to the syntax of `eprobit` and `xtprobit`.

[ERM] **Intro 9** walks you through an example that gives insight into the concepts of endogenous covariates, treatment assignment, and sample selection while fitting models with `eregress` that address these complications. Although the example uses `eregress`, the discussion applies equally to `eprobit`. This intro also demonstrates how to interpret results by using `margins` and `estat teffects`.

Additional examples are presented in [ERM] **Example 1a**–[ERM] **Example 9**. For examples using `eprobit`, see

[ERM] <b>Example 3a</b>	Probit regression with continuous endogenous covariate
[ERM] <b>Example 3b</b>	Probit regression with endogenous covariate and treatment
[ERM] <b>Example 4a</b>	Probit regression with endogenous sample selection
[ERM] <b>Example 4b</b>	Probit regression with endogenous treatment and sample selection
[ERM] <b>Example 5</b>	Probit regression with endogenous ordinal treatment
[ERM] <b>Example 9</b>	Probit regression with endogenous treatment and random effects

See *Examples* in [ERM] **Intro** for an overview of all the examples. All examples may be interesting because they handle complications in the same way.

`eprobit` and `xtprobit` fit many models discussed in the literature. This includes the probit model with continuous endogenous covariates (Newey 1987), the probit model with multiple endogenous binary covariates (Arendt and Holm 2006), the probit model with an endogenous treatment (Angrist 2001 and Pindyck and Rubinfeld 1998), and the random-effects probit model (Conway 1990). `eprobit` can also be used for probit models with selection, such as that discussed by Van de Ven and Van Pragg (1981), and for the model with a tobit selection equation, discussed in Wooldridge (2010, sec. 19.7).

`xtprobit` can be used for the random-effects probit model with selection discussed in Semykina and Wooldridge (2018). The bivariate probit model with random effects discussed in Mulkay (2015) may also be fit using `xtprobit`. Roodman (2011) investigated probit models with endogenous covariates and endogenous sample selection and demonstrated how multiple observational data complications could be addressed with a triangular model structure. He and Tamás Bartus showed how random effects could be used in the triangular model structure in Bartus and Roodman (2014). Roodman’s work has been used to model processes like the impact of finance on the probability of being an entrepreneur (Karymshakov, Sultakeev, and Sulaimanova 2015) and the impact of foreign direct investment on the probability of creating an innovative product (Vahter 2011).

## Stored results

eprobit stores the following in `e()`:

### Scalars

<code>e(N)</code>	number of observations
<code>e(N_selected)</code>	number of selected observations
<code>e(N_nonselected)</code>	number of nonselected observations
<code>e(k)</code>	number of parameters
<code>e(k_cat#)</code>	number of categories for the <code>#th depvar</code> , ordinal
<code>e(k_eq)</code>	number of equations in <code>e(b)</code>
<code>e(k_eq_model)</code>	number of equations in overall model test
<code>e(k_dv)</code>	number of dependent variables
<code>e(k_aux)</code>	number of auxiliary parameters
<code>e(df_m)</code>	model degrees of freedom
<code>e(ll)</code>	log likelihood
<code>e(N_clust)</code>	number of clusters
<code>e(chi2)</code>	$\chi^2$
<code>e(p)</code>	$p$ -value for model test
<code>e(n_quad)</code>	number of integration points for multivariate normal
<code>e(n_quad3)</code>	number of integration points for trivariate normal
<code>e(rank)</code>	rank of <code>e(V)</code>
<code>e(ic)</code>	number of iterations
<code>e(rc)</code>	return code
<code>e(converged)</code>	1 if converged, 0 otherwise

### Macros

<code>e(cmd)</code>	<code>eprobit</code>
<code>e(cmdline)</code>	command as typed
<code>e(depvar)</code>	names of dependent variables
<code>e(tsel_ll)</code>	left-censoring limit for tobit selection
<code>e(tsel_ul)</code>	right-censoring limit for tobit selection
<code>e(wtype)</code>	weight type
<code>e(wexp)</code>	weight expression
<code>e(title)</code>	title in estimation output
<code>e(clustvar)</code>	name of cluster variable
<code>e(offset#)</code>	offset for the <code>#th depvar</code> , where <code>#</code> is determined by equation order in output
<code>e(chi2type)</code>	Wald; type of model $\chi^2$ test
<code>e(vce)</code>	<code>vcetype</code> specified in <code>vce()</code>
<code>e(vcetype)</code>	title used to label Std. err.
<code>e(opt)</code>	type of optimization
<code>e(which)</code>	max or min; whether optimizer is to perform maximization or minimization
<code>e(ml_method)</code>	type of ml method
<code>e(user)</code>	name of likelihood-evaluator program
<code>e(technique)</code>	maximization technique
<code>e(properties)</code>	<code>b V</code>
<code>e(estat_cmd)</code>	program used to implement <code>estat</code>
<code>e(predict)</code>	program used to implement <code>predict</code>
<code>e(marginsok)</code>	predictions allowed by <code>margins</code>
<code>e(marginsnotok)</code>	predictions disallowed by <code>margins</code>
<code>e(asbalanced)</code>	factor variables <code>fvset</code> as <code>asbalanced</code>
<code>e(asobserved)</code>	factor variables <code>fvset</code> as <code>asobserved</code>

### Matrices

<code>e(b)</code>	coefficient vector
<code>e(cat#)</code>	categories for the <code>#th depvar</code> , ordinal
<code>e(Cns)</code>	constraints matrix
<code>e(ilog)</code>	iteration log (up to 20 iterations)
<code>e(gradient)</code>	gradient vector
<code>e(V)</code>	variance-covariance matrix of the estimators
<code>e(V_modelbased)</code>	model-based variance

### Functions

<code>e(sample)</code>	marks estimation sample
------------------------	-------------------------



In addition to the above, the following is stored in `r()`:

Matrices	
<code>r(table)</code>	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.

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

Scalars	
<code>e(N)</code>	number of observations
<code>e(N_g)</code>	number of groups
<code>e(N_selected)</code>	number of selected observations
<code>e(N_nonselected)</code>	number of nonselected observations
<code>e(k)</code>	number of parameters
<code>e(k_cat#)</code>	number of categories for the $\#$ th <i>depvar</i> , ordinal
<code>e(k_eq)</code>	number of equations in <code>e(b)</code>
<code>e(k_eq_model)</code>	number of equations in overall model test
<code>e(k_dv)</code>	number of dependent variables
<code>e(k_aux)</code>	number of auxiliary parameters
<code>e(df_m)</code>	model degrees of freedom
<code>e(ll)</code>	log likelihood
<code>e(N_clust)</code>	number of clusters
<code>e(chi2)</code>	$\chi^2$
<code>e(p)</code>	$p$ -value for model test
<code>e(n_quad)</code>	number of integration points for multivariate normal
<code>e(n_quad3)</code>	number of integration points for trivariate normal
<code>e(n_requad)</code>	number of integration points for random effects
<code>e(g_min)</code>	smallest group size
<code>e(g_avg)</code>	average group size
<code>e(g_max)</code>	largest group size
<code>e(rank)</code>	rank of <code>e(V)</code>
<code>e(ic)</code>	number of iterations
<code>e(rc)</code>	return code
<code>e(converged)</code>	1 if converged, 0 otherwise

#### Macros

<code>e(cmd)</code>	<code>xteprobit</code>
<code>e(cmdline)</code>	command as typed
<code>e(depvar)</code>	names of dependent variables
<code>e(tsel_ll)</code>	left-censoring limit for tobit selection
<code>e(tsel_ul)</code>	right-censoring limit for tobit selection
<code>e(ivar)</code>	variable denoting groups
<code>e(title)</code>	title in estimation output
<code>e(clustvar)</code>	name of cluster variable
<code>e(offset#)</code>	offset for the $\#$ th <i>depvar</i> , where $\#$ is determined by equation order in output
<code>e(chi2type)</code>	Wald; type of model $\chi^2$ test
<code>e(vce)</code>	<i>vcetype</i> specified in <code>vce()</code>
<code>e(vcetype)</code>	title used to label Std. err.
<code>e(reintmethod)</code>	integration method for random effects
<code>e(opt)</code>	type of optimization
<code>e(which)</code>	<code>max</code> or <code>min</code> ; whether optimizer is to perform maximization or minimization
<code>e(ml_method)</code>	type of <code>ml</code> method
<code>e(user)</code>	name of likelihood-evaluator program
<code>e(technique)</code>	maximization technique
<code>e(properties)</code>	<code>b V</code>
<code>e(estat_cmd)</code>	program used to implement <code>estat</code>
<code>e(predict)</code>	program used to implement <code>predict</code>
<code>e(marginsok)</code>	predictions allowed by <code>margins</code>
<code>e(marginsnotok)</code>	predictions disallowed by <code>margins</code>
<code>e(asbalanced)</code>	factor variables <code>fvset</code> as <code>asbalanced</code>
<code>e(asobserved)</code>	factor variables <code>fvset</code> as <code>asobserved</code>

Matrices	
<code>e(b)</code>	coefficient vector
<code>e(cat#)</code>	categories for the #th <i>depvar</i> , ordinal
<code>e(Cns)</code>	constraints matrix
<code>e(ilog)</code>	iteration log (up to 20 iterations)
<code>e(gradient)</code>	gradient vector
<code>e(V)</code>	variance–covariance matrix of the estimators
<code>e(V_modelbased)</code>	model-based variance
Functions	
<code>e(sample)</code>	marks estimation sample

In addition to the above, the following is stored in `r()`:

Matrices	
<code>r(table)</code>	matrix containing the coefficients with their standard errors, test statistics, <i>p</i> -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

The methods and formulas presented here are for the probit model. The estimators implemented in `eprobit` and `xteprobit` are maximum likelihood estimators covered by the results in chapter 13 of Wooldridge (2010) and White (1996).

The log-likelihood functions maximized by `eprobit` and `xteprobit` are implied by the triangular structure of the model. Specifically, the joint distribution of the endogenous 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 chapter 10 of Amemiya (1985); Heckman (1976, 1979); chapter 5 of Maddala (1983); Maddala and Lee (1976); sections 15.7.2, 15.7.3, 16.3.3, 17.5.2, and 19.7.1 in Wooldridge (2010); and Wooldridge (2014). Roodman (2011) and Bartus and Roodman (2014) used this result to derive the formulas discussed below.

