etregress — Linear regression with endogenous treatment effects

Syntax

Basic syntax

`etregress` *depvar* [*indepvars*] `,` `treat(`*depvar*$_t$ `=` *indepvars*$_t$`)` [`twostep`]

Full syntax for maximum likelihood estimates only

`etregress` *depvar* [*indepvars*] [*if*] [*in*] [*weight*] `,`
  `treat(`*depvar*$_t$ `=` *indepvars*$_t$ [`, noconstant`]`)` [*etregress_ml_options*]

Full syntax for two-step consistent estimates only

`etregress` *depvar* [*indepvars*] [*if*] [*in*] `,`
  `treat(`*depvar*$_t$ `=` *indepvars*$_t$ [`, noconstant`]`)` `twostep` [*etregress_ts_options*]

etregress_ml_options	Description
Model	
* `treat()`	equation for treatment effects
`noconstant`	suppress constant term
`constraints(`*constraints*`)`	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)`
`first`	report first-step probit estimates
`noskip`	perform likelihood-ratio test
`hazard(`*newvar*`)`	create *newvar* containing hazard from treatment equation
`nocnsreport`	do not display constraints
display_options	control column formats, row spacing, line width, display of omitted variables and base and empty cells, and factor-variable labeling
Maximization	
maximize_options	control the maximization process; seldom used
`coeflegend`	display legend instead of statistics

*`treat(`*depvar*$_t$ `=` *indepvars*$_t$ [`, noconstant`]`)` is required.

etregress_ts_options	Description
Model	
* treat()	equation for treatment effects
* twostep	produce two-step consistent estimate
noconstant	suppress constant term
SE	
vce(*vcetype*)	*vcetype* may be conventional, bootstrap, or jackknife
Reporting	
level(#)	set confidence level; default is level(95)
first	report first-step probit estimates
hazard(*newvar*)	create *newvar* containing hazard from treatment equation
display_options	control column formats, row spacing, line width, display of omitted variables and base and empty cells, and factor-variable labeling
coeflegend	display legend instead of statistics

*treat(*depvar*$_t$ = *indepvars*$_t$ [, noconstant]) and twostep are required.

indepvars and *indepvars*$_t$ may contain factor variables; see **[U] 11.4.3 Factor variables**.

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

bootstrap, by, fp, 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**.

aweights are not allowed with the jackknife prefix; see **[R] jackknife**.

twostep, vce(), first, noskip, hazard(), and weights are not allowed with the svy prefix; see **[SVY] svy**.

pweights, aweights, fweights, and iweights are allowed with maximum likelihood estimation; see **[U] 11.1.6 weight**. No weights are allowed if twostep is specified.

coeflegend does not appear in the dialog box.

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

Menu

Statistics > Treatment effects > Endogenous treatment estimators > Continuous outcome

Description

etregress estimates an average treatment effect (ATE) and the other parameters of a linear regression model augmented with an endogenous binary-treatment variable. Estimation is by either full maximum likelihood or a two-step consistent estimator.

In addition to the ATE, etregress can be used to estimate the average treatment effect on the treated (ATET) when the outcome may not be conditionally independent of the treatment.

etreg is a synonym for etregress.

Options for maximum likelihood estimates

Model

treat(*depvar*$_t$ = *indepvars*$_t$ [, noconstant]) specifies the variables and options for the treatment equation. It is an integral part of specifying a treatment-effects model and is required.

noconstant, constraints(*constraints*), collinear; see [R] **estimation options**.

SE/Robust

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

Reporting

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

first specifies that the first-step probit estimates of the treatment equation be displayed before estimation.

noskip specifies that a full maximum-likelihood model with only a constant for the regression equation be fit. This model is not displayed but is used as the base model to compute a likelihood-ratio test for the model test statistic displayed in the estimation header. By default, the overall model test statistic is an asymptotically equivalent Wald test that all the parameters in the regression equation are zero (except the constant). For many models, this option can substantially increase estimation time.

hazard(*newvar*) will create a new variable containing the hazard from the treatment equation. The hazard is computed from the estimated parameters of the treatment equation.

nocnsreport; see [R] **estimation options**.

