

[Description](#)[Menu](#)[Options for maximum likelihood estimates](#)[Options for control-function estimates](#)[Stored results](#)[References](#)[Quick start](#)[Syntax](#)[Options for two-step consistent estimates](#)[Remarks and examples](#)[Methods and formulas](#)[Also see](#)

## 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 full maximum likelihood, a two-step consistent estimator, or a control-function 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.

## Quick start

ATE and ATET from a linear regression model of `y` on `x` and endogenous binary treatment `treat` modeled by `x` and `w`

```
etregress y x, treat(treat = x w)
```

Same as above, but use a control-function estimator

```
etregress y x, treat(treat = x w) cfunction
```

With robust standard errors

```
etregress y x, treat(treat = x w) vce(robust)
```

Add the interaction between `treat` and continuous covariate `x` using [factor variables](#)

```
etregress y x i.treat#c.x, treat(treat = x w) vce(robust)
```

ATE after `etregress` with the required `vce(robust)` option and endogenous treatment interaction terms

```
margins r.treat, vce(unconditional)
```

Same as above, but calculate ATET

```
margins, vce(unconditional) predict(cte) subpop(if treat==1)
```

## Menu

Statistics > Causal inference/treatment effects > Endogenous treatment > Maximum likelihood estimator > Continuous outcomes

## Syntax

*Basic syntax*

```
etregress depvar [indepvars] , treat(depvart = indepvarst) [twostep | cfunction]
```

*Full syntax for maximum likelihood estimates only*

```
etregress depvar [indepvars] [if] [in] [weight] ,  
  treat(depvart = indepvarst [, noconstant]) [etregress_ml_options]
```

*Full syntax for two-step consistent estimates only*

```
etregress depvar [indepvars] [if] [in] ,  
  treat(depvart = indepvarst [, noconstant]) twostep [etregress_ts_options]
```

*Full syntax for control-function estimates only*

```
etregress depvar [indepvars] [if] [in] ,  
  treat(depvart = indepvarst [, noconstant]) cfunction [etregress_cf_options]
```

<i>etregress_ml_options</i>	Description
Model	
* <u>treat</u> ()	equation for treatment effects
<u>noconstant</u>	suppress constant term
<u>poutcomes</u>	use potential-outcome model with separate treatment and control group variance and correlation parameters
<u>constraints</u> ( <i>constraints</i> )	apply specified linear constraints
SE/Robust	
<u>vce</u> ( <i>vcetype</i> )	<i>vcetype</i> may be <u>oim</u> , <u>robust</u> , <u>cluster</u> <i>clustvar</i> , <u>opg</u> , <u>bootstrap</u> , or <u>jackknife</u>
Reporting	
<u>level</u> (#)	set confidence level; default is level(95)
<u>first</u>	report first-step probit estimates
<u>hazard</u> ( <i>newvar</i> )	create <i>newvar</i> containing hazard from treatment equation
<u>lrmodel</u>	perform the likelihood-ratio model test instead of the default Wald test
<u>nocnsreport</u>	do not display constraints
<u>display_options</u>	control columns and column formats, row spacing, line width, display of omitted variables and base and empty cells, and factor-variable labeling
Maximization	
<u>maximize_options</u>	control the maximization process; seldom used
<u>collinear</u>	keep collinear variables
<u>coeflegend</u>	display legend instead of statistics

\*treat(*depvar*<sub>*t*</sub> = *indepvars*<sub>*t*</sub> [, noconstant]) is required.

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

\*treat(*depvar*<sub>*t*</sub> = *indepvars*<sub>*t*</sub> [, noconstant]) and twostep are required.

<i>etregress _cf_options</i>	Description
Model	
* <u>treat()</u>	equation for treatment effects
* <u>cfunction</u>	produce control-function estimate
<u>noconstant</u>	suppress constant term
<u>poutcomes</u>	use potential-outcome model with separate treatment and control group variance and correlation parameters
SE	
<u>vce</u> ( <i>vcetype</i> )	<i>vcetype</i> may be <u>robust</u> , <u>bootstrap</u> , or <u>jackknife</u>
Reporting	
<u>level</u> (#)	set confidence level; default is level(95)
<u>first</u>	report first-step probit estimates
<u>hazard</u> ( <i>newvar</i> )	create <i>newvar</i> containing hazard from treatment equation
<i>display_options</i>	control columns and column formats, row spacing, line width, display of omitted variables and base and empty cells, and factor-variable labeling
Maximization	
<i>maximize_options</i>	control the maximization process; seldom used
<u>coeflegend</u>	display legend instead of statistics

\*treat(*depvar*<sub>*t*</sub> = *indepvars*<sub>*t*</sub> [ , noconstant ]) and cfunction are required.

*indepvars* and *indepvars*<sub>*t*</sub> may contain factor variables; see [U] 11.4.3 Factor variables.