Methods and formulas are presented under the following headings:

- Introduction*
- Endogenous covariates*
  - Continuous endogenous covariates*
  - Binary and ordinal endogenous covariates*
- Treatment*
- Endogenous sample selection*
  - Probit endogenous sample selection*
  - Tobit endogenous sample selection*
- Random effects*
- Combined model*
- Confidence intervals*
- Likelihood for multiequation models*

## Introduction

A probit regression of outcome  $y_i$  on covariates  $\mathbf{x}_i$  may be written as

$$y_i = 1 (\mathbf{x}_i\boldsymbol{\beta} + \epsilon_i > 0)$$

where the errors  $\epsilon_i$  are distributed as standard normal. The log likelihood is

$$\ln L = \sum_{i=1}^N w_i \{y_i \ln \Phi(\mathbf{x}_i\boldsymbol{\beta}) + (1 - y_i) \ln \Phi(-\mathbf{x}_i\boldsymbol{\beta})\}$$

where  $w_i$  are the weights. The conditional probability of success is

$$E(y_i|\mathbf{x}_i) = \Pr(y_i = 1|\mathbf{x}_i) = \Phi(\mathbf{x}_i\boldsymbol{\beta})$$

The standard normal cumulative distribution function  $\Phi(\cdot)$  used in these expressions is a one-sided probability that the random variable is below a certain point. In the models we describe later, it will be useful to use two-sided probabilities. For two-sided probabilities, we define  $\Phi_d^*$  with three inputs. The first two inputs are  $d$ -dimensional row vectors  $\mathbf{l}$  and  $\mathbf{u}$  that have values in  $\mathbb{R} \cup \{-\infty, \infty\}$ , the extended real line. The final input is a  $d \times d$  real-valued and positive-definite matrix  $\boldsymbol{\Sigma}$ .

$$\Phi_d^*(\mathbf{l}, \mathbf{u}, \boldsymbol{\Sigma}) = \int_{l_1}^{u_1} \dots \int_{l_d}^{u_d} \phi_d(\boldsymbol{\epsilon}, \boldsymbol{\Sigma}) d\epsilon_1 \dots d\epsilon_d$$

where  $\phi_d$  is the density of a mean 0, multivariate normal random variable. For details on the calculation of  $\Phi_d^*$ , see [M-5] `mvnormal()`. The probabilities are approximated using numeric integration. The number of integration or quadrature points can be varied to attain better approximations. For trivariate errors, we use the method of [Drezner \(1994\)](#). For four or more errors, we use the method of [Miwa, Hayter, and Kuriki \(2003\)](#).

The lower and upper limits  $l_{1i}$  and  $u_{1i}$  on the unobserved  $\epsilon_i$  are based on the observed values of  $y_i$  and  $\mathbf{x}_i$  and are defined as

$$l_{1i} = \begin{cases} -\infty & y_i = 0 \\ -\mathbf{x}_i\boldsymbol{\beta} & y_i = 1 \end{cases} \quad u_{1i} = \begin{cases} -\mathbf{x}_i\boldsymbol{\beta} & y_i = 0 \\ \infty & y_i = 1 \end{cases} \quad (1)$$

They let us rewrite the log likelihood concisely as

$$\ln L = \sum_{i=1}^N w_i \ln \Phi_1^*(l_{1i}, u_{1i}, 1)$$

The conditional probability of success can be written using similar notation:

$$\Pr(y_i = 1|\mathbf{x}_i) = \Phi_1^*(-\mathbf{x}_i\boldsymbol{\beta}, \infty, 1) \quad (2)$$

## Endogenous covariates

### Continuous endogenous covariates

A probit regression of  $y_i$  on exogenous covariates  $\mathbf{x}_i$  and  $C$  continuous endogenous covariates  $\mathbf{w}_{ci}$  has the form

$$y_i = 1 (\mathbf{x}_i\boldsymbol{\beta} + \mathbf{w}_{ci}\boldsymbol{\beta}_c + \epsilon_i > 0)$$

$$\mathbf{w}_{ci} = \mathbf{z}_{ci}\mathbf{A}_c + \epsilon_{ci}$$

The vector  $\mathbf{z}_{ci}$  contains variables from  $\mathbf{x}_i$  and other covariates that affect  $\mathbf{w}_{ci}$ . The unobserved errors  $\epsilon_i$  and  $\epsilon_{ci}$  are multivariate normal with mean 0 and covariance

$$\begin{bmatrix} 1 & \sigma'_{1c} \\ \sigma_{1c} & \boldsymbol{\Sigma}_c \end{bmatrix}$$

We can write the joint density of the dependent variables as a product:

$$f(y_i, \mathbf{w}_{ci} | \mathbf{x}_i, \mathbf{z}_{ci}) = f(y_i | \mathbf{w}_{ci}, \mathbf{x}_i, \mathbf{z}_{ci}) f(\mathbf{w}_{ci} | \mathbf{x}_i, \mathbf{z}_{ci})$$

The conditional density of  $\mathbf{w}_{ci}$  is

$$f(\mathbf{w}_{ci} | \mathbf{x}_i, \mathbf{z}_{ci}) = \phi_C(\mathbf{w}_{ci} - \mathbf{z}_{ci}\mathbf{A}_c, \boldsymbol{\Sigma}_c)$$

Note that

$$\Pr(y_i = 1 | \mathbf{w}_{ci}, \mathbf{x}_i, \mathbf{z}_{ci}) = \Pr(\mathbf{x}_i\boldsymbol{\beta} + \mathbf{w}_{ci}\boldsymbol{\beta}_c + \epsilon_i > 0 | \mathbf{w}_{ci}, \mathbf{x}_i, \mathbf{z}_{ci})$$

So the conditional density of  $y_i$  can be written as a probability for  $\epsilon_i$ . Thus, the conditional distribution of  $\epsilon_i$  can be used to find the conditional density of  $y_i$ . Conditional on the endogenous and exogenous covariates,  $\epsilon_i$  has mean and variance

$$E(\epsilon_i | \mathbf{w}_{ci}, \mathbf{x}_i, \mathbf{z}_{ci}) = \sigma'_{1c} \boldsymbol{\Sigma}_c^{-1} (\mathbf{w}_{ci} - \mathbf{z}_{ci}\mathbf{A}_c)'$$

$$\text{Var}(\epsilon_i | \mathbf{w}_{ci}, \mathbf{x}_i, \mathbf{z}_{ci}) = 1 - \sigma'_{1c} \boldsymbol{\Sigma}_c^{-1} \sigma_{1c}$$

The conditional mean is used in the lower and upper limits for the  $y_i$  probability, which are

$$l_{1i} = \begin{cases} -\infty & y_i = 0 \\ -\mathbf{x}_i\boldsymbol{\beta} - \sigma'_{1c} \boldsymbol{\Sigma}_c^{-1} (\mathbf{w}_{ci} - \mathbf{z}_{ci}\mathbf{A}_c)' & y_i = 1 \end{cases}$$

$$u_{1i} = \begin{cases} -\mathbf{x}_i\boldsymbol{\beta} - \sigma'_{1c} \boldsymbol{\Sigma}_c^{-1} (\mathbf{w}_{ci} - \mathbf{z}_{ci}\mathbf{A}_c)' & y_i = 0 \\ \infty & y_i = 1 \end{cases}$$

Using these limits, the conditional variance, and the conditional density of  $\mathbf{w}_{ci}$ , we obtain the log likelihood

$$\ln L = \sum_{i=1}^N w_i \{ \ln \Phi_1^*(l_{1i}, u_{1i}, 1 - \sigma'_{1c} \boldsymbol{\Sigma}_c^{-1} \sigma_{1c}) + \ln \phi_C(\mathbf{w}_{ci} - \mathbf{z}_{ci}\mathbf{A}_c, \boldsymbol{\Sigma}_c) \}$$

Letting

$$l_{1i1} = -\mathbf{x}_i\boldsymbol{\beta} - \sigma'_{1c} \boldsymbol{\Sigma}_c^{-1} (\mathbf{w}_{ci} - \mathbf{z}_{ci}\mathbf{A}_c)'$$

$$u_{1i1} = \infty$$

the conditional probability of success is

$$\Pr(y_i = 1 | \mathbf{w}_{ci}, \mathbf{x}_i, \mathbf{z}_{ci}) = \Phi_1^*(l_{1i1}, u_{1i1}, 1 - \sigma'_{1c} \boldsymbol{\Sigma}_c^{-1} \sigma_{1c})$$

## Binary and ordinal endogenous covariates

Here we begin by formulating the probit regression of  $y_i$  on exogenous covariates  $\mathbf{x}_i$  and  $B$  binary and ordinal endogenous covariates  $\mathbf{w}_{bi} = [w_{b1i}, \dots, w_{bBi}]$ . Indicator (dummy) variables for the levels of each binary and ordinal covariate are used in the model. You can also interact other covariates with the binary and ordinal endogenous covariates, as in treatment-effect models.

Let  $j = 1, \dots, B$ . We use a probit model for binary endogenous covariates

$$w_{bji} = 1 (\mathbf{z}_{bji}\boldsymbol{\alpha}_{bj} + \epsilon_{bji} > 0)$$

For ordinal endogenous covariate  $w_{bji}$  that takes values  $v_{bj1}, \dots, v_{bjB_j}$  with covariates  $\mathbf{z}_{bji}$ , we have the ordered probit model

$$w_{bji} = v_{bjh} \quad \text{iff} \quad \kappa_{bj(h-1)} < \mathbf{z}_{bji}\boldsymbol{\alpha}_{bj} + \epsilon_{bji} \leq \kappa_{bjh} \quad (3)$$

The values  $v_{bj1}, \dots, v_{bjB_j}$  are real numbers such that  $v_{bjh} < v_{bjm}$  for  $h < m$ .  $\kappa_{bj0}$  is taken as  $-\infty$  and  $\kappa_{bjB_j}$  is taken as  $+\infty$ . The errors  $\epsilon_{b1i}, \dots, \epsilon_{bBi}$  are multivariate normal with mean 0 and covariance

$$\boldsymbol{\Sigma}_b = \begin{bmatrix} 1 & \rho_{b12} & \cdots & \rho_{b1B} \\ \rho_{b12} & 1 & \cdots & \rho_{b2B} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{b1B} & \rho_{b2B} & \cdots & 1 \end{bmatrix}$$

Because the covariate  $w_{bji}$  is binary or ordinal, the effect of each category in the outcome equation is made with an indicator variable.

$$\mathbf{wind}_{bji} = \begin{bmatrix} 1(w_{bji} = v_{bj1}) \\ \vdots \\ 1(w_{bji} = v_{bjB_j}) \end{bmatrix}' \quad (4)$$

The model for the outcome can be formulated with or without different correlation parameters for each level of  $\mathbf{w}_{bi}$ . Level-specific parameters are obtained by specifying `pocorrelation` in the `endogenous()` option.

If the correlation parameters are not level specific, we have

$$y_i = 1(\mathbf{x}_i\boldsymbol{\beta} + \mathbf{wind}_{b1i}\boldsymbol{\beta}_{b1} + \cdots + \mathbf{wind}_{bBi}\boldsymbol{\beta}_{bB} + \epsilon_i > 0)$$

where the outcome error  $\epsilon_i$  and binary and ordinal endogenous errors  $\epsilon_{b1i}, \dots, \epsilon_{bBi}$  are multivariate normal with mean 0 and covariance

$$\boldsymbol{\Sigma} = \begin{bmatrix} 1 & \boldsymbol{\rho}'_{1b} \\ \boldsymbol{\rho}_{1b} & \boldsymbol{\Sigma}_b \end{bmatrix}$$

From here, we discuss the model with ordinal endogenous covariates. The results for binary endogenous covariates are similar.