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

Maximization

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

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

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

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

Options for two-step consistent estimates

Model

treat(*depvar*$_t$ = *indepvars*$_t$ [, noconstant]) specifies the variables and options for the treatment equation. It is an integral part of specifying a treatment-effects model and is required.

`twostep` specifies that two-step consistent estimates of the parameters, standard errors, and covariance matrix be produced, instead of the default maximum likelihood estimates.

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

SE

`vce(`*vcetype*`)` specifies the type of standard error reported, which includes types that are derived from asymptotic theory (`conventional`) and that use bootstrap or jackknife methods (`bootstrap`, `jackknife`); see [R] ***vce_option***.

`vce(conventional)`, the default, uses the conventionally derived variance estimator for the two-step estimator of the treatment-effects model.

Reporting

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

`first` specifies that the first-step probit estimates of the treatment equation be displayed before estimation.

`hazard(`*newvar*`)` will create a new variable containing the hazard from the treatment equation. The hazard is computed from the estimated parameters of the treatment equation.

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

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

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

Remarks and examples

stata.com

Remarks are presented under the following headings:

Overview
Basic example
Average treatment effect (ATE)
Average treatment effect on the treated (ATET)

Overview

`etregress` estimates an ATE and the other parameters of a linear regression model that also includes an endogenous binary-treatment variable. In addition to the ATE, the parameters estimated by `etregress` can be used to estimate the ATET when the outcome is not conditionally independent of the treatment.

We call the model fit by `etregress` an endogenous treatment-regression model, although it is also known as an endogenous binary-variable model or as an endogenous dummy-variable model. The endogenous treatment-regression model is a specific endogenous treatment-effects model; it uses a linear model for the outcome and a constrained normal distribution to model the deviation from the conditional independence assumption imposed by the estimators implemented in `teffects`; see [TE] **teffects intro**. In treatment-effects jargon, the endogenous binary-variable model is a linear potential-outcome model that allows for a specific correlation structure between the unobservables that affect the treatment and the unobservables that affect the potential outcomes. See [TE] **etpoisson** for an estimator that allows for a nonlinear outcome model and a similar model for the endogeneity of the treatment.

Heckman (1976, 1978) brought this model into the modern literature. Maddala (1983) derives the maximum likelihood and two-step estimators of the version implemented here, reviews some empirical applications of this model, and describes it as a constrained endogenous-switching model. Barnow, Cain, and Goldberger (1981) provide another useful derivation of this model. They concentrate on deriving the conditions for which the self-selection bias of the simple OLS estimator of the treatment effect, δ, is nonzero and of a specific sign. Cameron and Trivedi (2005, sec. 16.7 and 25.3.4) and Wooldridge (2010, sec. 21.4.1) discuss the endogenous binary-variable model as an endogenous treatment-effects model and link it to recent work.

More formally, the endogenous treatment-regression model is composed of an equation for the outcome y_j and an equation for the endogenous treatment t_j,

$$
\begin{aligned}
y_j &= \mathbf{x}_j\boldsymbol{\beta} + \delta t_j + \epsilon_j \\
t_j &= \begin{cases} 1, & \text{if } \mathbf{w}_j\boldsymbol{\gamma} + u_j > 0 \\ 0, & \text{otherwise} \end{cases}
\end{aligned}
$$

where $\mathbf{x}_j$ are the covariates used to model the outcome, $\mathbf{w}_j$ are the covariates used to model treatment assignment, and the error terms ϵ_j and u_j are bivariate normal with mean zero and covariance matrix

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

The covariates $\mathbf{x}_j$ and $\mathbf{w}_j$ are unrelated to the error terms; in other words, they are exogenous.

Basic example

When there are no interactions between the treatment variable and the outcome covariates, `etregress` directly estimates the ATE and the ATET.

▷ Example 1

We estimate the ATE of being a union member on wages of women in 1972 from a nonrepresentative extract of the National Longitudinal Survey on young women who were ages 14–26 in 1968. We will use the variables `wage` (wage), `grade` (years of schooling completed), `smsa` (an indicator for living in an SMSA—standard metropolitan statistical area), `black` (an indicator for being African-American), `tenure` (tenure at current job), and `south` (an indicator for living in the South).