*depvar*, *indepvars*, *depvar*<sub>*t*</sub>, and *indepvars*<sub>*t*</sub> may contain time-series operators; see [U] 11.4.4 Time-series varlists.

bayesboot, bootstrap, by, collect, 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, cfunction, vce( ), first, hazard( ), lrmodel, and weights are not allowed with the svy prefix; see [SVY] svy.

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

collinear and coeflegend do not appear in the dialog box.

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

## Options for maximum likelihood estimates

### Model

`treat(depvvart = indepvvarst [, 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`; see [R] [Estimation options](#).

`poutcomes` specifies that a potential-outcome model with separate variance and correlation parameters for each of the treatment and control groups be used.

`constraints(constraints)`; 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.

`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.

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

`display_options`: `nocl`, `nopvalues`, `noomitted`, `vsquish`, `noemptycells`, `baselevels`, `allbaselevels`, `nofvlabel`, `fvwrap(#)`, `fvwrapon(style)`, `cformat(%fmt)`, `pformat(%fmt)`, `sformat(%fmt)`, and `no1stretch`; 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 options are available with `etregress` but are not shown in the dialog box:

`collinear`, `coeflegend`; see [R] [Estimation options](#).

## Options for two-step consistent estimates

### Model

`treat(depvart = indepvarst[ , 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*: `noci`, `nopvalues`, `noomitted`, `vsquish`, `noemptycells`, `baselevels`, `allbaselevels`, `nofvlabel`, `fvwrap(#)`, `fvwrapon(style)`, `cformat(%fmt)`, `pformat(%fmt)`, `sformat(%fmt)`, and `no1stretch`; see [R] [Estimation options](#).

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

`coeflegend`; see [R] [Estimation options](#).

## Options for control-function estimates

### Model

`treat(depvart = indepvarst[ , noconstant ])` specifies the variables and options for the treatment equation. It is an integral part of specifying a treatment-effects model and is required.

`cfunction` specifies that control-function estimates of the parameters, standard errors, and covariance matrix be produced instead of the default maximum likelihood estimates. `cfunction` is required.

`noconstant`; see [R] [Estimation options](#).

`poutcomes` specifies that a potential-outcome model with separate variance and correlation parameters for each of the treatment and control groups be used.

### SE

`vce(vcetype)` specifies the type of standard error reported, which includes types that are robust to some kinds of misspecification (`robust`) 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.

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

## Maximization

`maximize_options`: `iterate(#)`, `[no]log`, and `from(init_specs)`; see [R] [Maximize](#). These options are seldom used.

`init_specs` is one of

`matname [ , skip copy]`

`# [ # ... ] copy`

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

`coeflegend`; see [R] [Estimation options](#).

## Remarks and examples

Remarks are presented under the following headings:

[Overview](#)

[Basic examples](#)

[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 normal distribution to model the deviation from the conditional independence assumption imposed by the estimators implemented in `teffects`; see [CAUSAL] [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 [CAUSAL] [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 the control-function (CF) estimators of the model. Maddala (1983) also reviews some empirical applications and describes it as an 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,  $\delta$ , 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.

etregress performs CF estimation in one step by using the generalized method of moments (GMM) with stacked moments. See Newey (1984) and Wooldridge (2010, sec. 14.2) for a description of this technique. Many econometric and statistical models can be expressed as conditions on the population moments. The parameter estimates produced by GMM estimators make the sample-moment conditions as true as possible given the data. See [R] gmm for further information on GMM estimation and how Stata performs it. Two-step CF estimation is also supported by etregress.

Formally, the endogenous treatment-regression model is composed of an equation for the outcome  $y_j$  and an equation for the endogenous treatment  $t_j$ . The variables  $\mathbf{x}_j$  are used to model the outcome. When there are no interactions between  $t_j$  and  $\mathbf{x}_j$ , we have

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

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

where  $\mathbf{w}_j$  are the covariates used to model treatment assignment, and the error terms  $\epsilon_j$  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. We call this the constrained model because the variance and correlation parameters are identical across the treatment and control groups.

This model can be generalized to a potential-outcome model with separate variance and correlation parameters for the treatment and control groups. The generalized model is

$$y_{0j} = \mathbf{x}_j\boldsymbol{\beta}_0 + \epsilon_{0j}$$

$$y_{1j} = \mathbf{x}_j\boldsymbol{\beta}_1 + \epsilon_{1j}$$

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

where  $y_{0j}$  is the outcome that person  $j$  obtains if person  $j$  selects treatment 0, and  $y_{1j}$  is the outcome that person  $j$  obtains if person  $j$  selects treatment 1. We never observe both  $y_{0j}$  and  $y_{1j}$ , only one or the other. We observe

$$y_j = t_j y_{1j} + (1 - t_j) y_{0j}$$



In this unconstrained model, the vector of error terms  $(\epsilon_{0j}, \epsilon_{1j}, u_j)'$  comes from a mean zero trivariate normal distribution with covariance matrix

$$\begin{bmatrix} \sigma_0^2 & \sigma_{01} & \sigma_0\rho_0 \\ \sigma_{01} & \sigma_1^2 & \sigma_1\rho_1 \\ \sigma_0\rho_0 & \sigma_1\rho_1 & 1 \end{bmatrix}$$