For  $j = 1, \dots, B$  and  $h = 0, \dots, B_j$ , let

$$c_{bjih} = \begin{cases} -\infty & h = 0 \\ \kappa_{bjh} - \mathbf{z}_{bji}\boldsymbol{\alpha}_{bj} & h = 1, \dots, B_j - 1 \\ \infty & h = B_j \end{cases}$$

The probability for  $w_{bji}$  has lower limit

$$l_{bji} = c_{bji(h-1)} \quad \text{if} \quad w_{bji} = v_{bjh} \quad (5)$$

and upper limit

$$u_{bji} = c_{bjih} \quad \text{if} \quad w_{bji} = v_{bjh} \quad (6)$$

Letting

$$c_{bi} = -\mathbf{x}_i\boldsymbol{\beta} - \mathbf{w}_{b1i}\boldsymbol{\beta}_{b1} - \cdots - \mathbf{w}_{bBi}\boldsymbol{\beta}_{bB}$$

the lower and upper limits for the  $y_i$  probability are

$$l_{1i} = \begin{cases} -\infty & y_i = 0 \\ c_{bi} & y_i = 1 \end{cases} \quad u_{1i} = \begin{cases} c_{bi} & y_i = 0 \\ \infty & y_i = 1 \end{cases}$$

and

$$\mathbf{l}_i = [l_{1i} \quad l_{b1i} \quad \cdots \quad l_{bBi}]$$

$$\mathbf{u}_i = [u_{1i} \quad u_{b1i} \quad \cdots \quad u_{bBi}]$$

The log likelihood for this model is

$$\ln L = \sum_{i=1}^N w_i \ln \Phi_{B+1}^*(\mathbf{l}_i, \mathbf{u}_i, \boldsymbol{\Sigma})$$

Now let

$$\mathbf{l}_{bi} = [l_{b1i} \quad \cdots \quad l_{bBi}]$$

$$\mathbf{u}_{bi} = [u_{b1i} \quad \cdots \quad u_{bBi}]$$

$$\mathbf{l}_{i1} = [-\infty \quad \mathbf{l}_{bi}]$$

$$\mathbf{u}_{i1} = [c_{bi} \quad \mathbf{u}_{bi}]$$

The conditional probability of success is

$$\Pr(y_i = 1 | \mathbf{x}_i, \mathbf{z}_{b1i}, \dots, \mathbf{z}_{bBi}, \mathbf{w}_{bi}) = \frac{\Phi_{B+1}^*(\mathbf{l}_{i1}, \mathbf{u}_{i1}, \boldsymbol{\Sigma})}{\Phi_B^*(\mathbf{l}_{bi}, \mathbf{u}_{bi}, \boldsymbol{\Sigma}_b)}$$

When the endogenous ordinal variables are different treatments, holding the correlation parameters constant over the treatment levels is a constrained form of the potential-outcome model. In an unconstrained potential-outcome model, the correlations between the outcome and the treatments—the endogenous ordinal regressors  $\mathbf{w}_{bi}$ —vary over the levels of each treatment.

In this unconstrained model, there is a different potential-outcome error for each level of each treatment. For example, when the endogenous treatment variable  $w_1$  has three levels (0, 1, and 2) and the endogenous treatment variable  $w_2$  has four levels (0, 1, 2, and 3), the unconstrained model has  $12 = 3 \times 4$  outcome errors. Because there is a different correlation between each potential outcome and each endogenous treatment, there are  $2 \times 12$  correlation parameters between the potential outcomes and the treatments in this example model.

We denote the number of different combinations of values for the endogenous treatments  $\mathbf{w}_{bi}$  by  $M$ , and we denote the vector of values in each combination by  $\mathbf{v}_j$  ( $j \in \{1, 2, \dots, M\}$ ). Letting  $k_{wp}$  be the number of levels of endogenous ordinal treatment variable  $p \in \{1, 2, \dots, B\}$  implies that  $M = k_{w1} \times k_{w2} \times \cdots \times k_{wB}$ .

Denoting the outcome errors  $\epsilon_{1i}, \dots, \epsilon_{Mi}$ , we have

$$\begin{aligned} y_{1i} &= 1(\mathbf{x}_i\boldsymbol{\beta} + \mathbf{wind}_{b1i}\boldsymbol{\beta}_{b1} + \dots + \mathbf{wind}_{bBi}\boldsymbol{\beta}_{bB} + \epsilon_{1i} > 0) \\ &\vdots \\ y_{Mi} &= 1(\mathbf{x}_i\boldsymbol{\beta} + \mathbf{wind}_{b1i}\boldsymbol{\beta}_{b1} + \dots + \mathbf{wind}_{bBi}\boldsymbol{\beta}_{bB} + \epsilon_{Mi} > 0) \\ y_i &= \sum_{j=1}^M 1(\mathbf{w}_{bi} = \mathbf{v}_j)y_{ji} \end{aligned}$$

For  $j = 1, \dots, M$ , the outcome error  $\epsilon_{ji}$  and the endogenous errors  $\epsilon_{b1i}, \dots, \epsilon_{bBi}$  are multivariate normal with 0 mean and covariance

$$\boldsymbol{\Sigma}_j = \begin{bmatrix} 1 & \boldsymbol{\rho}'_{j1b} \\ \boldsymbol{\rho}_{j1b} & \boldsymbol{\Sigma}_b \end{bmatrix}$$

Now let

$$\boldsymbol{\Sigma}_{i,b} = \sum_{j=1}^M 1(\mathbf{w}_{bi} = \mathbf{v}_j)\boldsymbol{\Sigma}_j$$

Now the log likelihood for this model is

$$\ln L = \sum_{i=1}^N w_i \ln \Phi_{B+1}^*(\mathbf{1}_i, \mathbf{u}_i, \boldsymbol{\Sigma}_{i,b})$$

The conditional probability of success is

$$\Pr(y_i = 1 | \mathbf{x}_i, \mathbf{z}_{b1i}, \dots, \mathbf{z}_{bBi}, \mathbf{w}_{bi}) = \frac{\Phi_{B+1}^*(\mathbf{1}_{i1}, \mathbf{u}_{i1}, \boldsymbol{\Sigma}_{i,b})}{\Phi_B^*(\mathbf{1}_{bi}, \mathbf{u}_{bi}, \boldsymbol{\Sigma}_b)}$$

## Treatment

In the potential-outcomes framework, the treatment  $t_i$  is a discrete variable taking  $T$  values, indexing the  $T$  potential outcomes of the outcome  $y_i$ :  $y_{1i}, \dots, y_{Ti}$ .

When we observe treatment  $t_i$  with levels  $v_1, \dots, v_T$ , we have

$$y_i = \sum_{j=1}^T 1(t_i = v_j)y_{ji}$$

So for each observation, we observe only the potential outcome associated with that observation's treatment value.

For exogenous treatments, our approach is equivalent to the regression adjustment treatment-effect estimation method. See [CAUSAL] **teffects intro advanced**. We do not model the treatment assignment process. The formulas for the treatment effects and potential-outcome means (POMs) are equivalent to what we provide here for endogenous treatments. The treatment effect on the treated for  $\mathbf{x}_i$  for an exogenous treatment is equivalent to what we provide here for the endogenous treatment when the correlation parameter between the outcome and treatment errors is set to 0. The average treatment effects (ATES) and POMs for exogenous treatments are estimated as predictive margins in an analogous manner to what we describe here for endogenous treatments.

From here, we assume an endogenous treatment  $t_i$ . For ordinal treatment  $t_i$  with covariates  $\mathbf{z}_{ti}$ , we have the ordered probit model

$$t_i = v_h \quad \text{iff} \quad \kappa_{h-1} < \mathbf{z}_{ti}\boldsymbol{\alpha}_t + \epsilon_{ti} \leq \kappa_h \quad (7)$$

The treatment values  $v_1, \dots, v_T$  are real numbers such that  $v_h < v_m$  for  $h < m$ .  $\kappa_0$  is taken as  $-\infty$  and  $\kappa_T$  is taken as  $+\infty$ . The treatment error  $\epsilon_{ti}$  is standard normal.

We use a probit model for binary treatments that take values in  $\{0, 1\}$ ,

$$t_i = 1 (\mathbf{z}_{ti}\boldsymbol{\alpha}_t + \epsilon_{ti} > 0)$$

A probit regression of  $y_i$  on exogenous covariates  $\mathbf{x}_i$  and endogenous treatment  $t_i$  taking values  $v_1, \dots, v_T$  has the form

$$\begin{aligned} y_{1i} &= 1 (\mathbf{x}_i\boldsymbol{\beta}_1 + \epsilon_{1i} > 0) \\ &\vdots \\ y_{Ti} &= 1 (\mathbf{x}_i\boldsymbol{\beta}_T + \epsilon_{Ti} > 0) \\ y_i &= \sum_{j=1}^T 1(t_i = v_j)y_{ji} \end{aligned}$$

This model can be formulated with or without different correlation parameters for each potential outcome. Potential-outcome specific parameters are obtained by specifying `pocorrelation` in the `entreat()` option.

If the correlation parameters are not potential-outcome specific, for  $j = 1, \dots, T$ ,  $\epsilon_{ji}$  and  $\epsilon_{ti}$  are bivariate normal with mean 0 and covariance

$$\boldsymbol{\Sigma} = \begin{bmatrix} 1 & \rho_{1t} \\ \rho_{1t} & 1 \end{bmatrix}$$

The treatment is exogenous if  $\rho_{1t} = 0$ . Note that we did not specify the structure of the correlations between the potential-outcome errors. We do not need information about these correlations to estimate POMs and treatment effects because all covariates and the outcome are observed in observations from each group.

From here, we discuss a model with an ordinal endogenous treatment. The results for binary treatment models are similar. Because the unobserved errors are bivariate normal, we can express the log likelihood in terms of the  $\Phi_2^*$  function.

For  $j = 1, \dots, T$ , let

$$c_{1ij} = -\mathbf{x}_i\boldsymbol{\beta}_j$$

The lower and upper limits for the  $y_i$  probability are

$$l_{1i} = \begin{cases} -\infty & y_i = 0 \\ c_{1ij} & y_i = 1, t_i = v_j \end{cases} \quad u_{1i} = \begin{cases} c_{1ij} & y_i = 0, t_i = v_j \\ \infty & y_i = 1 \end{cases}$$



For  $j = 0, \dots, T$ , define

$$c_{tij} = \begin{cases} -\infty & j = 0 \\ \kappa_j - \mathbf{z}_{ti}\boldsymbol{\alpha}_t & j = 1, \dots, T-1 \\ \infty & j = T \end{cases}$$

So for the  $t_i$  probability, we have lower limit

$$l_{ti} = c_{ti(j-1)} \quad \text{if } t_i = v_j \quad (8)$$

and upper limit

$$u_{ti} = c_{tij} \quad \text{if } t_i = v_j \quad (9)$$

The log likelihood for the model is

$$\ln L = \sum_{i=1}^N w_i \ln \Phi_2^*([l_{1i} \quad l_{ti}], [u_{1i} \quad u_{ti}], \boldsymbol{\Sigma})$$

The conditional probability of obtaining treatment level  $v_h$  is

$$\Pr(t_i = v_h | \mathbf{z}_{ti}) = \Phi_1^*(c_{ti(h-1)}, c_{tih}, 1)$$

The conditional probability of success at treatment level  $v_j$  is

$$\Pr(y_i = 1 | \mathbf{x}_i, \mathbf{z}_{ti}, t_i = v_j) = \frac{\Phi_2^*([c_{1ij} \quad c_{ti(j-1)}], [\infty \quad c_{tij}], \boldsymbol{\Sigma})}{\Phi_1^*(c_{ti(j-1)}, c_{tij}, 1)}$$

The conditional POM for treatment group  $j$  is

$$\text{POM}_j(\mathbf{x}_i) = E(y_{ji} | \mathbf{x}_i) = \Phi_1^*(c_{1ij}, \infty, 1)$$

Conditional on the covariates  $\mathbf{x}_i$  and  $\mathbf{z}_{ti}$  and the treatment  $t_i = v_h$ , the POM for treatment group  $j$  is

$$\begin{aligned} \text{POM}_j(\mathbf{x}_i, \mathbf{z}_{ti}, t_i = v_h) &= E(y_{ji} | \mathbf{x}_i, \mathbf{z}_{ti}, t_i = v_h) \\ &= \frac{\Phi_2^*([c_{1ij} \quad c_{ti(h-1)}], [\infty \quad c_{tih}], \boldsymbol{\Sigma})}{\Phi_1^*(c_{ti(h-1)}, c_{tih}, 1)} \end{aligned}$$

The treatment effect  $y_{ji} - y_{1i}$  is the difference in the outcome for individual  $i$  if the individual receives the treatment  $t_i = v_j$  instead of the control  $t_i = v_1$  and what the difference would have been if the individual received the control treatment instead.