We use `etregress` to estimate the parameters of the endogenous treatment-regression model.

```
. use http://www.stata-press.com/data/r13/union3
(National Longitudinal Survey.  Young Women 14-26 years of age in 1968)
. etregress wage age grade smsa black tenure, treat(union = south black tenure)
Iteration 0:   log likelihood = -3097.9871
Iteration 1:   log likelihood = -3052.5988
Iteration 2:   log likelihood = -3051.5789
Iteration 3:   log likelihood =  -3051.575
Iteration 4:   log likelihood =  -3051.575
Linear regression with endogenous treatment       Number of obs    =       1210
Estimator: maximum likelihood                     Wald chi2(6)     =     681.89
Log likelihood =  -3051.575                       Prob > chi2      =     0.0000
```

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
wage						
age	.1487409	.0193291	7.70	0.000	.1108566	.1866252
grade	.4205658	.0293577	14.33	0.000	.3630258	.4781058
smsa	.9117045	.1249041	7.30	0.000	.6668969	1.156512
black	-.7882471	.1367078	-5.77	0.000	-1.056189	-.5203047
tenure	.1524015	.0369596	4.12	0.000	.0799621	.2248409
union	2.945815	.2749624	10.71	0.000	2.406898	3.484731
_cons	-4.351572	.5283952	-8.24	0.000	-5.387208	-3.315936
union						
south	-.5807419	.0851111	-6.82	0.000	-.7475567	-.4139271
black	.4557499	.0958042	4.76	0.000	.2679772	.6435226
tenure	.0871536	.0232483	3.75	0.000	.0415878	.1327195
_cons	-.8855759	.0724506	-12.22	0.000	-1.027576	-.7435754
/athrho	-.6544344	.0910315	-7.19	0.000	-.8328529	-.4760159
/lnsigma	.7026768	.0293372	23.95	0.000	.645177	.7601767
rho	-.5746476	.0609711			-.6820049	-.4430472
sigma	2.01915	.0592362			1.906324	2.138654
lambda	-1.1603	.1495099			-1.453334	-.867266

```
LR test of indep. eqns. (rho = 0):   chi2(1) =    19.84   Prob > chi2 = 0.0000
```

All the covariates are statistically significant, and the likelihood-ratio test in the footer indicates that we can reject the null hypothesis of no correlation between the treatment errors and the outcome errors. The estimated ATE of being a union member is 2.95. The ATET is the same as the ATE in this case because the treatment indicator variable has not been interacted with any of the outcome covariates.

Although we discuss some details about this parameter below, the estimated correlation between the treatment-assignment errors and the outcome errors is -0.575, indicating that unobservables that raise observed wages tend to occur with unobservables that lower union membership.

The results for the two ancillary parameters require explanation. For numerical stability during optimization, `etregress` does not directly estimate ρ or σ. Instead, `etregress` estimates the inverse hyperbolic tangent of ρ,

$$\operatorname{atanh}\rho = \frac{1}{2}\ln\left(\frac{1+\rho}{1-\rho}\right)$$

and $\ln\sigma$. Also `etregress` reports $\lambda = \rho\sigma$, along with an estimate of the standard error of the estimate and the confidence interval.

◁

Average treatment effect (ATE)

When there is a treatment variable and outcome covariate interaction, the parameter estimates from `etregress` can be used by `margins` to estimate the ATE, the average difference of the treatment potential outcomes and the control potential outcomes.

▷ Example 2

In example 1, the coefficients on the outcome covariates do not vary by treatment level. The differences in wages between union members and nonmembers are modeled as a level shift captured by the coefficient on the indicator for union membership. In this example, we use factor-variable notation to allow some of the coefficients to vary over treatment level and then use `margins` (see [R] **margins**) to estimate the ATE. (See [U] **11.4.3 Factor variables** for an introduction to factor-variable notation.)

