Title

stata.com

eprobit — Extended probit regression

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.

Quick start

Probit regression of `y` on `x` with continuous endogenous covariate `y2` modeled by `x` and `z`

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

As above, but adding continuous endogenous covariate `y3` modeled by `x` and `z2`

```
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)
```

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

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

As above, but adding endogenous covariate `y2` modeled by `x` and `z2`

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

As above, but adding endogenous treatment recorded in `trtvar` and modeled by `x` and `z3`

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

Menu

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

Syntax

Basic probit regression with endogenous covariates

`eprobit` *depvar* [*indepvars*] ,
  `endogenous(`*depvars*$_{\mathrm{en}}$ = *varlist*$_{\mathrm{en}}$`)` [*options*]

Basic probit regression with endogenous treatment assignment

`eprobit` *depvar* [*indepvars*] ,
  `entreat(`*depvar*$_{\mathrm{tr}}$ [= *varlist*$_{\mathrm{tr}}$]`)` [*options*]

Basic probit regression with exogenous treatment assignment

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

Basic probit regression with sample selection

`eprobit` *depvar* [*indepvars*] ,
  `select(`*depvar*$_s$ = *varlist*$_s$`)` [*options*]

Basic probit regression with tobit sample selection

`eprobit` *depvar* [*indepvars*] ,
  `tobitselect(`*depvar*$_s$ = *varlist*$_s$`)` [*options*]

Probit regression combining endogenous covariates, treatment, and selection

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

extensions	Description
Model	
endogenous(*enspec*)	model for endogenous covariates; may be repeated
entreat(*entrspec*)	model for endogenous treatment assignment
extreat(*extrspec*)	exogenous treatment
select(*selspec*)	probit model for selection
tobitselect(*tselspec*)	tobit model for selection

options	Description
Model	
noconstant	suppress constant term
offset(*varname*$_o$)	include *varname*$_o$ in model with coefficient constrained to 1
constraints(*numlist*)	apply specified linear constraints
collinear	keep collinear variables
SE/Robust	
vce(*vcetype*)	*vcetype* may be oim, robust, cluster *clustvar*, opg, bootstrap, or jackknife
Reporting	
level(#)	set confidence level; default is level(95)
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	
intpoints(#)	set the number of integration (quadrature) points for integration over four or more dimensions; default is intpoints(128)
triintpoints(#)	set the number of integration (quadrature) points for integration over three dimensions; default is triintpoints(10)
Maximization	
maximize_options	control the maximization process; seldom used
coeflegend	display legend instead of statistics

enspec is *depvars*$_{\text{en}}$ = *varlist*$_{\text{en}}$ [, *enopts*]

where *depvars*$_{\text{en}}$ is a list of endogenous covariates. Each variable in *depvars*$_{\text{en}}$ specifies an endogenous covariate model using the common *varlist*$_{\text{en}}$ and options.

entrspec is *depvar*$_{\text{tr}}$ [= *varlist*$_{\text{tr}}$] [, *entropts*]

where *depvar*$_{\text{tr}}$ is a variable indicating treatment assignment. *varlist*$_{\text{tr}}$ is a list of covariates predicting treatment assignment.

extrspec is *tvar* [, *extropts*]

where *tvar* is a variable indicating treatment assignment.

selspec is *depvar*$_s$ = *varlist*$_s$ [, `noconstant` `offset(`*varname*$_o$`)`]

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

tselspec is *depvar*$_s$ = *varlist*$_s$ [, *tselopts*]