For treatment group  $j$ , the treatment effect (TE) conditioned on  $\mathbf{x}_i$  is

$$\text{TE}_j(\mathbf{x}_i) = E(y_{ji} - y_{1i} | \mathbf{x}_i) = \text{POM}_j(\mathbf{x}_i) - \text{POM}_1(\mathbf{x}_i)$$

For treatment group  $j$ , the treatment effect on the treated (TET) in treatment group  $h$  conditioned on  $\mathbf{x}_i$  and  $\mathbf{z}_{ti}$  is

$$\begin{aligned} \text{TET}_j(\mathbf{x}_i, \mathbf{z}_{ti}, t_i = v_h) &= E(y_{ji} - y_{1i} | \mathbf{x}_i, \mathbf{z}_{ti}, t_i = v_h) \\ &= \text{POM}_j(\mathbf{x}_i, \mathbf{z}_{ti}, t_i = v_h) - \text{POM}_1(\mathbf{x}_i, \mathbf{z}_{ti}, t_i = v_h) \end{aligned}$$

We can take the expectation of these conditional predictions over the covariates to get population average parameters. The `margins` command is used to estimate the expectations as predictive margins once the model is fit with `eprobit`. The POM for treatment group  $j$  is

$$\text{POM}_j = E(y_{ji}) = E\{\text{POM}_j(\mathbf{x}_i)\}$$

The ATE for treatment group  $j$  is

$$\text{ATE}_j = E(y_{ji} - y_{1i}) = E\{\text{TE}_j(\mathbf{x}_i)\}$$

For treatment group  $j$ , the average treatment effect on the treated (ATET) in treatment group  $h$  is

$$\begin{aligned} \text{ATET}_{jh} &= E(y_{ji} - y_{1i} | t_i = v_h) \\ &= E\{\text{TET}_j(\mathbf{x}_i, \mathbf{z}_{ti}, t_i = v_h) | t_i = v_h\} \end{aligned}$$

If the correlation parameters are potential-outcome specific, for  $j = 1, \dots, T$ ,  $\epsilon_{ji}$  and  $\epsilon_{ti}$  are bivariate normal with mean 0 and covariance

$$\Sigma_j = \begin{bmatrix} 1 & \rho_{j1t} \\ \rho_{j1t} & 1 \end{bmatrix}$$

Now define

$$\Sigma_i = \sum_{j=1}^T 1(t_i = v_j) \Sigma_j$$

The log likelihood for the potential-outcome specification correlation model is

$$\ln L = \sum_{i=1}^N w_i \ln \Phi_2^*([l_{1i} \quad l_{ti}], [u_{1i} \quad u_{ti}], \Sigma_i)$$

The conditional probability of success at treatment level  $v_j$  is

$$\Pr(y_i = 1 | \mathbf{x}_i, \mathbf{z}_{ti}, t_i = v_j) = \frac{\Phi_2^*([c_{1ij} \quad c_{ti(j-1)}], [\infty \quad c_{tij}], \Sigma_j)}{\Phi_1^*(c_{ti(j-1)}, c_{tij}, 1)}$$

The conditional POM for exogenous covariates  $\mathbf{x}_i$  and treatment group  $j$  has the same definition as in the single correlation case. However, when we also condition on the treatment level  $t_i = v_h$  and  $\mathbf{z}_{ti}$ , the POM for treatment group  $j$  is

$$\begin{aligned} \text{POM}_j(\mathbf{x}_i, \mathbf{z}_{ti}, t_i = v_h) &= E(y_{ji} | \mathbf{x}_i, \mathbf{z}_{ti}, t_i = v_h) \\ &= \frac{\Phi_2^*([c_{1ij} \quad c_{ti(h-1)}], [\infty \quad c_{tih}], \Sigma_j)}{\Phi_1^*(c_{ti(h-1)}, c_{tih}, 1)} \end{aligned}$$

Treatment effects are formulated as in the single correlation case but using these updated POM definitions. We can take the expectation of these conditional predictions over the covariates to get population-averaged parameters. The `estat teffects` or `margins` command is used to estimate the expectations as predictive margins once the model is fit with `eprobit`.

## Endogenous sample selection

### Probit endogenous sample selection

A probit model for outcome  $y_i$  with selection on  $s_i$  has the form

$$\begin{aligned} y_i &= 1 (\mathbf{x}_i\boldsymbol{\beta} + \epsilon_i > 0) \\ s_i &= 1 (\mathbf{z}_{si}\boldsymbol{\alpha}_s + \epsilon_{si} > 0) \end{aligned}$$

where  $\mathbf{x}_i$  are covariates that affect the outcome and  $\mathbf{z}_{si}$  are covariates that affect selection. The outcome  $y_i$  is observed if  $s_i = 1$  and not observed if  $s_i = 0$ . The unobserved errors  $\epsilon_i$  and  $\epsilon_{si}$  are normal with mean 0 and covariance

$$\boldsymbol{\Sigma} = \begin{bmatrix} 1 & \rho_{1s} \\ \rho_{1s} & 1 \end{bmatrix}$$

The lower and upper limits for the  $y_i$  probability,  $l_{1i}$  and  $u_{1i}$ , are as defined in (1). For the selection indicator, we have lower and upper limits

$$l_{si} = \begin{cases} -\infty & s_i = 0 \\ -\mathbf{z}_{si}\boldsymbol{\alpha}_s & s_i = 1 \end{cases} \quad u_{si} = \begin{cases} -\mathbf{z}_{si}\boldsymbol{\alpha}_s & s_i = 0 \\ \infty & s_i = 1 \end{cases} \quad (10)$$

The log likelihood for the model is

$$\begin{aligned} \ln L &= \sum_{i \in S} w_i \ln \Phi_2^*([l_{1i} \quad l_{si}], [u_{1i} \quad u_{si}], \boldsymbol{\Sigma}) + \\ &\quad \sum_{i \notin S} w_i \ln \Phi_1^*(l_{si}, u_{si}, 1) \end{aligned}$$

where  $S$  is the set of observations for which  $y_i$  is observed.

In this model, the probability of success is usually predicted conditional on the covariates  $\mathbf{x}_i$  and not on the selection status  $s_i$ . The formulas for the conditional probability are thus the same as in (2).

The conditional probability of selection is

$$\Pr(s_i = 1 | \mathbf{z}_{si}) = \Phi_1^*(-\mathbf{z}_{si}\boldsymbol{\alpha}_s, \infty, 1)$$

### Tobit endogenous sample selection

Instead of constraining the selection indicator to be binary, tobit endogenous sample selection uses a censored continuous sample-selection indicator. We allow the selection variable to be left- or right-censored.

A probit model for outcome  $y_i$  with tobit selection on  $s_i$  has the form

$$y_i = 1 (\mathbf{x}_i\boldsymbol{\beta} + \epsilon_i > 0)$$

We observe the selection indicator  $s_i$ , which indicates the censoring status of the latent selection variable  $s_i^*$ ,

$$s_i^* = \mathbf{z}_{si}\boldsymbol{\alpha}_s + \epsilon_{si}$$

$$s_i = \begin{cases} l_i & s_i^* \leq l_i \\ s_i^* & l_i < s_i^* < u_i \\ u_i & s_i^* \geq u_i \end{cases}$$

where  $\mathbf{z}_{si}$  are covariates that affect selection and  $l_i$  and  $u_i$  are fixed lower and upper limits.

The outcome  $y_i$  is observed when  $s_i^*$  is not censored ( $l_i < s_i^* < u_i$ ). The outcome  $y_i$  is not observed when  $s_i^*$  is left-censored ( $s_i^* \leq l_i$ ) or  $s_i^*$  is right-censored ( $s_i^* \geq u_i$ ). The unobserved errors  $\epsilon_i$  and  $\epsilon_{si}$  are normal with mean 0 and covariance

$$\begin{bmatrix} 1 & \rho_{1s}\sigma_s \\ \rho_{1s}\sigma_s & \sigma_s^2 \end{bmatrix}$$

For the selected observations, we can treat  $s_i$  as a continuous endogenous regressor, as in *Continuous endogenous covariates*. In fact,  $s_i$  may even be used as a regressor for  $y_i$  in `eprobit` (specify `tobitselect(... main)`). On the nonselected observations, we treat  $s_i$  like the probit endogenous sample-selection indicator in *Probit endogenous sample selection*.

For nonselected observations, we have

$$\begin{aligned} \Pr(s_i^* \leq l_i | \mathbf{z}_{si}, \mathbf{x}_i) &= \Pr(\mathbf{z}_{si}\boldsymbol{\alpha}_s + \epsilon_{si} \leq l_i) \\ &= \Phi\left(\frac{l_i - \mathbf{z}_{si}\boldsymbol{\alpha}_s}{\sigma_s}\right) \end{aligned}$$

and

$$\begin{aligned} \Pr(s_i^* \geq u_i | \mathbf{z}_{si}, \mathbf{x}_i) &= \Pr(\mathbf{z}_{si}\boldsymbol{\alpha}_s + \epsilon_{si} \geq u_i) \\ &= \Phi\left(\frac{\mathbf{z}_{si}\boldsymbol{\alpha}_s - u_i}{\sigma_s}\right) \end{aligned}$$

The lower and upper limits for the  $s_i$  probability for nonselected observations where  $s_i^*$  is left-censored are

$$l_{li} = -\infty$$

$$u_{li} = \frac{l_i - \mathbf{z}_{si}\boldsymbol{\alpha}_s}{\sigma_s}$$

The lower and upper limits for the  $s_i$  probability for nonselected observations where  $s_i^*$  is right-censored are

$$l_{ui} = \frac{u_i - \mathbf{z}_{si}\boldsymbol{\alpha}_s}{\sigma_s}$$

$$u_{ui} = \infty$$

Now we consider the selected observations. For  $s_i = s_i^* = S_i$ , we can write the joint density of the dependent variables as a product,

$$f(y_i, s_i = S_i | \mathbf{x}_i, \mathbf{z}_{si}) = f(y_i | s_i = S_i, \mathbf{x}_i, \mathbf{z}_{si})f(s_i = S_i | \mathbf{x}_i, \mathbf{z}_{si})$$

The marginal density of  $s_i = S_i$  is

$$f(s_i = S_i | \mathbf{x}_i, \mathbf{z}_{s,i}) = \phi(S_i - \mathbf{z}_{s,i} \boldsymbol{\alpha}_s, \sigma_s^2)$$

The conditional density of  $y_i$  can be written as a probability for  $\epsilon_i$ . Thus, the conditional distribution of  $\epsilon_i$  can be used to find the conditional density of  $y_i$ . Conditional on  $s_i = S_i$ ,  $\epsilon_i$  has mean and variance

$$\begin{aligned} E(\epsilon_i | s_i = S_i, \mathbf{x}_i, \mathbf{z}_{s,i}) &= \rho_{1s} \sigma_s^{-1} (S_i - \mathbf{z}_{s,i} \boldsymbol{\alpha}) \\ \text{Var}(\epsilon_i | s_i = S_i, \mathbf{x}_i, \mathbf{z}_{s,i}) &= 1 - \rho_{1,s}^2 \end{aligned}$$

The conditional mean is used in the lower and upper limits for the  $y_i$  probability for selected observations, which are

$$\begin{aligned} l_{1i} &= \begin{cases} -\infty & y_i = 0 \\ -\mathbf{x}_i \boldsymbol{\beta} - \rho_{1s} \sigma_s^{-1} (s_i - \mathbf{z}_{s,i} \boldsymbol{\alpha}) & y_i = 1 \end{cases} \\ u_{1i} &= \begin{cases} -\mathbf{x}_i \boldsymbol{\beta} - \rho_{1s} \sigma_s^{-1} (s_i - \mathbf{z}_{s,i} \boldsymbol{\alpha}) & y_i = 0 \\ \infty & y_i = 1 \end{cases} \end{aligned}$$

It follows that the log likelihood is

$$\begin{aligned} \ln L &= \sum_{i \in S} w_i \{ \ln \Phi_1^*(l_{1i}, u_{1i}, 1 - \rho_{1s}^2) + \ln \phi(s_i - \mathbf{z}_{s,i} \boldsymbol{\alpha}_s, \sigma_s^2) \} \\ &+ \sum_{i \in L} w_i \ln \Phi_1^*(l_{1i}, u_{1i}, 1) \\ &+ \sum_{i \in U} w_i \ln \Phi_1^*(l_{ui}, u_{ui}, 1) \end{aligned}$$

where  $S$  is the set of observations for which  $y_i$  is observed,  $L$  is the set of observations where  $s_i^*$  is left-censored, and  $U$  is the set of observations where  $s_i^*$  is right-censored.