We begin be estimating the parameters of the model in which the coefficients on `black` and `tenure` differ for union members and nonmembers. We specify the `vce(robust)` option because we need to specify `vce(unconditional)` when we use `margins` below.

```
. etregress wage age grade smsa i.union#c.(black tenure),
> treat(union = south black tenure) vce(robust)
Iteration 0:   log pseudolikelihood = -3104.7035
Iteration 1:   log pseudolikelihood = -3053.2128
Iteration 2:   log pseudolikelihood = -3049.3217
Iteration 3:   log pseudolikelihood = -3049.2838
Iteration 4:   log pseudolikelihood = -3049.2838
Linear regression with endogenous treatment       Number of obs   =       1210
Estimator: maximum likelihood                     Wald chi2(8)    =     493.40
Log pseudolikelihood = -3049.2838                 Prob > chi2     =     0.0000
```

	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
wage						
age	.1489075	.0207283	7.18	0.000	.1082809	.1895342
grade	.4200493	.0377621	11.12	0.000	.346037	.4940616
smsa	.9232614	.1201486	7.68	0.000	.6877746	1.158748
union#c.black						
0	-.6685584	.1444213	-4.63	0.000	-.9516189	-.3854979
1	-1.183101	.2574817	-4.59	0.000	-1.687756	-.6784459
union#c.tenure						
0	.1687459	.0503107	3.35	0.001	.0701387	.2673531
1	.0836366	.0903669	0.93	0.355	-.0934793	.2607526
union	3.342862	.5586856	5.98	0.000	2.247859	4.437866
_cons	-4.42566	.6493003	-6.82	0.000	-5.698265	-3.153054
union						
south	-.5844679	.0833069	-7.02	0.000	-.7477464	-.4211894
black	.4740688	.093241	5.08	0.000	.2913198	.6568179
tenure	.0874297	.0253892	3.44	0.001	.0376678	.1371916
_cons	-.8910483	.0746329	-11.94	0.000	-1.037326	-.7447705
/athrho	-.673316	.2215326	-3.04	0.002	-1.107512	-.23912
/lnsigma	.7055908	.0749711	9.41	0.000	.5586502	.8525315
rho	-.5871569	.1451585			-.8031811	-.2346644
sigma	2.025043	.1518197			1.748311	2.345577
lambda	-1.189018	.3631074			-1.900695	-.4773404

```
Wald test of indep. eqns. (rho = 0): chi2(1) =    22.35    Prob > chi2 = 0.0000
```

The results indicate that the coefficients on `black` differ by union membership and that the coefficient on `tenure` for nonmembers is positive, while the coefficient on `tenure` for members is 0. The model fits well overall, so we proceed with interpretation. Because we interacted the treatment variable with two of the covariates, the estimated coefficient on the treatment level is not an estimate of the ATE. Below we use `margins` to estimate the ATE from these results. We specify the `vce(unconditional)` option to obtain the standard errors for the population ATE instead of the sample ATE. We specify the `contrast(nowald)` option to suppress the Wald tests, which `margins` displays by default for contrasts.

```
. margins r.union, vce(unconditional) contrast(nowald)
Contrasts of predictive margins
Expression   : Linear prediction, predict()
```

	Contrast	Unconditional Std. Err.	[95% Conf.	Interval]
union (1 vs 0)	3.042691	.5305143	2.002902	4.08248

The ATE estimate is essentially the same as the one produced by the constrained model in example 1.

◁

Average treatment effect on the treated (ATET)

When there is a treatment variable and outcome covariate interaction, the parameter estimates from `etregress` can be used by `margins` to estimate the ATET, the average difference of the treatment potential outcomes and the control potential outcomes on the treated population.

▷ Example 3

The ATET may differ from the ATE in the previous example because the interaction between the treatment variable and some outcome covariates makes the ATE and the ATET vary over outcome covariate values. Below we use `margins` to estimate the ATET by specifying the `subpop(union)` option, which restricts the sample used by `margins` to union members.

```
. margins r.union, vce(unconditional) contrast(nowald) subpop(union)
Contrasts of predictive margins
Expression   : Linear prediction, predict()
```

	Contrast	Unconditional Std. Err.	[95% Conf.	Interval]
union (1 vs 0)	2.96898	.5358449	1.918744	4.019217

The estimated ATET and ATE are close, indicating that the average predicted outcome for the treatment group is similar to the average predicted outcome for the whole population.