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

enopts	Description
Model	
`probit`	treat endogenous covariate as binary
`oprobit`	treat endogenous covariate as ordinal
`pocorrelation`	estimate different correlations for each level of a binary or an ordinal endogenous covariate
`nomain`	do not add endogenous covariate to main equation
`noconstant`	suppress constant term

entropts	Description
Model	
`pocorrelation`	estimate different correlations for each potential outcome
`nomain`	do not add treatment indicator to main equation
`nointeract`	do not interact treatment with covariates in main equation
`noconstant`	suppress constant term
`offset(`*varname*$_o$`)`	include *varname*$_o$ in model with coefficient constrained to 1

extropts	Description
Model	
`pocorrelation`	estimate different correlations for each potential outcome
`nomain`	do not add treatment indicator to main equation
`nointeract`	do not interact treatment with covariates in main equation

tselopts	Description
Model	
`ll(`*varname* | *#*`)`	left-censoring variable or limit
`ul(`*varname* | *#*`)`	right-censoring variable or limit
`main`	add censored selection variable to main equation
`noconstant`	suppress constant term
`offset(`*varname*$_o$`)`	include *varname*$_o$ in model with coefficient constrained to 1

indepvars, $varlist_{en}$, $varlist_{tr}$, and $varlist_s$ may contain factor variables; see [U] **11.4.3 Factor variables**.

depvar, *indepvars*, $depvars_{en}$, $varlist_{en}$, $depvar_{tr}$, $varlist_{tr}$, *tvar*, $depvar_s$, and $varlist_s$ may contain time-series operators; see [U] **11.4.4 Time-series varlists**.

`bootstrap`, `by`, `jackknife`, `rolling`, `statsby`, and `svy` are allowed; 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**.

`fweight`s, `iweight`s, and `pweight`s are allowed; see [U] **11.1.6 weight**.

`coeflegend` does not appear in the dialog box.

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

Options

Model

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

`noconstant`, `offset(`$varname_o$`)`, `constraints(`*numlist*`)`, `collinear`; see [R] **estimation options**.

SE/Robust

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

Reporting

`level(`*#*`)`, `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**.

Integration

`intpoints(`*#*`)`, `triintpoints(`*#*`)`; 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**.

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

The following option is available with `eprobit` but is not shown in the dialog box.

`coeflegend`; see [R] **estimation options**.

Remarks and examples

stata.com

`eprobit` fits models that we refer to as "extended probit regression models", meaning that they accommodate endogenous covariates, nonrandom treatment assignment, and endogenous sample selection. `eprobit` can account for these complications whether they arise individually or in combination.

In this entry, you will find information on the eprobit command syntax. You can see *Methods and formulas* for a full description of the models that can be fit with eprobit 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. Below, we provide a guide to help you locate the ones that will be helpful to you.

For an introduction to eprobit and the other extended regression commands (eregress, eintreg, and eoprobit), see [ERM] **intro 1**–[ERM] **intro 8**.

> [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** discusses interpretation of results. You can interpret coefficients from eprobit 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 7** will be particularly helpful if you are familiar with ivprobit, heckprobit, and other commands that address endogenous covariates, sample selection, or nonrandom treatment assignment. This introduction is a Rosetta stone that maps the syntax of those commands to the syntax of eprobit.
>
> [ERM] **intro 8** 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 6b**. For examples using eprobit, see

[ERM] **example 3a**	Probit regression with continuous endogenous covariate
[ERM] **example 3b**	Probit regression with endogenous covariate and treatment
[ERM] **example 4a**	Probit regression with endogenous sample selection
[ERM] **example 4b**	Probit regression with endogenous treatment and sample selection
[ERM] **example 5**	Probit regression with endogenous ordinal treatment

See *Examples* in [ERM] **intro** for an overview of all the examples. These examples demonstrate all four extended regression commands, and all may be interesting because they handle complications in the same way.

You can also find in literature discussion and examples of many models that eprobit can fit. This includes discussion of the probit model with continuous endogenous covariates (Newey 1987), the probit model with multiple endogenous binary covariates (Arendt and Holm 2006), and the probit model with an endogenous treatment (Angrist [2001] and Pindyck and Rubinfeld [1998]). 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). 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. His 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

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

Macros

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

Matrices

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

Functions

`e(sample)`	marks estimation sample

Methods and formulas

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

The log-likelihood function maximized by `eprobit` is 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) used this result to derive the formulas discussed below.

Methods and formulas are presented under the following headings:

Introduction

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

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

where the errors ϵ_i are distributed as standard normal. The log likelihood is

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

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

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

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 Φ_d^* with three inputs. The first two inputs are d-dimensional row vectors $\mathbf{l}$ and $\mathbf{u}$ that have values in $I\!R \cup \{-\infty, \infty\}$, the extended real line. The final input is a $d \times d$ real-valued and positive-definite matrix $\mathbf{\Sigma}$.