The probability of success conditional on  $s_i = s_i^* = S_i$  is

$$\Pr(y_i = 1 | \mathbf{x}_i, s_i = s_i^* = S_i) = \Phi_1^* \{ -\mathbf{x}_i \boldsymbol{\beta} - \rho_{1s} \sigma_s^{-1} (S_i - \mathbf{z}_{s,i} \boldsymbol{\alpha}), \infty, 1 - \rho_{1s}^2 \}$$

If we do not include  $s_i$  in the main outcome equation, the probability of success is calculated as (2) again.

## Random effects

For a probit regression with random effects, we observe panel data. For panel  $i = 1, \dots, N$  and observation  $j = 1, \dots, N_i$ , a probit regression of outcome  $y_{ij}$  on covariates  $\mathbf{x}_{ij}$  may be written as

$$y_{ij} = 1 (\mathbf{x}_{ij} \boldsymbol{\beta} + \epsilon_{ij} + u_i > 0)$$

The random effect  $u_i$  is normal with mean 0 and variance  $\sigma_u^2$ . It is independent of the observation-level error  $\epsilon_{ij}$ , which is standard normal.

We derive the likelihood by using the conditional density of  $y_{ij}$  on the random effect  $u_i$  and the marginal density of  $u_i$ . Multiplying them together, we have the joint density, which is integrated over  $u_i$ .

Let

$$l_{ij}(u) = y_{ij}\Phi(\mathbf{x}_{ij}\boldsymbol{\beta} + u) + (1 - y_{ij})\Phi(-\mathbf{x}_{ij}\boldsymbol{\beta} - u)$$

The likelihood for panel  $i$  is

$$L_i = \int_{-\infty}^{\infty} \phi\left(\frac{u_i}{\sigma_u}\right) \prod_{j=1}^{N_i} l_{ij}(u_i) du_i$$

We can approximate this integral using Gauss–Hermite quadrature. For  $q$ -point Gauss–Hermite quadrature, let the abscissa and weight pairs be denoted by  $(a_{ki}, w_{ki})$ ,  $k = 1, \dots, q$ . Then, the Gauss–Hermite quadrature approximation is

$$\int_{-\infty}^{\infty} f(x) \exp(-x^2) dx \approx \sum_{k=1}^q w_{ki} f(a_{ki})$$

The default approximation used by `xteprobit` is mean–variance adaptive Gauss–Hermite quadrature. This chooses optimal abscissa and weights for each panel. See [Likelihood for multiequation models](#) in [ERM] **eprobit** for more information on the use of mean–variance adaptive Gauss–Hermite quadrature.

Using the quadrature approximation, the log likelihood is

$$\ln L = \sum_{i=1}^N \ln \left\{ \sum_{k=1}^q w_{ki} \prod_{j=1}^{N_i} l_{ij}(\sigma_u a_{ki}) \right\}$$

Now we will derive the conditional probability of success. This is similar to what was given in [Introduction](#), but the variance input to  $\Phi_1^*$  is the variance of the random effect plus the observation-level error.

First, let

$$\xi_{ij} = \epsilon_{ij} + u_i$$

where  $\xi_{ij}$  is normal with mean 0 and variance  $\sigma_\xi^2 = 1 + \sigma_u^2$ .

Then, the conditional probability of success is

$$\Pr(y_{ij} = 1 | \mathbf{x}_{ij}) = \Phi_1^*(-\mathbf{x}_{ij}\boldsymbol{\beta}, \infty, \sigma_\xi^2)$$

## Combined model

Here we present the likelihood for the probit model with continuous endogenous covariates, ordinal endogenous covariates, an ordinal endogenous treatment, and endogenous sample selection. This combines all the extensions to the standard probit model that are supported by `eprobit`. In [Likelihood for multiequation models](#), we describe the general framework for ERMs with multiple features and show how random effects may be combined with other features, how `xteprobit` can support the other ERM features.

Deriving the combined model with tobit rather than probit endogenous sample selection is straightforward. On selected observations, the selection indicator would be treated like a continuous endogenous covariate. On nonselected observations, the model would be identical to the combined model with probit selection. The correlations between the outcome errors and other errors are also the same between treatment groups and levels of ordinal endogenous covariates. Deriving the model with different correlations for the treatment groups and endogenous covariate groups is straightforward. Take the likelihood given here in this section, and use a different covariance matrix depending on the levels of treatment and the ordinal endogenous covariates.

In this model, the treatment  $t_i$  takes  $T$  values, indexing the potential outcomes of the main outcome  $y_i$ :  $y_{1i}, \dots, y_{Ti}$ . The relationship between the ordinal treatment  $t_i$ , treatment covariates  $\mathbf{z}_{t,i}$ , and error  $\epsilon_{ti}$  is described in (7). For  $j = 1, \dots, B$ , the relationship between the ordinal endogenous covariates  $w_{bji}$ , exogenous covariates  $\mathbf{z}_{bji}$ , and error  $\epsilon_{bji}$  is given in (3). The model also uses the  $\mathbf{wind}_{bji}$  terms that are defined in (4).

The probit regression of  $y_i$  on exogenous covariates  $\mathbf{x}_i$ ,  $C$  continuous endogenous covariates  $\mathbf{w}_{ci}$ , and  $B$  ordinal endogenous covariates  $\mathbf{w}_{bi} = [w_{b1i}, \dots, w_{bBi}]$  with endogenous treatment  $t_i$  and endogenous sample selection on  $s_i$  has the form

$$\begin{aligned} y_{1i} &= 1(\mathbf{x}_i\beta_1 + \mathbf{w}_{ci}\beta_{c1} + \mathbf{wind}_{b1i}\beta_{b11} + \dots + \mathbf{wind}_{bBi}\beta_{bB1} + \epsilon_{1i} > 0) \\ &\vdots \\ y_{Ti} &= 1(\mathbf{x}_i\beta_T + \mathbf{w}_{ci}\beta_{cT} + \mathbf{wind}_{b1i}\beta_{b1T} + \dots + \mathbf{wind}_{bBi}\beta_{bBT} + \epsilon_{Ti} > 0) \\ y_i &= \sum_{j=1}^T 1(t_i = v_j)y_{ji} \\ \mathbf{w}_{ci} &= \mathbf{z}_{ci}\mathbf{A}_c + \epsilon_{ci} \\ s_i &= 1(\mathbf{z}_{si}\alpha_s + \epsilon_{si} > 0) \end{aligned}$$

where  $\mathbf{z}_{si}$  are covariates that affect selection and  $\mathbf{z}_{ci}$  are covariates that affect the continuous endogenous covariates. The outcome  $y_i$  is observed if  $s_i = 1$  and is not observed if  $s_i = 0$ .

For  $j = 1, \dots, T$ , the unobserved errors  $\epsilon_{ji}, \epsilon_{si}, \epsilon_{ti}, \epsilon_{b1i}, \dots, \epsilon_{bBi}, \epsilon_{ci}$  are multivariate normal with mean 0 and covariance

$$\Sigma = \begin{bmatrix} 1 & \rho_{1s} & \rho_{1t} & \rho'_{1b} & \sigma'_{1c} \\ \rho_{1s} & 1 & \rho_{st} & \rho'_{sb} & \sigma'_{sc} \\ \rho_{1t} & \rho_{st} & 1 & \rho'_{tb} & \sigma'_{tc} \\ \rho_{1b} & \rho_{sb} & \rho_{tb} & \Sigma_b & \Sigma'_{bc} \\ \sigma_{1c} & \sigma_{sc} & \sigma_{tc} & \Sigma_{bc} & \Sigma_c \end{bmatrix}$$

As in *Continuous endogenous covariates*, we can write the joint density of the dependent variables as a product. We have

$$\begin{aligned} &f(y_i, s_i, t_i, \mathbf{w}_{bi}, \mathbf{w}_{ci} | \mathbf{x}_i, \mathbf{z}_{si}, \mathbf{z}_{ti}, \mathbf{z}_{b1i}, \dots, \mathbf{z}_{bBi}, \mathbf{z}_{ci}) = \\ &f(y_i, s_i, t_i, \mathbf{w}_{bi} | \mathbf{w}_{ci}, \mathbf{x}_i, \mathbf{z}_{si}, \mathbf{z}_{ti}, \mathbf{z}_{b1i}, \dots, \mathbf{z}_{bBi}, \mathbf{z}_{ci}) f(\mathbf{w}_{ci} | \mathbf{z}_{ci}) \end{aligned}$$

We can then use the conditional distribution of  $\epsilon_{ji}, \epsilon_{si}, \epsilon_{ti}, \epsilon_{b1i}, \dots, \epsilon_{bBi}$  to obtain the conditional density of  $y_i, s_i, t_i$ , and  $\mathbf{w}_{bi}$ .