The covariance  $\sigma_{01}$  cannot be identified because we never observe both  $y_{1j}$  and  $y_{0j}$ . However, identification of  $\sigma_{01}$  is not necessary to estimate the other parameters because all covariates and the outcome are observed in observations from each group. We normalize the treatment error variance to be 1 because we observe only whether an outcome occurs under treatment. More details are found in [Methods and formulas](#).

Rather than showing two separate regression equations, `etregress` reports one outcome equation with interaction terms between the treatment and outcome covariates. `etregress` can fit the constrained and generalized potential-outcome models using either the default maximum likelihood estimator or the one-step CF estimator obtained with option `cfunction`. The two-step CF estimator provides consistent estimates for the constrained model.

## Basic examples

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

### ► Example 1: Basic example

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 https://www.stata-press.com/data/r19/union3
(NLS Women 14-24 in 1968)

. etregress wage age grade smsa black tenure, treat(union = south black tenure)

Iteration 0:  Log likelihood = -3140.811
Iteration 1:  Log likelihood = -3053.6629
Iteration 2:  Log likelihood = -3051.5847
Iteration 3:  Log likelihood = -3051.575
Iteration 4:  Log likelihood = -3051.575

Linear regression with endogenous treatment          Number of obs = 1,210
Estimator: Maximum likelihood                      Wald chi2(6) = 681.89
Log likelihood = -3051.575                          Prob > chi2 = 0.0000
```

	Coefficient	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	.9117044	.1249041	7.30	0.000	.6668969	1.156512
black	-.7882471	.1367078	-5.77	0.000	-1.05619	-.5203048
tenure	.1524015	.0369596	4.12	0.000	.0799621	.2248409
1.union	2.945815	.2749621	10.71	0.000	2.4069	3.484731
_cons	-4.351572	.5283952	-8.24	0.000	-5.387208	-3.315936
union						
south	-.5807419	.0851111	-6.82	0.000	-.7475566	-.4139271
black	.4557499	.0958042	4.76	0.000	.2679771	.6435226
tenure	.0871536	.0232483	3.75	0.000	.0415878	.1327195
_cons	-.8855758	.0724506	-12.22	0.000	-1.027576	-.7435753
/athrho	-.6544347	.0910314	-7.19	0.000	-.832853	-.4760164
/lnsigma	.7026769	.0293372	23.95	0.000	.645177	.7601767
rho	-.5746478	.060971			-.682005	-.4430476
sigma	2.019151	.0592362			1.906325	2.138654
lambda	-1.1603	.1495097			-1.453334	-.8672668

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

The likelihood-ratio test in the footer indicates that we can reject the null hypothesis of no correlation between the treatment-assignment errors and the outcome errors. The estimated correlation between the treatment-assignment errors and the outcome errors,  $\rho$ , is  $-0.575$ . The negative relationship indicates that unobservables that raise observed wages tend to occur with unobservables that lower union membership. We discuss some details about this parameter in the technical note [below](#).

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, and the correlation and variance parameters are identical across the control and treatment groups.

## □ Technical note

The results for the ancillary parameters  $\rho$  and  $\sigma$  require explanation. For numerical stability during optimization, `etregress` does not directly estimate  $\rho$  or  $\sigma$ . Instead, `etregress` estimates the inverse hyperbolic tangent of  $\rho$ ,

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

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

□

In contrast to the constrained model, `etregress` directly estimates the ATE only when there are no interactions between the treatment variable and the outcome covariates in the unconstrained model.

## ▷ Example 2: Allowing group-specific variance and correlation

We estimate the ATE of having health insurance on the natural logarithm of total out-of-pocket prescription drug expenditures from a simulated random sample of individuals between the ages of 25 and 64. We will use the variables `lndrug` (natural logarithm of spending on prescription drugs), `age` (age of the individual), `chron` (whether the individual has a chronic condition), `lninc` (natural logarithm of income), `married` (marriage status), and `work` (employment status). Our treatment is whether the person has health insurance, `ins`. We allow the outcome error variance and correlation parameters to vary between the treated (insured) and control (uninsured) groups in this example, rather than constraining them to be equal as in [example 1](#).

We use `etregress` to estimate the parameters of the endogenous treatment-effects model. To estimate separate variance and correlation parameters for each of the control and treatment groups, we specify the `poutcomes` option. We specify the `cfunction` option to use the CF estimator.