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

where ϕ_d is the density of a mean 0, multivariate normal random variable. For details on the calculation of Φ_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 ϵ_i are based on the observed values of y_i and $\mathbf{x}_i$ and are defined as

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

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:

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

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

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

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

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

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 ϵ_i. Thus, the conditional distribution of ϵ_i can be used to find the conditional density of y_i. Conditional on the endogenous and exogenous covariates, ϵ_i has mean and variance

$$\begin{aligned} E\left(\epsilon_i|\mathbf{w}_{ci},\mathbf{x}_i,\mathbf{z}_{ci}\right) &= \boldsymbol{\sigma}'_{1c}\boldsymbol{\Sigma}_c^{-1}\left(\mathbf{w}_{ci} - \mathbf{z}_{ci}\mathbf{A}_c\right)' \\ \mathrm{Var}\left(\epsilon_i|\mathbf{w}_{ci},\mathbf{x}_i,\mathbf{z}_{ci}\right) &= 1 - \boldsymbol{\sigma}'_{1c}\boldsymbol{\Sigma}_c^{-1}\boldsymbol{\sigma}_{1c} \end{aligned}$$

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

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

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

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

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

Letting

$$\begin{aligned} l_{1i1} &= -\mathbf{x}_i\boldsymbol{\beta} - \boldsymbol{\sigma}'_{1c}\boldsymbol{\Sigma}_c^{-1}\left(\mathbf{w}_{ci} - \mathbf{z}_{ci}\mathbf{A}_c\right)' \\ u_{1i1} &= \infty \end{aligned}$$

the conditional probability of success is

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

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 $\mathbf{w}_{bi} = [w_{b1i}, \ldots, w_{bBi}]$. 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 $j = 1, \ldots, B$. We use a probit model for binary endogenous covariates

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

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

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

The values $v_{bj1}, \ldots, v_{bjB_j}$ are real numbers such that $v_{bjh} < v_{bjm}$ for $h < m$. κ_{bj0} is taken as $-\infty$ and κ_{bjB_j} is taken as $+\infty$. The errors $\epsilon_{b1i}, \ldots, \epsilon_{bBi}$ 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.

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

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 ϵ_i and binary and ordinal endogenous errors $\epsilon_{b1i}, \ldots, \epsilon_{bBi}$ 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 $j = 1, \ldots, B$ and $h = 0, \ldots, B_j$, let

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

The probability for w_{bji} has lower limit

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

and upper limit

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

Letting

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

the lower and upper limits for the y_i probability are

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

and

$$\mathbf{l}_i = [\, l_{1i} \quad l_{b1i} \quad \ldots \quad l_{bBi} \,]$$
$$\mathbf{u}_i = [\, u_{1i} \quad u_{b1i} \quad \ldots \quad u_{bBi} \,]$$

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

$$\mathbf{l}_{bi} = [\, l_{b1i} \quad \ldots \quad l_{bBi} \,]$$
$$\mathbf{u}_{bi} = [\, u_{b1i} \quad \ldots \quad u_{bBi} \,]$$
$$\mathbf{l}_{i1} = [\, -\infty \quad \mathbf{l}_{bi} \,]$$
$$\mathbf{u}_{i1} = [\, c_{bi} \quad \mathbf{u}_{bi} \,]$$

The conditional probability of success is

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

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 2×12 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 $\mathbf{v}_j$ ($j \in \{1, 2, \ldots, M\}$). Letting k_{wp} be the number of levels of endogenous ordinal treatment variable $p \in \{1, 2, \ldots, B\}$ implies that $M = k_{w1} \times k_{w2} \times \cdots \times k_{wB}$.

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

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

For $j = 1, \ldots, M$, the outcome error ϵ_{ji} and the endogenous errors $\epsilon_{b1i}, \ldots, \epsilon_{bBi}$ 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

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

The conditional probability of success is

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

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: $y_{1i}, \ldots, y_{Ti}$.