For  $j = 1, \dots, T$ , conditional on  $\mathbf{w}_{ci}$  and the exogenous covariates,  $\epsilon_{ji}$  has mean

$$\begin{aligned} e_{1i} &= E(\epsilon_{ji} | \mathbf{w}_{ci}, \mathbf{x}_i, \mathbf{z}_{si}, \mathbf{z}_{ti}, \mathbf{z}_{b1i}, \dots, \mathbf{z}_{bBi}, \mathbf{z}_{ci}) \\ &= \boldsymbol{\sigma}'_{1,c} \boldsymbol{\Sigma}_c^{-1} (\mathbf{w}_{ci} - \mathbf{z}_{c,i} \mathbf{A}_c)' \end{aligned}$$

Now, for  $j = 1, \dots, T$ , let

$$c_{1ij} = \begin{cases} -\mathbf{x}_i \boldsymbol{\beta}_1 - \mathbf{w}_{ci} \boldsymbol{\beta}_{c,1} - \mathbf{wind}_{b1i} \boldsymbol{\beta}_{b11} - \dots - \mathbf{wind}_{bBi} \boldsymbol{\beta}_{bB1} - e_{1i} & j = 1 \\ \vdots \\ -\mathbf{x}_i \boldsymbol{\beta}_T - \mathbf{w}_{ci} \boldsymbol{\beta}_{cT} - \mathbf{wind}_{b1i} \boldsymbol{\beta}_{b1T} - \dots - \mathbf{wind}_{bBi} \boldsymbol{\beta}_{bBT} - e_{1i} & j = T \end{cases}$$

The lower and upper limits for the  $y_i$  probability are

$$l_{1i} = \begin{cases} -\infty & y_i = 0 \\ c_{1ij} & y_i = 1, t_i = v_j \end{cases} \quad u_{1i} = \begin{cases} c_{1ij} & y_i = 0, t_i = v_j \\ \infty & y_i = 1 \end{cases}$$

The conditional means of the unobserved errors  $\epsilon_{si}, \epsilon_{ti}, \epsilon_{b1i}, \dots, \epsilon_{bBi}$  have similar forms to  $e_{1i}$ . Denote these means by  $e_{si}, e_{ti}, e_{b1i}, \dots, e_{bBi}$ . The lower and upper probability limits for  $s_i, t_i$ , and the ordinal endogenous covariates are obtained by subtracting the means from the limits defined in (10), (8), (9), (5), and (6).

$$\begin{aligned} l_{si}^* &= l_{si} - e_{si} \\ u_{si}^* &= u_{si} - e_{si} \\ l_{ti}^* &= l_{ti} - e_{ti} \\ u_{ti}^* &= u_{ti} - e_{ti} \\ l_{b1i}^* &= l_{b1i} - e_{b1i} \\ u_{b1i}^* &= u_{b1i} - e_{b1i} \\ &\vdots \\ l_{bBi}^* &= l_{bBi} - e_{bBi} \\ u_{bBi}^* &= u_{bBi} - e_{bBi} \end{aligned}$$

We have lower and upper limits; we need a conditional covariance and the conditional density of  $\mathbf{w}_{ci}$  to formulate the likelihood. For  $j = 1, \dots, T$ , conditional on  $\mathbf{w}_{ci}$  and the exogenous covariates,  $\epsilon_{ji}, \epsilon_{si}, \epsilon_{ti}, \epsilon_{b1i}, \dots, \epsilon_{bBi}$  have covariance

$$\boldsymbol{\Sigma}_{o|c} = \begin{bmatrix} 1 & \rho_{1s} & \rho_{1t} & \boldsymbol{\rho}'_{1b} \\ \rho_{1s} & 1 & \rho_{st} & \boldsymbol{\rho}'_{sb} \\ \rho_{1t} & \rho_{st} & 1 & \boldsymbol{\rho}'_{tb} \\ \boldsymbol{\rho}_{1b} & \boldsymbol{\rho}_{sb} & \boldsymbol{\rho}_{tb} & \boldsymbol{\Sigma}_b \end{bmatrix} - \begin{bmatrix} \boldsymbol{\sigma}'_{1c} \\ \boldsymbol{\sigma}'_{sc} \\ \boldsymbol{\sigma}'_{tc} \\ \boldsymbol{\Sigma}'_{bc} \end{bmatrix} \boldsymbol{\Sigma}_c^{-1} \begin{bmatrix} \boldsymbol{\sigma}'_{1c} \\ \boldsymbol{\sigma}'_{sc} \\ \boldsymbol{\sigma}'_{tc} \\ \boldsymbol{\Sigma}'_{bc} \end{bmatrix}'$$

The conditional density of  $\mathbf{w}_{ci}$  is

$$f(\mathbf{w}_{ci} | \mathbf{z}_{ci}) = \phi_C(\mathbf{w}_{ci} - \mathbf{z}_{ci} \mathbf{A}_c, \boldsymbol{\Sigma}_c)$$



Let

$$\begin{aligned} \mathbf{l}_{1i} &= [l_{1i} \quad l_{si}^* \quad l_{ti}^* \quad l_{b1i}^* \quad \dots \quad l_{bBi}^*] \\ \mathbf{u}_{1i} &= [u_{1i} \quad u_{si}^* \quad u_{ti}^* \quad u_{b1i}^* \quad \dots \quad u_{bBi}^*] \\ \mathbf{l}_i &= [l_{si}^* \quad l_{ti}^* \quad l_{b1i}^* \quad \dots \quad l_{bBi}^*] \\ \mathbf{u}_i &= [u_{si}^* \quad u_{ti}^* \quad u_{b1i}^* \quad \dots \quad u_{bBi}^*] \end{aligned}$$

The log likelihood of the model is

$$\begin{aligned} \ln L &= \sum_{i \in S} w_i \ln \Phi_{3+B}^* (\mathbf{l}_{1i}, \mathbf{u}_{1i}, \boldsymbol{\Sigma}_{o|c}) + \\ &\quad \sum_{i \notin S} w_i \ln \Phi_{2+B}^* (\mathbf{l}_i, \mathbf{u}_i, \boldsymbol{\Sigma}_{o|c,-1}) + \\ &\quad \sum_{i=1}^N w_i \ln \phi_C(\mathbf{w}_{ci} - \mathbf{z}_{ci} \mathbf{A}_c, \boldsymbol{\Sigma}_c) \end{aligned}$$

where  $S$  is the set of observations where  $y_i$  is observed and  $\boldsymbol{\Sigma}_{o|c,-1}$  is  $\boldsymbol{\Sigma}_{o|c}$  with the first row and column removed.

As in previous sections, we use the joint and marginal probabilities to determine conditional probabilities.

For  $j = 1, \dots, T$  and  $i$  such that  $t_i = v_j$ , let

$$\begin{aligned} \mathbf{l}_{i11} &= [c_{1ij} \quad l_{ti}^* \quad l_{b1i}^* \quad \dots \quad l_{bBi}^*] \\ \mathbf{u}_{i11} &= [\infty \quad u_{ti}^* \quad u_{b1i}^* \quad \dots \quad u_{bBi}^*] \\ \mathbf{l}_{i12} &= [l_{ti}^* \quad l_{b1i}^* \quad \dots \quad l_{bBi}^*] \\ \mathbf{u}_{i12} &= [u_{ti}^* \quad u_{b1i}^* \quad \dots \quad u_{bBi}^*] \end{aligned}$$

Let  $\boldsymbol{\Sigma}_{o|c,-s}$  be  $\boldsymbol{\Sigma}_{o|c}$  with the second row and column removed. This is the conditional covariance matrix without the endogenous sample-selection equation components. Let  $\boldsymbol{\Sigma}_{o|c,-s-1}$  be  $\boldsymbol{\Sigma}_{o|c,-s}$  with the first row and column removed.

The conditional probability of success at treatment level  $t_i = v_j$  is

$$\Pr(y_i = 1 | t_i = v_j, \mathbf{w}_{bi}, \mathbf{w}_{ci}, \mathbf{x}_i, \mathbf{z}_{si}, \mathbf{z}_{ti}, \mathbf{z}_{b1i}, \dots, \mathbf{z}_{bBi}, \mathbf{z}_{ci}) = \frac{\Phi_{2+B}^* (\mathbf{l}_{i11}, \mathbf{u}_{i11}, \boldsymbol{\Sigma}_{o|c,-s})}{\Phi_{1+B}^* (\mathbf{l}_{i12}, \mathbf{u}_{i12}, \boldsymbol{\Sigma}_{o|c,-s-1})}$$

The conditional probabilities of treatment, selection, and the ordinal endogenous covariates are derived in similar ways. We condition on the treatment and the other endogenous covariates together with the exogenous covariates that affect the outcome. POMS and treatment effects are conditioned on the endogenous and exogenous covariates. See *Predictions using the full model* in [ERM] **eprobit postestimation** for more details.

## Confidence intervals

The estimated variances will always be nonnegative, and the estimated correlations will always fall in  $(-1, 1)$ . We use transformations to obtain confidence intervals that accommodate these ranges.

We use the log transformation to obtain the confidence intervals for variance parameters. Let  $\hat{\sigma}^2$  be a point estimate for the variance parameter  $\sigma^2$ , and let  $\widehat{\text{SE}}(\hat{\sigma}^2)$  be its standard error. The  $(1 - \alpha) \times 100\%$  confidence interval for  $\ln(\sigma^2)$  is

$$\ln(\hat{\sigma}^2) \pm z_{\alpha/2} \frac{\widehat{\text{SE}}(\hat{\sigma}^2)}{\hat{\sigma}^2}$$

where  $z_{\alpha/2}$  is the  $1 - \alpha/2$  quantile of the standard normal distribution. Let  $k_u$  be the upper endpoint of this interval, and let  $k_l$  be the lower. The  $(1 - \alpha) \times 100\%$  confidence interval for  $\sigma^2$  is then given by

$$(e^{k_l}, e^{k_u})$$

We use the inverse hyperbolic tangent transformation to obtain confidence intervals for correlation parameters; for details on the hyperbolic functions, see [FN] **Trigonometric functions**. Let  $\hat{\rho}$  be a point estimate for the correlation parameter  $\rho$ , and let  $\widehat{\text{SE}}(\hat{\rho})$  be its standard error. The  $(1 - \alpha) \times 100\%$  confidence interval for  $\text{atanh}(\rho)$  is

$$\text{atanh}(\hat{\rho}) \pm z_{\alpha/2} \widehat{\text{SE}}(\hat{\rho}) \frac{1}{1 - \hat{\rho}^2}$$

where  $z_{\alpha/2}$  is the  $1 - \alpha/2$  quantile of the standard normal distribution. Let  $k_u$  be the upper endpoint of this interval, and let  $k_l$  be the lower. The  $(1 - \alpha) \times 100\%$  confidence interval for  $\rho$  is then given by

$$\{\tanh(k_l), \tanh(k_u)\}$$

## Likelihood for multiequation models

The general framework for ERMs is formulated such that it accommodates multiple features. Binary and ordinal endogenous covariates may occur together with continuous endogenous covariates in ERMs. Endogenous covariates may also occur together with endogenous sample selection or treatments in ERMs. Random effects may occur in any combination with the other features as well.

Here we show how the log likelihood is formulated when we have multiple auxiliary equations. We begin with the cross-sectional case, where there are no random effects.

Suppose that we have  $H$  auxiliary equations with endogenous outcomes  $y_{1i}, \dots, y_{Hi}$ . We will treat the main outcome  $y_i$  as stage  $J = H + 1$ , so  $y_{Ji} = y_i$ . The ERMs that we fit with `eintreg`, `eoprobit`, `eprobit`, and `eregress` are triangular, so we can order the equations such that the first depends only on exogenous covariates—say,  $\mathbf{w}_{1i} = \mathbf{z}_i$ —and for  $j = 2, \dots, J$ , equation  $j$  depends only on the exogenous covariates  $\mathbf{z}_i$  and the endogenous covariates from equation  $h = j - 1$  and  $y_{1i}, \dots, y_{hi}$  below. These are stored together in  $\mathbf{w}_{ji}$ .

So we have

$$\begin{aligned} y_{1i} &= g_{1i}(\mathbf{w}_{1i}\beta_1 + v_{1i}) \\ &\vdots \\ y_{Hi} &= g_{Hi}(\mathbf{w}_{Hi}\beta_H + v_{Hi}) \\ y_i &= y_{Ji} = g_{Ji}(\mathbf{w}_{Ji}\beta_J + v_{Ji}) \end{aligned}$$

where the form of the functions  $g_{ji}(\cdot)$  is determined by whether the outcome  $y_{ji}$  has a linear, probit, or interval model. The errors  $v_{1i}, \dots, v_{Ji}$  are multivariate normal with mean 0 and covariance  $\Sigma$ .