◁

Stored results

`etregress` (maximum likelihood) stores the following in `e()`:

Scalars

`e(N)`	number of observations
`e(k)`	number of parameters
`e(k_eq)`	number of equations in `e(b)`
`e(k_eq_model)`	number of equations in overall model test
`e(k_aux)`	number of auxiliary parameters
`e(k_dv)`	number of dependent variables
`e(df_m)`	model degrees of freedom
`e(ll)`	log likelihood
`e(ll_0)`	log likelihood, constant-only model (`noskip` only)
`e(N_clust)`	number of clusters
`e(lambda)`	λ
`e(selambda)`	standard error of λ
`e(sigma)`	estimate of sigma
`e(chi2)`	χ^2
`e(chi2_c)`	χ^2 for comparison test
`e(p_c)`	p-value for comparison test
`e(p)`	significance
`e(rho)`	ρ
`e(rank)`	rank of `e(V)`
`e(rank0)`	rank of `e(V)` for constant-only model
`e(ic)`	number of iterations
`e(rc)`	return code
`e(converged)`	`1` if converged, `0` otherwise

Macros

`e(cmd)`	`etregress`
`e(cmdline)`	command as typed
`e(depvar)`	name of dependent variable
`e(hazard)`	variable containing hazard
`e(wtype)`	weight type
`e(wexp)`	weight expression
`e(title)`	title in estimation output
`e(clustvar)`	name of cluster variable
`e(chi2type)`	`Wald` or `LR`; type of model χ^2 test
`e(chi2_ct)`	`Wald` or `LR`; type of model χ^2 test corresponding to `e(chi2_c)`
`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(method)`	`ml`
`e(ml_method)`	type of `ml` method
`e(user)`	name of likelihood-evaluator program
`e(technique)`	maximization technique
`e(properties)`	`b V`
`e(predict)`	program used to implement `predict`
`e(footnote)`	program used to implement the footnote display
`e(marginsok)`	predictions allowed by `margins`
`e(asbalanced)`	factor variables `fvset` as `asbalanced`
`e(asobserved)`	factor variables `fvset` as `asobserved`

Matrices

`e(b)`	coefficient vector
`e(Cns)`	constraints matrix
`e(ilog)`	iteration log (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

`etregress` (two-step) stores the following in `e()`:

Scalars	
`e(N)`	number of observations
`e(df_m)`	model degrees of freedom
`e(lambda)`	λ
`e(selambda)`	standard error of λ
`e(sigma)`	estimate of sigma
`e(chi2)`	χ^2
`e(p)`	significance
`e(rho)`	ρ
`e(rank)`	rank of `e(V)`
Macros	
`e(cmd)`	`etregress`
`e(cmdline)`	command as typed
`e(depvar)`	name of dependent variable
`e(title)`	title in estimation output
`e(chi2type)`	`Wald` or `LR`; type of model χ^2 test
`e(vce)`	*vcetype* specified in `vce()`
`e(vcetype)`	title used to label Std. Err.
`e(hazard)`	variable specified in `hazard()`
`e(method)`	`ml` or `twostep`
`e(properties)`	`b V`
`e(predict)`	program used to implement `predict`
`e(footnote)`	program used to implement the footnote display
`e(marginsok)`	predictions allowed by `margins`
`e(marginsnotok)`	predictions disallowed by `margins`
`e(asbalanced)`	factor variables `fvset` as `asbalanced`
`e(asobserved)`	factor variables `fvset` as `asobserved`
Matrices	
`e(b)`	coefficient vector
`e(V)`	variance–covariance matrix of the estimators
Functions	
`e(sample)`	marks estimation sample

Methods and formulas

Maddala (1983, 117–122) derives both the maximum likelihood and the two-step estimator implemented here. Greene (2012, 890–894) also provides an introduction to the treatment-effects model. Cameron and Trivedi (2005, sections 16.7 and 25.3.4) and Wooldridge (2010, section 21.4.1) discuss the endogenous binary-variable model as an endogenous treatment-effects model and link it to recent work.