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

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

So for each observation, we only observe 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 [TE] **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

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

The treatment values $v_1, \ldots, v_T$ are real numbers such that $v_h < v_m$ for $h < m$. κ_0 is taken as $-\infty$ and κ_T is taken as $+\infty$. The treatment error ϵ_{ti} is standard normal.

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

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

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

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

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 $j = 1, \ldots, T$, ϵ_{ji} and ϵ_{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 Φ_2^* function.

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

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

The lower and upper limits for the y_i probability are

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

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

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

So for the t_i probability, we have lower limit

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

and upper limit

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

The log likelihood for the model is

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

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

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

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

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

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

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

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

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

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

$$\boldsymbol{\Sigma}_j = \begin{bmatrix} 1 & \rho_{j1t} \\ \rho_{j1t} & 1 \end{bmatrix}$$

Now, define

$$\boldsymbol{\Sigma}_i = \sum_{j=1}^{T} 1(t_i = v_j)\boldsymbol{\Sigma}_j$$

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

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

The conditional probability of success at treatment level v_j is

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

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

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

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

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

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 ϵ_i and ϵ_{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

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

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

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

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

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

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

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

$$s_i = \begin{cases} l_i & s_i^\star \le l_i \\ \\ s_i^\star & l_i < s_i^\star < u_i \\ \\ u_i & s_i^\star \ge u_i \end{cases}$$

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 $s_i^\star$ is not censored ($l_i < s_i^\star < u_i$). The outcome y_i is not observed when $s_i^\star$ is left-censored ($s_i^\star \le l_i$) or $s_i^\star$ is right-censored ($s_i^\star \ge u_i$). The unobserved errors ϵ_i and ϵ_{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

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

and

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

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

$$\begin{aligned} l_{li} &= -\infty \\ u_{li} &= \frac{l_i - \mathbf{z}_{si}\boldsymbol{\alpha}_s}{\sigma_s} \end{aligned}$$

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

$$\begin{aligned} l_{ui} &= \frac{u_i - \mathbf{z}_{si}\boldsymbol{\alpha}_s}{\sigma_s} \\ u_{ui} &= \infty \end{aligned}$$

Now, we consider the selected observations. For $s_i = s_i^\star = S_i$, 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

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

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

$$E\left(\epsilon_i|s_i = S_i, \mathbf{x}_i, \mathbf{z}_{si}\right) = \rho_{1s}\sigma_s^{-1}\left(S_i - \mathbf{z}_{si}\boldsymbol{\alpha}\right)$$
$$\text{Var}\left(\epsilon_i|s_i = S_i, \mathbf{x}_i, \mathbf{z}_{si}\right) = 1 - \rho_{1,s}^2$$

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

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

It follows that the log likelihood is

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

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

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

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

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

Combined model

The probit model with continuous endogenous covariates, ordinal endogenous covariates, an ordinal endogenous treatment, and endogenous sample selection combines all the extensions to the standard probit model that are supported by `eprobit`. The formulation of other combinations of model features can be easily derived from this combined model. In *Likelihood for multiequation models*, we describe the general framework for ERMs with multiple 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: $y_{1i}, \ldots, y_{Ti}$. The relationship between the ordinal treatment t_i, treatment covariates $\mathbf{z}_{t,i}$, and error ϵ_{ti} is described in (7). For $j = 1, \ldots, B$, the relationship between the ordinal endogenous covariates w_{bji}, exogenous covariates $\mathbf{z}_{bji}$, and error ϵ_{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 $\mathbf{w}_{bi} = [w_{b1i}, \ldots, w_{bBi}]$ with endogenous treatment t_i and endogenous sample selection on s_i has the form