The covariates  $\mathbf{w}_{ji}$  and the outcome  $y_{ji}$  determine a range for the error  $v_{ji}$ . For example, if  $y_{ji}$  has a linear model, then  $v_{ji} = y_{ji} - \mathbf{w}_{ji}\beta_j$ , the residual. If  $y_{ji} = 1$  and  $y_{ji}$  has a probit model, then  $v_{ji}$  is in the range  $(-\mathbf{w}_{ji}\beta_j, \infty)$ . If  $y_{ji}$  is left-censored at  $l_i$ , then  $v_{ji}$  is in the range  $(-\infty, l_i - \mathbf{w}_{ji}\beta_j)$ .

The density of the endogenous variables can be represented using a multivariate normal density function that is evaluated at the residuals for the continuous outcomes and integrated over the error ranges of the noncontinuous outcomes.

The conditional density of the error  $v_{ji}$  on  $\mathbf{w}_{ji}$  has the form

$$f(v_{ji}|\mathbf{w}_{ji}) = \frac{\int_{\mathbf{V}_{hi}^*} \phi_j(v_{1i}, \dots, v_{ji}, \Sigma_j) d\mathbf{v}_{hi}^*}{\int_{\mathbf{V}_{hi}^*} \phi_h(v_{1i}, \dots, v_{hi}, \Sigma_h) d\mathbf{v}_{hi}^*}$$

where  $\Sigma_j$  is the covariance of  $v_{1i}, \dots, v_{ji}$  and  $\Sigma_h$  is the covariance of  $v_{1i}, \dots, v_{hi}$ , where  $h = j - 1$ . The vector  $\mathbf{v}_{hi}^*$  contains the errors that correspond to binary, ordinal, or censored outcomes in  $y_{1i}, \dots, y_{hi}$ . These outcomes induce the error ranges  $\mathbf{V}_{hi}^*$ , which we integrate over. The other errors are determined by the outcomes and covariates as residuals.

If  $y_{ji}$  is continuous, then

$$f(y_{ji}|\mathbf{w}_{ji}) = f(v_{ji}|\mathbf{w}_{ji}) \quad (11)$$

When  $y_{ji}$  is a binary, ordinal, or censored outcome, we have

$$f(y_{ji}|\mathbf{w}_{ji}) = \frac{\int_{\mathbf{V}_{ji}^*} \phi_j(v_{1i}, \dots, v_{ji}, \Sigma_j) d\mathbf{v}_{ji}^*}{\int_{\mathbf{V}_{hi}^*} \phi_h(v_{1i}, \dots, v_{hi}, \Sigma_h) d\mathbf{v}_{hi}^*} \quad (12)$$

So we also integrate over the range of the error  $v_{ji}$  when  $y_{ji}$  is not continuous.

We can express the joint density of the main outcome and the endogenous covariates in terms of the marginal and conditional densities. The denominator in (11) or (12) in the higher stage will cancel out the numerator of (11) or (12) in the lower stage, so we have

$$f(y_{1i}, \dots, y_{ji}|\mathbf{z}_i) = \int_{\mathbf{V}_{ji}^*} \phi_j(v_{1i}, \dots, v_{ji}, \Sigma_j) d\mathbf{v}_{ji}^* \quad (13)$$

If we have only continuous endogenous variables, we have

$$f(y_{1i}, \dots, y_{ji}|\mathbf{z}_i) = \phi_j(v_{1i}, \dots, v_{ji}, \Sigma_j)$$

If  $\mathbf{V}_{ji}^*$  has dimension  $j$ , we can calculate the integral given in (13) by using the  $\Phi_j^*$ . Let  $\mathbf{l}_i$  contain the lower endpoints and  $\mathbf{u}_i$  contain the upper endpoints for  $\mathbf{V}_{ji}^*$ . When we do not have continuous endogenous covariates, we have

$$f(y_{1i}, \dots, y_{ji}|\mathbf{z}_i) = \Phi_j^*(\mathbf{l}_i, \mathbf{u}_i, \Sigma_j)$$

Now suppose that we have  $C < j$  continuous outcomes in  $y_{1i}, \dots, y_{ji}$ , so the dimension of  $\mathbf{V}_{ji}^*$  is  $j - C$ . Without loss of generality, these  $C$  correspond to the last  $C$  endogenous covariates  $y_{(j-C+1)i}, \dots, y_{ji}$ . The covariates can be reordered as needed.

We partition the covariance

$$\Sigma_j = \begin{bmatrix} \Sigma_{11} & \Sigma'_{12} \\ \Sigma_{12} & \Sigma_{22} \end{bmatrix}$$

where  $\Sigma_{22}$  is the covariance of the last  $C$  errors.

Conditional on  $v_{(j-C+1)i}, \dots, v_{ji}$ , the errors  $v_{1i}, \dots, v_{(j-C)i}$  have mean and variance

$$\begin{aligned} \mu_{1|2,i} &= \Sigma_{12} \Sigma_{22}^{-1} \begin{bmatrix} v_{(j-C+1)i} \\ \vdots \\ v_{ji} \end{bmatrix} \\ \Sigma_{1|2} &= \Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma'_{12} \end{aligned}$$

By conditioning on  $v_{(j-C+1)i}, \dots, v_{ji}$ , we can express the density in terms of  $\phi_C$  and  $\Phi_{j-C}^*$ . We can write the joint density in terms of the marginal and conditional densities to obtain

$$f(y_{1i}, \dots, y_{ji} | \mathbf{z}_i) = \phi_C(v_{(j-C+1)i}, \dots, v_{ji}, \Sigma_{22}) \Phi_{j-C}^*(\mathbf{I}_i - \mu_{1|2,i}, \mathbf{u}_i - \mu_{1|2,i}, \Sigma_{1|2})$$

The natural logarithm of the density  $f(y_{1i}, \dots, y_{ji} | \mathbf{z}_i)$  is the log likelihood of the model. We maximize the log likelihood to estimate the model parameters.

We can relax the assumption that the errors  $v_{1i}, \dots, v_{ji}$  are multivariate normal with mean 0 and covariance  $\Sigma$ . We will allow the covariance matrix to vary based on the  $M$  different levels of the binary or ordinal endogenous covariates  $\mathbf{w}_{poi}$ :  $\omega_1, \dots, \omega_M$ . These are the different combinations of values for the covariates  $\mathbf{w}_{poi}$ .

We use a potential-outcome framework for the outcome errors  $v_{Ji}$ . For the potential-outcome errors  $v_{1Ji}, \dots, v_{MJi}$ , we have

$$v_{Ji} = \sum_{m=1}^M \mathbf{1}(\mathbf{w}_{poi} = \omega_m) v_{mJi}$$

For  $m = 1, \dots, M$ ,  $v_{mJi}$  and  $v_{1i}, \dots, v_{Hi}$  are multivariate normal mean 0 and covariance

$$\Sigma_m = \begin{bmatrix} \sigma_m^2 & \sigma'_{mo} \\ \sigma_{mo} & \Sigma_o \end{bmatrix}$$

For observations where  $\mathbf{w}_{poi} = \omega_m$ , the log likelihood can be derived with  $\Sigma_m$  in place of  $\Sigma$ . The log likelihoods from the different potential-outcome group observations can then be summed together to get the log likelihood of the model.

Now we assume that we have random effects in each equation and a panel-data structure. This discussion applies to the models fit by `xteintreg`, `xteprobit`, `xteprobit`, and `xteregress`. For simplicity, we assume that the errors do not follow a potential-outcome framework. We have  $N$  panels. For panel  $i = 1, \dots, N$ , there are  $N_i$  observations, and for  $t = 1, \dots, N_i$ , we have

$$\begin{aligned}
 y_{1it} &= g_{1it}(\mathbf{w}_{1it}\boldsymbol{\beta}_1 + v_{1it} + u_{1i}) \\
 &\vdots \\
 y_{Hit} &= g_{Hit}(\mathbf{w}_{Hit}\boldsymbol{\beta}_H + v_{Hit} + u_{Hi}) \\
 y_{it} &= y_{Jit} = g_{Jit}(\mathbf{w}_{Jit}\boldsymbol{\beta}_J + v_{Jit} + u_{Ji})
 \end{aligned}$$

The observation-level errors  $v_{1it}, \dots, v_{Jit}$  are multivariate normal with mean 0 and covariance  $\boldsymbol{\Sigma}$ . They are independent of the panel-level errors, or random effects  $u_{1i}, \dots, u_{Ji}$ , which are multivariate normal with mean 0 and covariance  $\boldsymbol{\Sigma}_u$ . We further assume that the observation-level errors are independent within panels.

Now the covariates  $\mathbf{w}_{jit}$ , random effect  $u_{ji}$ , and the outcome  $y_{jit}$  determine a range for the error  $v_{jit}$ . For example, if  $y_{jit}$  has a linear model, then  $v_{jit} = y_{jit} - \mathbf{w}_{jit}\boldsymbol{\beta}_j - u_{ji}$ , the residual. If  $y_{jit} = 1$  and  $y_{jit}$  has a probit model, then  $v_{jit}$  is in the range  $(-\mathbf{w}_{jit}\boldsymbol{\beta}_j - u_{ji}, \infty)$ . If  $y_{jit}$  is left-censored at  $l_{it}$ , then  $v_{jit}$  is in the range  $(-\infty, l_{it} - \mathbf{w}_{jit}\boldsymbol{\beta}_j - u_{ji})$ .

Conditional on the random effects  $u_{1i}, \dots, u_{Ji}$ , the density of the endogenous variables can be represented using a multivariate normal density function that is evaluated at the residuals for the continuous outcomes and integrated over the error ranges of the noncontinuous outcomes. So the conditional density is formulated as in the cross-sectional case. The random effects are essentially added to the covariates  $\mathbf{w}_{1it}, \dots, \mathbf{w}_{Jit}$ .

Note that each panel has the same random effects for every observation. So if panel  $i$  has random effects  $\mathbf{u}_i = (u_{1i}, \dots, u_{Ji})$ , its likelihood is

$$L_i = \int_{\mathbb{R}^J} \left\{ \prod_{t=1}^{N_i} f(y_{1it}, \dots, y_{Jit} | \mathbf{z}_{it}, \mathbf{u}_i) \right\} \phi_J(\mathbf{u}_i, \boldsymbol{\Sigma}_u) d\mathbf{u}_i \quad (14)$$

This multivariate integral is generally not tractable. We can use a change-of-variables technique to transform the multivariate integral in (14) into a set of nested univariate integrals. Let  $\mathbf{L}$  be the Cholesky decomposition of  $\boldsymbol{\Sigma}_u$ ; that is,  $\boldsymbol{\Sigma}_u = \mathbf{L}\mathbf{L}'$ . It follows that  $\mathbf{u}_i = \mathbf{L}\boldsymbol{\psi}_i$ , where  $\boldsymbol{\psi}_i$  is a vector of independent standard normal random variables.