```
. use https://www.stata-press.com/data/r19/drugexp
(Prescription drug expenditures)
. etregress lndrug chron age lninc, treat(ins=age married lninc work) poutcomes
> cfunction
Iteration 0: GMM criterion Q(b) = 2.279e-15
Iteration 1: GMM criterion Q(b) = 1.842e-28
Linear regression with endogenous treatment          Number of obs = 6,000
Estimator: Control function
```

	Robust					
	Coefficient	std. err.	z	P> z	[95% conf. interval]	
lndrug						
chron	.4671725	.0319731	14.61	0.000	.4045064	.5298387
age	.1021359	.00292	34.98	0.000	.0964128	.1078589
lninc	.0550672	.0225036	2.45	0.014	.0109609	.0991735
1.ins	-.8598836	.3483648	-2.47	0.014	-1.542666	-.1771011
_cons	1.665539	.2527527	6.59	0.000	1.170153	2.160925
ins						
age	.021142	.0022961	9.21	0.000	.0166416	.0256424
married	.084631	.0359713	2.35	0.019	.0141286	.1551334
lninc	.1023032	.0225009	4.55	0.000	.0582022	.1464041
work	.288418	.0372281	7.75	0.000	.2154522	.3613837
_cons	-.622993	.108795	-5.73	0.000	-.8362273	-.4097587
/athrho0	.4035094	.1724539	2.34	0.019	.0655059	.7415129
/lnsigma0	.3159269	.0500476	6.31	0.000	.2178353	.4140184
/athrho1	.7929459	.2986601	2.66	0.008	.2075829	1.378309
/lnsigma1	.1865347	.0613124	3.04	0.002	.0663646	.3067048
rho0	.3829477	.1471637			.0654124	.6300583
sigma0	1.37153	.0686418			1.243382	1.512885
lambda0	.5252243	.226367			.0815532	.9688954
rho1	.6600746	.1685343			.2046518	.880572
sigma1	1.205066	.0738855			1.068616	1.35894
lambda1	.7954338	.2513036			.3028878	1.28798

Wald test of indep. (rho0 = rho1 = 0): chi2(2) = 8.88 Prob > chi2 = 0.0118

The Wald test reported in the footer indicates that we can reject the null hypothesis of no correlation between the treatment-assignment errors and the outcome errors for the control and treatment groups. The estimate of the correlation of the treatment-assignment errors for the control group ( $\rho_0$ ) is positive, indicating that unobservables that increase spending on prescription drugs tend to occur with unobservables that increase health insurance coverage. Because  $\rho_1$  is also positive, we make the same interpretation for individuals with insurance. The estimate  $\rho_1$  is larger than the estimate  $\rho_0$ , indicating a stronger relationship between the unobservables and treatment outcomes in the treated group.

The estimated ATE of having health insurance is  $-0.86$ . Note that while the ATE and ATET were the same in [example 1](#), that is not the case here. We show how to calculate the ATET for a potential-outcome model in [example 6](#).

The estimate of the outcome error standard-deviation parameter for the control group ( $\sigma_0$ ) is slightly larger than that of the treatment-group parameter ( $\sigma_1$ ), indicating a greater variability in the unobservables among the untreated group.

## 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 3: Allowing interactions between treatment and outcome covariates, ATE

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 by 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.

```
. use https://www.stata-press.com/data/r19/union3
(NLS Women 14-24 in 1968)

. etregress wage age grade smsa i.union#c.(black tenure),
> treat(union = south black tenure) vce(robust)

Iteration 0:  Log pseudolikelihood = -3614.6714
Iteration 1:  Log pseudolikelihood = -3218.8152
Iteration 2:  Log pseudolikelihood = -3057.0115
Iteration 3:  Log pseudolikelihood = -3049.3081
Iteration 4:  Log pseudolikelihood = -3049.2838
Iteration 5:  Log pseudolikelihood = -3049.2838

Linear regression with endogenous treatment
Estimator: Maximum likelihood
Log pseudolikelihood = -3049.2838