$$
\begin{aligned}
y_{1i} &= 1\left(\mathbf{x}_i\boldsymbol{\beta}_1 + \mathbf{w}_{ci}\boldsymbol{\beta}_{c1} + \mathbf{wind}_{b1i}\boldsymbol{\beta}_{b11} + \cdots + \mathbf{wind}_{bBi}\boldsymbol{\beta}_{bB1} + \epsilon_{1i} > 0\right) \\
&\ \ \vdots \\
y_{Ti} &= 1\left(\mathbf{x}_i\boldsymbol{\beta}_T + \mathbf{w}_{ci}\boldsymbol{\beta}_{cT} + \mathbf{wind}_{b1i}\boldsymbol{\beta}_{b1T} + \cdots + \mathbf{wind}_{bBi}\boldsymbol{\beta}_{bBT} + \epsilon_{Ti} > 0\right) \\
y_i &= \sum_{j=1}^{T} 1(t_i = v_j) y_{ji} \\
\mathbf{w}_{ci} &= \mathbf{z}_{ci}\mathbf{A}_c + \boldsymbol{\epsilon}_{ci} \\
s_i &= 1\left(\mathbf{z}_{si}\boldsymbol{\alpha}_s + \epsilon_{si} > 0\right)
\end{aligned}
$$

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

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

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

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

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

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

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

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

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

The lower and upper limits for the y_i probability are

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

The conditional means of the unobserved errors ϵ_{si}, ϵ_{ti}, $\epsilon_{b1i}, \ldots, \epsilon_{bBi}$ have similar forms to e_{1i}. Denote these means by e_{si}, e_{ti}, $e_{b1i}, \ldots, e_{bBi}$. 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 $j = 1, \ldots, T$, conditional on $\mathbf{w}_{ci}$ and the exogenous covariates, $\epsilon_{ji}, \epsilon_{si}, \epsilon_{ti}, \epsilon_{b1i}, \ldots, \epsilon_{bBi}$ 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

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

The log likelihood of the model is

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

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

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

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

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

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

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

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

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 $\widehat{\sigma}^2$ be a point estimate for the variance parameter σ^2, and let $\widehat{\mathrm{SE}}(\widehat{\sigma}^2)$ be its standard error. The $(1-\alpha) \times 100\%$ confidence interval for $\ln(\sigma^2)$ is

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

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 σ^2 is then given by

$$\left(e^{k_l}, e^{k_u}\right)$$

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 $\widehat{\rho}$ be a point estimate for the correlation parameter ρ, and let $\widehat{\mathrm{SE}}(\widehat{\rho})$ be its standard error. The $(1-\alpha)\times 100\%$ confidence interval for `atanh`(ρ) is

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

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 ρ is then given by

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

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.

Here, we show how the log likelihood is formulated when we have multiple auxiliary equations.

Suppose that we have H auxiliary equations with endogenous outcomes $y_{1i},\ldots,y_{Hi}$. 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 $j=2,\ \ldots,\ J$, equation j depends only on the exogenous covariates $\mathbf{z}_i$ and the endogenous covariates from equation $h=j-1$ and below $y_{1i},\ \ldots,\ y_{hi}$. These are stored together in $\mathbf{w}_{ji}$.

So we have

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

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 $v_{1i},\ \ldots,\ v_{Ji}$ are multivariate normal with mean 0 and covariance $\boldsymbol{\Sigma}$.

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 $v_{ji}=y_{ji}-\mathbf{w}_{ji}\boldsymbol{\beta}_j$, the residual. If $y_{ji}=1$ and y_{ji} has a probit model, then v_{ji} is in the range $(-\mathbf{w}_{ji}\boldsymbol{\beta}_j,\infty)$. If y_{ij} is left-censored at l_i, then v_{ji} is in the range $(-\infty, l_i-\mathbf{w}_{ji}\boldsymbol{\beta}_j)$.

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

$$f(v_{ji}|\mathbf{w}_{ji}) = \frac{\int_{\mathbf{V}^\star_{hi}} \phi_j(v_{1i},\ldots,v_{ji},\boldsymbol{\Sigma}_j)d\mathbf{v}^\star_{hi}}{\int_{\mathbf{V}^\star_{hi}} \phi_h(v_{1i},\ldots,v_{hi},\boldsymbol{\Sigma}_h)d\mathbf{v}^\star_{hi}}$$