So we can rewrite (14) as

$$L_i = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \left\{ \prod_{t=1}^{N_i} f(y_{1it}, \dots, y_{Jit} | \mathbf{z}_{it}, \mathbf{u}_i = \mathbf{L}\boldsymbol{\psi}_i) \right\} \phi(\psi_{1i}) \dots \phi(\psi_{Ji}) d\psi_{1i} \dots d\psi_{Ji} \quad (15)$$

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  $(a_k^*, w_k^*)$ ,  $k = 1, \dots, q$ . Then, the GHQ approximation is

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

Consider a  $J$ -dimensional quadrature grid containing  $q$  quadrature points in each dimension. Let the vector of abscissas  $\mathbf{a}_k = (a_{k1}, \dots, a_{kJ})'$  be a point in this grid, and let  $\mathbf{w}_k = (w_{k1}, \dots, w_{kJ})'$  be the vector of corresponding weights. The GHQ approximation to the likelihood for a given panel is

$$L_i = \sum_{k_1=1}^q \dots \sum_{k_J=1}^q \left[ \left\{ \prod_{t=1}^{N_i} f(y_{1it}, \dots, y_{Jit} | \mathbf{z}_{it}, \mathbf{u}_i = \mathbf{L}\mathbf{a}_k) \right\} \left\{ \prod_{s=1}^J \omega_{k_s} \right\} \right] \quad (16)$$

Rather than using regular GHQ, we can use mean–variance adaptive Gauss–Hermite quadrature. Fixing the observed variables and model parameters in the integrand of (14), we see the posterior density for  $\psi_i$  is proportional to

$$\left\{ \prod_{t=1}^{N_i} f(y_{1it}, \dots, y_{Jit} | \mathbf{z}_{it}, \mathbf{u}_i = \mathbf{L}\psi_i) \right\} \phi(\psi_i)$$

It is reasonable to assume that this posterior density can be approximated by a multivariate normal density with mean vector  $\boldsymbol{\mu}_{vi}$  and variance matrix  $\boldsymbol{\tau}_{vi}$ . 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,

$$L_i = \int_{\mathbb{R}^J} \frac{\left\{ \prod_{t=1}^{N_i} f(y_{1it}, \dots, y_{Jit} | \mathbf{z}_{it}, \mathbf{u}_i = \mathbf{L}\psi_i) \right\} \phi(\psi_i)}{\phi(\psi_i, \boldsymbol{\mu}_{vi}, \boldsymbol{\tau}_{vi})} \phi(\psi_i, \boldsymbol{\mu}_{vi}, \boldsymbol{\tau}_{vi}) d\psi_i$$

The likelihood is then approximated by

$$L_i = \sum_{k_1=1}^q \dots \sum_{k_J=1}^q \left[ \left\{ \prod_{t=1}^{N_i} f(y_{1it}, \dots, y_{Jit} | \mathbf{z}_{it}, \mathbf{u}_i = \mathbf{L}\boldsymbol{\alpha}_k) \right\} \left\{ \prod_{s=1}^J \omega_{k_s} \right\} \right] \quad (17)$$

where  $\boldsymbol{\alpha}_k$  and  $\omega_{k_s}$  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  $\boldsymbol{\mu}_{vi}$  and posterior variances  $\boldsymbol{\tau}_{vi}$  are computed iteratively by updating the posterior moments by using the mean–variance adaptive Gauss–Hermite approximation, starting with a 0 mean vector and identity variance matrix.

Then, the log likelihood for all panels is

$$\ln L = \sum_{i=1}^N \left( \ln \sum_{k_1=1}^q \dots \sum_{k_J=1}^q \left[ \left\{ \prod_{t=1}^{N_i} f(y_{1it}, \dots, y_{Jit} | \mathbf{z}_{it}, \mathbf{u}_i = \mathbf{L}\boldsymbol{\alpha}_k) \right\} \left\{ \prod_{s=1}^J \omega_{k_s} \right\} \right] \right) \quad (18)$$

As in the cross-sectional case, we can relax the assumption that the errors  $v_{1it}, \dots, v_{Jit}$  are multivariate normal with mean 0 and covariance  $\boldsymbol{\Sigma}$ . We will allow the covariance matrix to vary based on the  $M$  different levels of the binary or ordinal endogenous covariates  $\mathbf{w}_{poit}$ :  $\boldsymbol{\omega}_1, \dots, \boldsymbol{\omega}_M$ . These are the different combinations of values for the covariates  $\mathbf{w}_{poit}$ .

We use a potential-outcome framework for the outcome errors  $v_{Jit}$ . For the potential-outcome errors  $v_{1Jit}, \dots, v_{MJit}$ , we have

$$v_{Jit} = \sum_{m=1}^M 1(\mathbf{w}_{poit} = \boldsymbol{\omega}_m) v_{mJit}$$

For  $m = 1, \dots, M$ ,  $v_{mJit}$  and  $v_{1it}, \dots, v_{Hit}$  are multivariate normal mean 0 and covariance

$$\Sigma_m = \begin{bmatrix} \sigma_m^2 & \sigma'_{mo} \\ \sigma_{mo} & \Sigma_o \end{bmatrix}$$

For observations where  $\mathbf{w}_{poit} = \boldsymbol{\omega}_m$ , the likelihood can be derived with  $\Sigma_m$  in place of  $\Sigma$ .

## References

- Amemiya, T. 1985. *Advanced Econometrics*. Cambridge, MA: Harvard University Press.
- Angrist, J. D. 2001. Estimation of limited dependent variable models with dummy endogenous regressors: Simple strategies for empirical practice. *Journal of Business and Economic Statistics* 19: 2–28. <https://doi.org/10.1198/07350010152472571>.
- Arendt, J. N., and A. Holm. 2006. Probit models with binary endogenous regressors. Discussion Papers on Business and Economics No. 4/2006, Department of Business and Economics, University of Southern Denmark, Odense, Demark. [https://www.sdu.dk/da/om\\_sdu/institutter\\_centre/ivoe\\_virksomhedsledelse\\_og\\_oekonomi/forskning/forskningspublikationer/~media/Files/Om\\_SDU/Institutter/Ivoe/Disc\\_papers/Disc\\_2006/dpbe4%202006%20pdf.ashx](https://www.sdu.dk/da/om_sdu/institutter_centre/ivoe_virksomhedsledelse_og_oekonomi/forskning/forskningspublikationer/~media/Files/Om_SDU/Institutter/Ivoe/Disc_papers/Disc_2006/dpbe4%202006%20pdf.ashx).
- Bartus, T., and D. Roodman. 2014. Estimation of multiprocess survival models with cmp. *Stata Journal* 14: 756–777.
- Conway, M. R. 1990. A random effects model for binary data. *Biometrics* 46: 317–328. <https://doi.org/10.2307/2531437>.
- Drezner, Z. 1994. Computation of the trivariate normal integral. *Mathematics of Computation* 62: 289–294. <http://doi.org/10.2307/2153409>.
- Hasebe, T. 2022. Endogenous models of binary choice outcomes: Copula-based maximum-likelihood estimation and treatment effects. *Stata Journal* 22: 734–771.
- Heckman, J. 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.
- . 1979. Sample selection bias as a specification error. *Econometrica* 47: 153–161. <https://doi.org/10.2307/1912352>.
- Karymshakov, K., K. Sultakeev, and B. Sulaimanova. 2015. The impact of microfinance on entrepreneurship in Kyrgyzstan. *International Conference on Eurasian Economics*, paper ID 1412. Eurasian Economists Association: Kazan, Russia. <http://www.avekon.org/papers/1412.pdf>.
- Maddala, G. S. 1983. *Limited-Dependent and Qualitative Variables in Econometrics*. Cambridge: Cambridge University Press.
- Maddala, G. S., and L.-F. Lee. 1976. Recursive models with qualitative endogenous variables. *Annals of Economic and Social Measurement* 5: 525–545.
- Miwa, T., A. J. Hayter, and S. Kuriki. 2003. The evaluation of general non-centred orthant probabilities. *Journal of the Royal Statistical Society, Series B* 65: 223–234. <https://doi.org/10.1111/1467-9868.00382>.
- Miwa, T., A. J. Hayter, and W. Liu. 2000. Calculations of level probabilities for normal random variables with unequal variances with applications to Bartholomew’s test in unbalanced one-way models. *Computational Statistics and Data Analysis* 34: 17–32. [https://doi.org/10.1016/S0167-9473\(99\)00075-4](https://doi.org/10.1016/S0167-9473(99)00075-4).
- Mulkay, B. 2015. Bivariate probit estimation for panel data: A two-step Gauss–Hermite quadrature approach with an application to product and process innovations for France. Working paper, Université de Montpellier, Faculté d’Economie, Montpellier, France. <https://afse2015.sciencesconf.org/61055/document>.
- Naylor, J. C., and A. F. M. Smith. 1982. Applications of a method for the efficient computation of posterior distributions. *Journal of the Royal Statistical Society, Series C* 31: 214–225. <https://doi.org/10.2307/2347995>.
- Newey, W. K. 1987. Efficient estimation of limited dependent variable models with endogenous explanatory variables. *Journal of Econometrics* 36: 231–250. [https://doi.org/10.1016/0304-4076\(87\)90001-7](https://doi.org/10.1016/0304-4076(87)90001-7).
- Pindyck, R. S., and D. L. Rubinfeld. 1998. *Econometric Models and Economic Forecasts*. 4th ed. New York: McGraw–Hill.
- Roodman, D. 2011. Fitting fully observed recursive mixed-process models with cmp. *Stata Journal* 11: 159–206.
- Semykina, A., and J. M. Wooldridge. 2018. Binary response panel data models with sample selection and self-selection. *Journal of Applied Econometrics* 33: 179–197. <https://doi.org/10.1002/jae.2592>.

- Skroddal, A., and S. Rabe-Hesketh. 2004. *Generalized Latent Variable Modeling: Multilevel, Longitudinal, and Structural Equation Models*. Boca Raton, FL: Chapman and Hall/CRC.
- Vahter, P. 2011. Does FDI spur productivity, knowledge sourcing and innovation by incumbent firms? Evidence from manufacturing industry in Estonia. *World Economy* 34: 1308–1326. <http://doi.org/10.1111/j.1467-9701.2011.01379.x>.
- Van de Ven, W. P. M. M., and B. M. S. Van Pragg. 1981. The demand for deductibles in private health insurance: A probit model with sample selection. *Journal of Econometrics* 17: 229–252. [https://doi.org/10.1016/0304-4076\(81\)90028-2](https://doi.org/10.1016/0304-4076(81)90028-2).
- White, H. L., Jr. 1996. *Estimation, Inference and Specification Analysis*. Cambridge: Cambridge University Press.
- Wooldridge, J. M. 2010. *Econometric Analysis of Cross Section and Panel Data*. 2nd ed. Cambridge, MA: MIT Press.
- . 2014. Quasi-maximum likelihood estimation and testing for nonlinear models with endogenous explanatory variables. *Journal of Econometrics* 182: 226–234. <https://doi.org/10.1016/j.jeconom.2014.04.020>.

## Also see

- [ERM] **eprobit postestimation** — Postestimation tools for eprobit and xtprobit
- [ERM] **eprobit predict** — predict after eprobit and xtprobit
- [ERM] **predict advanced** — predict’s advanced features
- [ERM] **predict treatment** — predict for treatment statistics
- [ERM] **estat teffects** — Average treatment effects for extended regression models
- [ERM] **Intro 9** — Conceptual introduction via worked example
- [R] **biprobit** — Bivariate probit regression
- [R] **heckprobit** — Probit model with sample selection
- [R] **ivprobit** — Probit model with continuous endogenous covariates
- [R] **probit** — Probit regression
- [SVY] **svy estimation** — Estimation commands for survey data
- [XT] **xtprobit** — Random-effects and population-averaged probit models
- [U] **20 Estimation and postestimation commands**

Stata, Stata Press, and Mata are registered trademarks of StataCorp LLC. Stata and Stata Press are registered trademarks with the World Intellectual Property Organization of the United Nations. Other brand and product names are registered trademarks or trademarks of their respective companies. Copyright © 1985–2023 StataCorp LLC, College Station, TX, USA. All rights reserved.