Number of obs = 1,210
Wald chi2(8) = 493.40
Prob > chi2 = 0.0000
```

	Coefficient	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	.3460371	.4940616
smsa	.9232615	.1201486	7.68	0.000	.6877746	1.158748
union#						
c.black						
0	-.6685582	.1444213	-4.63	0.000	-.9516187	-.3854977
1	-1.1831	.2574817	-4.59	0.000	-1.687755	-.6784455
union#						
c.tenure						
0	.168746	.0503107	3.35	0.001	.0701388	.2673532
1	.0836367	.0903669	0.93	0.355	-.0934792	.2607526
1.union	3.342859	.5586863	5.98	0.000	2.247854	4.437864
_cons	-4.42566	.6493003	-6.82	0.000	-5.698265	-3.153055
union						
south	-.5844678	.0833069	-7.02	0.000	-.7477464	-.4211893
black	.4740688	.093241	5.08	0.000	.2913197	.6568178
tenure	.0874297	.0253892	3.44	0.001	.0376678	.1371916
_cons	-.8910484	.0746329	-11.94	0.000	-1.037326	-.7447706
/athrho	-.6733149	.2215328	-3.04	0.002	-1.107511	-.2391185
/lnsigma	.7055907	.0749711	9.41	0.000	.55865	.8525313
rho	-.5871562	.1451589			-.8031809	-.234663
sigma	2.025042	.1518197			1.748311	2.345577
lambda	-1.189016	.3631079			-1.900695	-.4773378

Wald test of indep. eqns. (rho = 0): chi2(1) = 9.24 Prob > chi2 = 0.0024

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                                Number of obs = 1,210
Expression: Linear prediction, predict()
```

	Unconditional			
	Contrast	std. err.	[95% conf. interval]	
union (1 vs 0)	3.042688	.5305151	2.002898	4.082478

The ATE estimate is essentially the same as the one produced by the constrained model in [example 1](#). ◀

We can use the same methods above to obtain the ATE in an unconstrained model.

➤ Example 4: Treatment interactions and group-specific variance and correlation, ATE

In [example 2](#), the coefficients on the outcome covariates do not vary by treatment level. Suppose we believe that the effect of having a chronic condition on out-of-pocket spending differs between the insured and uninsured. Again, we use an interaction term. Because we are using a CF estimator, the variance–covariance of the estimator (VCE) is already robust so we do not specify `vce(robust)`.

```
. use https://www.stata-press.com/data/r19/drugexp
(Prescription drug expenditures)
. etregress lndrug i.ins#i.chron age lninc, treat(ins=age married lninc work)
> poutcomes cfunction
Iteration 0: GMM criterion Q(b) = 2.279e-15
Iteration 1: GMM criterion Q(b) = 2.883e-28
Linear regression with endogenous treatment          Number of obs = 6,000
Estimator: Control function
```

	Robust					
	Coefficient	std. err.	z	P> z	[95% conf. interval]	
lndrug						
ins#chron						
0 1	.3798705	.0720713	5.27	0.000	.2386132	.5211277
1 1	.4957773	.0352571	14.06	0.000	.4266746	.5648801
age	.1022045	.0029228	34.97	0.000	.0964758	.1079331
lninc	.0548917	.0225219	2.44	0.015	.0107497	.0990337
1.ins	-.89703	.3493058	-2.57	0.010	-1.581657	-.2124031
_cons	1.691336	.2531222	6.68	0.000	1.195225	2.187446
ins						
age	.021142	.0022961	9.21	0.000	.0166416	.0256424
married	.084631	.0359713	2.35	0.019	.0141286	.1551334
lninc	.1023032	.0225009	4.55	0.000	.0582022	.1464041
work	.288418	.0372281	7.75	0.000	.2154522	.3613837
_cons	-.622993	.108795	-5.73	0.000	-.8362273	-.4097587
/athrho0	.4046007	.1725597	2.34	0.019	.0663899	.7428115
/lnsigma0	.3157561	.0501956	6.29	0.000	.2173746	.4141376
/athrho1	.7950592	.2992825	2.66	0.008	.2084763	1.381642
/lnsigma1	.1868903	.0614281	3.04	0.002	.0664934	.3072871
rho0	.3838786	.1471308			.0662925	.6308408
sigma0	1.371296	.0688329			1.24281	1.513065
lambda0	.5264111	.2264197			.0826366	.9701856
rho1	.6612655	.1684146			.2055076	.8813184
sigma1	1.205495	.0740512			1.068754	1.359731
lambda1	.7971523	.2514293			.3043599	1.289945

Wald test of indep. (rho0 = rho1 = 0): chi2(2) = 8.90 Prob > chi2 = 0.0117

The results indicate that the coefficient on `chron` differs by whether an individual has insurance. The model fits well overall, so we proceed with interpretation.

Because we interacted the treatment variable with one 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.



```
. margins r.ins, vce(unconditional) contrast(nowald)
Contrasts of predictive margins                                Number of obs = 6,000
Expression: Linear prediction, predict()
```

	Unconditional			
	Contrast	std. err.	[95% conf. interval]	
ins (1 vs 0)	-.8632045	.3484924	-1.546237	-.1801718

The ATE estimate is similar to the one produced by the constrained model in [example 2](#).