The primary regression equation of interest is

$$y_j = \mathbf{x}_j\boldsymbol{\beta} + \delta t_j + \epsilon_j$$

where t_j is a binary-treatment variable that is assumed to stem from an unobservable latent variable:

$$t_j^* = \mathbf{w}_j\boldsymbol{\gamma} + u_j$$

The decision to obtain the treatment is made according to the rule

$$t_j = \begin{cases} 1, & \text{if } t_j^* > 0 \\ 0, & \text{otherwise} \end{cases}$$

where ϵ and u are bivariate normal with mean zero and covariance matrix

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

The likelihood function for this model is given in Maddala (1983, 122). Greene (2000, 180) discusses the standard method of reducing a bivariate normal to a function of a univariate normal and the correlation ρ. The following is the log likelihood for observation j,

$$\ln L_j = \begin{cases} \ln\Phi\left\{\dfrac{\mathbf{w}_j\boldsymbol{\gamma} + (y_j - \mathbf{x}_j\boldsymbol{\beta} - \delta)\rho/\sigma}{\sqrt{1-\rho^2}}\right\} - \dfrac{1}{2}\left(\dfrac{y_j - \mathbf{x}_j\boldsymbol{\beta} - \delta}{\sigma}\right)^2 - \ln(\sqrt{2\pi}\sigma) & t_j = 1 \\ \ln\Phi\left\{\dfrac{-\mathbf{w}_j\boldsymbol{\gamma} - (y_j - \mathbf{x}_j\boldsymbol{\beta})\rho/\sigma}{\sqrt{1-\rho^2}}\right\} - \dfrac{1}{2}\left(\dfrac{y_j - \mathbf{x}_j\boldsymbol{\beta}}{\sigma}\right)^2 - \ln(\sqrt{2\pi}\sigma) & t_j = 0 \end{cases}$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution.

In the maximum likelihood estimation, σ and ρ are not directly estimated. Rather $\ln\sigma$ and $\operatorname{atanh}\rho$ are directly estimated, where

$$\operatorname{atanh}\rho = \frac{1}{2}\ln\left(\frac{1+\rho}{1-\rho}\right)$$

The standard error of $\lambda = \rho\sigma$ is approximated through the delta method, which is given by

$$\mathrm{Var}(\lambda) \approx \mathbf{D}\,\mathrm{Var}\big\{(\operatorname{atanh}\rho \quad \ln\sigma)\big\}\,\mathbf{D}'$$

where $\mathbf{D}$ is the Jacobian of λ with respect to $\operatorname{atanh}\rho$ and $\ln\sigma$.

With maximum likelihood estimation, this command supports the Huber/White/sandwich estimator of the variance and its clustered version using `vce(robust)` and `vce(cluster` *clustvar*`)`, respectively. See [P] **_robust**, particularly *Maximum likelihood estimators* and *Methods and formulas*.

The maximum likelihood version of `etregress` also supports estimation with survey data. For details on VCEs with survey data, see [SVY] **variance estimation**.

Maddala (1983, 120–122) also derives the two-step estimator. In the first stage, probit estimates are obtained of the treatment equation

$$\Pr(t_j = 1 \mid \mathbf{w}_j) = \Phi(\mathbf{w}_j\boldsymbol{\gamma})$$

From these estimates, the hazard, h_j, for each observation j is computed as

$$h_j = \begin{cases} \phi(\mathbf{w}_j\widehat{\boldsymbol{\gamma}})/\Phi(\mathbf{w}_j\widehat{\boldsymbol{\gamma}}) & t_j = 1 \\ -\phi(\mathbf{w}_j\widehat{\boldsymbol{\gamma}})/\big\{1 - \Phi(\mathbf{w}_j\widehat{\boldsymbol{\gamma}})\big\} & t_j = 0 \end{cases}$$

where ϕ is the standard normal density function. If

$$d_j = h_j(h_j + \widehat{\boldsymbol{\gamma}}\mathbf{w}_j)$$

then

$$E\left(y_j \mid t_j, \mathbf{x}_j, \mathbf{w}_j\right) = \mathbf{x}_j\boldsymbol{\beta} + \delta t_j + \rho\sigma h_j$$
$$\mathrm{Var}\left(y_j \mid t_j, \mathbf{x}_j, \mathbf{w}_j\right) = \sigma^2\left(1 - \rho^2 d_j\right)$$