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

If y_{ji} is continuous, then

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

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

$$f(y_{ji}|\mathbf{w}_{ji}) = \frac{\int_{\mathbf{V}^\star_{ji}} \phi_j(v_{1i},\ldots,v_{ji},\boldsymbol{\Sigma}_j)d\mathbf{v}^\star_{ji}}{\int_{\mathbf{V}^\star_{hi}} \phi_h(v_{1i},\ldots,v_{hi},\boldsymbol{\Sigma}_h)d\mathbf{v}^\star_{hi}} \tag{12}$$

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

$$f(y_{1i},\ldots,y_{ji}|\mathbf{z}_i) = \int_{\mathbf{V}^\star_{ji}} \phi_j(v_{1i},\ldots,v_{ji},\boldsymbol{\Sigma}_j)d\mathbf{v}^\star_{ji} \tag{13}$$

If we only have continuous endogenous variables, we have

$$f(y_{1i},\ldots,y_{ji}|\mathbf{z}_i) = \phi_j(v_{1i},\ldots,v_{ji},\boldsymbol{\Sigma}_j)$$

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

$$f(y_{1i},\ldots,y_{ji}|\mathbf{z}_i) = \Phi^*_j(\mathbf{l}_i,\mathbf{u}_i,\boldsymbol{\Sigma}_j)$$

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

We partition the covariance

$$\boldsymbol{\Sigma}_j = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}'_{12} \\ \boldsymbol{\Sigma}_{12} & \boldsymbol{\Sigma}_{22} \end{bmatrix}$$

where $\boldsymbol{\Sigma}_{22}$ is the covariance of the last C errors.

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

$$\boldsymbol{\mu}_{1|2,i} = \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\begin{bmatrix} v_{(j-C+1)i} \\ \vdots \\ v_{ji} \end{bmatrix}$$

$$\boldsymbol{\Sigma}_{1|2} = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}'_{12}$$

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

$$f(y_{1i},\ldots,y_{ji}|\mathbf{z}_i) = \phi_C(v_{(j-C+1)i},\ldots,v_{ji},\mathbf{\Sigma}_{22})\Phi^*_{j-C}(\mathbf{l}_i - \boldsymbol{\mu}_{1|2,i}, \mathbf{u}_i - \boldsymbol{\mu}_{1|2,i}, \mathbf{\Sigma}_{1|2})$$

The natural logarithm of the density $f(y_{1i},\ldots,y_{Ji}|\mathbf{z}_i)$ 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}, $\ldots$, v_{Ji} are multivariate normal with mean 0 and covariance $\mathbf{\Sigma}$. We will allow the covariance matrix to vary based on the M different levels of the binary or ordinal endogenous covariates $\mathbf{w}_{poi}$: $\boldsymbol{\omega}_1$, $\ldots$,$\boldsymbol{\omega}_M$. 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}, …, $_{MJi}$, we have

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

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

$$\mathbf{\Sigma}_m = \begin{bmatrix} \sigma^2_m & \boldsymbol{\sigma}'_{mo} \\ \boldsymbol{\sigma}_{mo} & \mathbf{\Sigma}_o \end{bmatrix}$$

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

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: 2–16.

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. static.sdu.dk/mediafiles/Files/Om_SDU/Institutter/Ivoe/Disc_papers/Disc_2006/dpbe4 2006 pdf.pdf.

Drezner, Z. 1994. Computation of the trivariate normal integral. *Mathematics of Computation* 62: 289–294.

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

——. 1979. Sample selection bias as a specification error. *Econometrica* 47: 153–161.

Karymshakov, K., K. Sultakeev, and B. Sulaimanova. 2015. The impact of microfinance on entrepreneurship in Kyrgyzstan. *International Conference on Eurasian Economies*, 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.

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 & Data Analysis* 34: 17–32.

Newey, W. K. 1987. Efficient estimation of limited dependent variable models with endogenous explanatory variables. *Journal of Econometrics* 36: 231–250.

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.

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.

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.

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.

Also see

[ERM] **eprobit postestimation** — Postestimation tools for eprobit

[ERM] **eprobit predict** — predict after eprobit

[ERM] **estat teffects** — Average treatment effects for extended regression models

[ERM] **intro 8** — 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

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