## 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 5: Allowing interactions between treatment and outcome covariates, ATET

The ATET may differ from the ATE in [example 3](#) 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.

```
. use https://www.stata-press.com/data/r19/union3
(NLS Women 14-24 in 1968)
. etregress wage age grade smsa i.union#c.(black tenure),
> treat(union = south black tenure) vce(robust)
(output omitted)
. margins r.union, vce(unconditional) contrast(nowald) subpop(union)
Contrasts of predictive margins                                Number of obs = 1,210
Subpop. no. obs = 253
Expression: Linear prediction, predict()
```

	Unconditional			
	Contrast	std. err.	[95% conf. interval]	
union (1 vs 0)	2.968977	.5358456	1.918739	4.019215

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.



## ➤ Example 6: Treatment interactions and group-specific variance and correlation, ATET

The ATET may differ from the ATE in [example 4](#) because the interaction between the treatment variable and some outcome covariates makes the ATE and the ATET vary over values of the covariate in the outcome equation. Even if there is no interaction between treatment assignment and a covariate in the outcome equation, the estimated ATE and ATET will differ if the variances of the outcome errors and their correlations with the treatment-assignment errors differ across the control and treatment groups.

We can estimate the ATET of having health insurance by using the conditional treatment effect (conditional on exogenous covariates and treatment level) obtained using the `predict`, `cte` and the `margins` commands; see [Methods and formulas](#) below and [\[CAUSAL\] etregress postestimation](#) for more details about the use of `predict` after `etregress`.

We restrict estimation to the treated subpopulation by specifying the `subpop(ins)` option with `margins`.

```
. use https://www.stata-press.com/data/r19/drugexp
(Prescription drug expenditures)
. etregress ln drug i.ins#i.chron age lninc,
> treat(ins = age married lninc work) poutcomes cfunction
(output omitted)
. margins, predict(cte) subpop(ins) vce(unconditional)
Predictive margins                                Number of obs   = 6,000
                                                    Subpop. no. obs = 4,556
Expression: Conditional treatment effect, predict(cte)
```

	Unconditional				
	Margin	std. err.	z	P> z	[95% conf. interval]
_cons	-.7558373	.3827579	-1.97	0.048	-1.506029 -.0056457

In absolute value, the treatment effect on the treated of  $-0.76$  is smaller than the population average effect of  $-0.86$  that we found in [example 4](#).



## Stored results

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

### Scalars

<code>e(N)</code>	number of observations
<code>e(k)</code>	number of parameters
<code>e(k_eq)</code>	number of equations in <code>e(b)</code>
<code>e(k_eq_model)</code>	number of equations in overall model test
<code>e(k_aux)</code>	number of auxiliary parameters
<code>e(k_dv)</code>	number of dependent variables
<code>e(df_m)</code>	model degrees of freedom
<code>e(ll)</code>	log likelihood
<code>e(ll_0)</code>	log likelihood, constant-only model (1rmodel only)
<code>e(N_clust)</code>	number of clusters
<code>e(lambda)</code>	estimate of $\lambda$ in constrained model
<code>e(selambda)</code>	standard error of $\lambda$ in constrained model
<code>e(sigma)</code>	estimate of $\sigma$ in constrained model
<code>e(lambda0)</code>	estimate of $\lambda_0$ in potential-outcome model
<code>e(selambda0)</code>	standard error of $\lambda_0$ in potential-outcome model

e(sigma0)	estimate of $\sigma_0$ in potential-outcome model
e(lambda1)	estimate of $\lambda_1$ in potential-outcome model
e(selambda1)	standard error of $\lambda_1$ in potential-outcome model
e(sigma1)	estimate of $\sigma_1$ in potential-outcome model
e(chi2)	$\chi^2$
e(chi2_c)	$\chi^2$ for comparison test
e(p)	$p$ -value for model test
e(p_c)	$p$ -value for comparison test
e(rho)	estimate of $\rho$ in constrained model
e(rho0)	estimate of $\rho_0$ in potential-outcome model
e(rho1)	estimate of $\rho_1$ in potential-outcome model
e(rank)	rank of $\mathbf{e}(V)$
e(rank0)	rank of $\mathbf{e}(V)$ 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(title2)	secondary title in estimation output
e(clustvar)	name of cluster variable
e(chi2type)	Wald or LR; type of model $\chi^2$ test
e(chi2_ct)	Wald or LR; type of model $\chi^2$ test corresponding to e(chi2_c)
e(vce)	<i>vcetype</i> specified in <code>vce()</code>
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)	m1
e(m1_method)	type of m1 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 <code>fvset</code> as <code>asbalanced</code>
e(asobserved)	factor variables <code>fvset</code> as <code>asobserved</code>

#### Matrices

e(b)	coefficient vector
e(Cns)	constraints matrix
e(ilog)	iteration log (up to 20 iterations)
e(gradient)	gradient vector
e(V)	variance–covariance matrix of the estimators
e(V_modelbased)	model-based variance

#### Functions

e(sample)	marks estimation sample
-----------	-------------------------

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

Matrices	
<code>r(table)</code>	matrix containing the coefficients with their standard errors, test statistics, $p$ -values, and confidence intervals