The two-step parameter estimates of $\boldsymbol{\beta}$ and δ are obtained by augmenting the regression equation with the hazard h. Thus the regressors become $[\,\mathbf{x}\ \mathbf{t}\ h\,]$, and the additional parameter estimate β_h is obtained on the variable containing the hazard. A consistent estimate of the regression disturbance variance is obtained using the residuals from the augmented regression and the parameter estimate on the hazard

$$\widehat{\sigma}^{\,2} = \frac{\mathbf{e}'\mathbf{e} + \beta_h^2 \sum_{j=1}^N d_j}{N}$$

The two-step estimate of ρ is then

$$\widehat{\rho} = \frac{\beta_h}{\widehat{\sigma}}$$

To understand how the consistent estimates of the coefficient covariance matrix based on the augmented regression are derived, let $\mathbf{A} = [\,\mathbf{x}\ \mathbf{t}\ h\,]$ and $\mathbf{D}$ be a square diagonal matrix of size N with $(1 - \widehat{\rho}^{\,2} d_j)$ on the diagonal elements. The conventional VCE is

$$\mathbf{V}_{\text{twostep}} = \widehat{\sigma}^{\,2}(\mathbf{A}'\mathbf{A})^{-\mathbf{1}}(\mathbf{A}'\mathbf{D}\mathbf{A} + \mathbf{Q})(\mathbf{A}'\mathbf{A})^{-\mathbf{1}}$$

where

$$\mathbf{Q} = \widehat{\rho}^{\,2}(\mathbf{A}'\mathbf{D}\mathbf{A})\mathbf{V}_{\mathbf{p}}(\mathbf{A}'\mathbf{D}\mathbf{A})$$

and $\mathbf{V}_{\mathbf{p}}$ is the variance–covariance estimate from the probit estimation of the treatment equation.

References

Barnow, B. S., G. G. Cain, and A. S. Goldberger. 1981. Issues in the analysis of selectivity bias. In Vol. 5 of *Evaluation Studies Review Annual*, ed. E. W. Stromsdorfer and G. Farkas, 123–126. Beverly Hills: Sage.

Brave, S., and T. Walstrum. 2014. Estimating marginal treatment effects using parametric and semiparametric methods. *Stata Journal* 14: 191–217.

Brown, G. K., and T. Mergoupis. 2011. Treatment interactions with nonexperimental data in Stata. *Stata Journal* 11: 545–555.

Cameron, A. C., and P. K. Trivedi. 2005. *Microeconometrics: Methods and Applications*. New York: Cambridge University Press.

Cong, R., and D. M. Drukker. 2000. sg141: Treatment effects model. *Stata Technical Bulletin* 55: 25–33. Reprinted in *Stata Technical Bulletin Reprints*, vol. 10, pp. 159–169. College Station, TX: Stata Press.

Greene, W. H. 2000. *Econometric Analysis*. 4th ed. Upper Saddle River, NJ: Prentice Hall.

———. 2012. *Econometric Analysis*. 7th ed. Upper Saddle River, NJ: Prentice Hall.

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.

———. 1978. Dummy endogenous variables in a simultaneous equation system. *Econometrica* 46: 931–959.

Maddala, G. S. 1983. *Limited-Dependent and Qualitative Variables in Econometrics*. Cambridge: Cambridge University Press.

Nannicini, T. 2007. Simulation-based sensitivity analysis for matching estimators. *Stata Journal* 7: 334–350.

Nichols, A. 2007. Causal inference with observational data. *Stata Journal* 7: 507–541.

Wooldridge, J. M. 2010. *Econometric Analysis of Cross Section and Panel Data*. 2nd ed. Cambridge, MA: MIT Press.

Also see

[TE] **etregress postestimation** — Postestimation tools for etregress

[TE] **etpoisson** — Poisson regression with endogenous treatment effects

[R] **heckman** — Heckman selection model

[R] **probit** — Probit regression

[R] **regress** — Linear regression

[SVY] **svy estimation** — Estimation commands for survey data

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