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

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

Scalars	
<code>e(N)</code>	number of observations
<code>e(df_m)</code>	model degrees of freedom
<code>e(lambda)</code>	$\lambda$
<code>e(selambda)</code>	standard error of $\lambda$
<code>e(sigma)</code>	estimate of sigma
<code>e(chi2)</code>	$\chi^2$
<code>e(p)</code>	$p$ -value for model test
<code>e(rho)</code>	$\rho$
<code>e(rank)</code>	rank of $e(V)$
Macros	
<code>e(cmd)</code>	<code>etregress</code>
<code>e(cmdline)</code>	command as typed
<code>e(depvar)</code>	name of dependent variable
<code>e(hazard)</code>	variable containing hazard
<code>e(title)</code>	title in estimation output
<code>e(title2)</code>	secondary title in estimation output
<code>e(chi2type)</code>	Wald or LR; type of model $\chi^2$ test
<code>e(vce)</code>	<code>vcetype</code> specified in <code>vce()</code>
<code>e(method)</code>	<code>twostep</code>
<code>e(properties)</code>	<code>b V</code>
<code>e(predict)</code>	program used to implement predict
<code>e(footnote)</code>	program used to implement the footnote display
<code>e(marginsok)</code>	predictions allowed by <code>margins</code>
<code>e(marginsnotok)</code>	predictions disallowed by <code>margins</code>
<code>e(asbalanced)</code>	factor variables <code>fvset</code> as <code>asbalanced</code>
<code>e(asobserved)</code>	factor variables <code>fvset</code> as <code>asobserved</code>
Matrices	
<code>e(b)</code>	coefficient vector
<code>e(V)</code>	variance–covariance matrix of the estimators
Functions	
<code>e(sample)</code>	marks estimation sample

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

Matrices	
<code>r(table)</code>	matrix containing the coefficients with their standard errors, test statistics, $p$ -values, and confidence intervals

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

etregress (control-function) stores the following in `e()`:

#### Scalars

<code>e(N)</code>	number of observations
<code>e(k)</code>	number of parameters
<code>e(k_eq)</code>	number of equations in <code>e(b)</code>
<code>e(k_aux)</code>	number of auxiliary parameters
<code>e(k_dv)</code>	number of dependent variables
<code>e(lambda)</code>	estimate of $\lambda$ in constrained model
<code>e(selambda)</code>	standard error of $\lambda$ in constrained model
<code>e(sigma)</code>	estimate of $\sigma$ in constrained model
<code>e(lambda0)</code>	estimate of $\lambda_0$ in potential-outcome model
<code>e(selambda0)</code>	standard error of $\lambda_0$ in potential-outcome model
<code>e(sigma0)</code>	estimate of $\sigma_0$ in potential-outcome model
<code>e(lambda1)</code>	estimate of $\lambda_1$ in potential-outcome model
<code>e(selambda1)</code>	standard error of $\lambda_1$ in potential-outcome model
<code>e(sigma1)</code>	estimate of $\sigma_1$ in potential-outcome model
<code>e(chi2_c)</code>	$\chi^2$ for comparison test
<code>e(p_c)</code>	$p$ -value for comparison test
<code>e(rho)</code>	estimate of $\rho$ in constrained model
<code>e(rho0)</code>	estimate of $\rho_0$ in potential-outcome model
<code>e(rho1)</code>	estimate of $\rho_1$ in potential-outcome model
<code>e(rank)</code>	rank of <code>e(V)</code>
<code>e(converged)</code>	1 if converged, 0 otherwise

#### Macros

<code>e(cmd)</code>	etregress
<code>e(cmdline)</code>	command as typed
<code>e(depvar)</code>	name of dependent variable
<code>e(hazard)</code>	variable containing hazard
<code>e(wtype)</code>	weight type
<code>e(wexp)</code>	weight expression
<code>e(title)</code>	title in estimation output
<code>e(title2)</code>	secondary title in estimation output
<code>e(chi2_ct)</code>	Wald; type of model $\chi^2$ test corresponding to <code>e(chi2_c)</code>
<code>e(vce)</code>	<i>vce</i> type specified in <code>vce()</code>
<code>e(vcetype)</code>	title used to label Std. err.
<code>e(method)</code>	cfunction
<code>e(properties)</code>	b V
<code>e(predict)</code>	program used to implement predict
<code>e(footnote)</code>	program used to implement the footnote display
<code>e(marginsok)</code>	predictions allowed by margins
<code>e(asbalanced)</code>	factor variables fvset as asbalanced
<code>e(asobserved)</code>	factor variables fvset as asobserved

#### Matrices

<code>e(b)</code>	coefficient vector
<code>e(V)</code>	variance–covariance matrix of the estimators

#### Functions

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

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

#### Matrices

<code>r(table)</code>	matrix containing the coefficients with their standard errors, test statistics, $p$ -values, and confidence intervals
-----------------------	---

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

## Methods and formulas

Maddala (1983, 117–122 and 223–228) derives both the maximum likelihood and the CF estimators 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.

Methods and formulas are presented under the following headings:

*Constrained model*  
*General potential-outcome model*  
*Average treatment effect*  
*Average treatment effect on the treated*

### Constrained model

The primary regression equation of interest is

$$y_j = \mathbf{x}_j\boldsymbol{\beta} + \delta t_j + \epsilon_j \quad (1)$$

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  $\epsilon$  and  $u$  are bivariate normal with mean zero and covariance matrix

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

Interactions between  $\mathbf{x}_j$  and the treatment  $t_j$  are also allowed in (1). 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  $\rho$ . The following is the log likelihood for observation  $j$ ,

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

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

In the maximum likelihood estimation,  $\sigma$  and  $\rho$  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

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

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

Maddala (1983, 120–122) also derives the CF estimator as a two-step estimator. This estimator is implemented here. We will discuss it and then discuss the one-step CF estimator that is also implemented.

For the two-step estimator, probit estimates of the treatment equation

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

are obtained in the first stage. From these estimates, the hazard,  $h_j$ , for each observation  $j$  is computed as

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

where  $\phi$  is the standard normal density function. If

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

then

$$\begin{aligned} E(y_j \mid t_j, \mathbf{x}_j, \mathbf{w}_j) &= \mathbf{x}_j \boldsymbol{\beta} + \delta t_j + \rho \sigma h_j \\ \operatorname{Var}(y_j \mid t_j, \mathbf{x}_j, \mathbf{w}_j) &= \sigma^2 (1 - \rho^2 d_j) \end{aligned}$$

The two-step parameter estimates of  $\boldsymbol{\beta}$  and  $\delta$  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  $\beta_h$  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

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

The two-step estimate of  $\rho$  is then

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

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  $(1 - \hat{\rho}^2 d_j)$  on the diagonal elements. The conventional VCE is

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

where

$$\mathbf{Q} = \hat{\rho}^2 (\mathbf{A}'\mathbf{D}\mathbf{A}) \mathbf{V}_p (\mathbf{A}'\mathbf{D}\mathbf{A})$$

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

The one-step CF estimator is a GMM estimator with stacked moments. See [Newey \(1984\)](#) and [Wooldridge \(2010, sec. 14.2\)](#) for a description of this technique. Many econometric and statistical models can be expressed as conditions on the population moments. The parameter estimates produced by GMM estimators make the sample-moment conditions as true as possible given the data.

Under CF estimation, as in maximum likelihood estimation, we directly estimate  $\text{atanh } \rho$  and  $\ln \sigma$  rather than  $\rho$  and  $\sigma$ , so the parameter vector is

$$\boldsymbol{\theta} = (\boldsymbol{\beta}', \delta, \boldsymbol{\gamma}', \text{atanh } \rho, \ln \sigma)'$$

In this case, we have separate error functions for the treatment assignment

$$u_t(t_j, \mathbf{w}_j, \boldsymbol{\theta}) = \begin{cases} \phi(\mathbf{w}_j \boldsymbol{\gamma}) / \Phi(\mathbf{w}_j \boldsymbol{\gamma}) & t_j = 1 \\ -\phi(\mathbf{w}_j \boldsymbol{\gamma}) / \{1 - \Phi(\mathbf{w}_j \boldsymbol{\gamma})\} & t_j = 0 \end{cases}$$

for the outcome mean

$$u_m(y_j, t_j, \mathbf{x}_j, \mathbf{w}_j, \boldsymbol{\theta}) = y_j - \mathbf{x}_j \boldsymbol{\beta} - \delta t_j - \rho \sigma u_{t,j}$$

and for the outcome variance

$$u_v(y_j, t_j, \mathbf{x}_j, \mathbf{w}_j, \boldsymbol{\theta}) = u_{m,j}^2 - \sigma^2 [1 - \rho^2 \{u_{t,j}(u_{t,j} + \mathbf{w}_j \boldsymbol{\gamma})\}]$$

We calculate the hazard,  $h_j$ , prior to estimation from a probit regression of the treatment  $t_j$  on the treatment covariates  $\mathbf{w}_j$ . Let  $\tilde{\mathbf{z}}_j = (\mathbf{x}_j, t_j, h_j)$ . Now we define

$$\mathbf{Z}_j = \begin{bmatrix} \tilde{\mathbf{z}}_j & \mathbf{0} & 0 \\ 0 & \mathbf{w}_j & 0 \\ 0 & \mathbf{0} & 1 \end{bmatrix}$$

and

$$s_j(y_j, t_j, \mathbf{x}_j, \mathbf{w}_j, \boldsymbol{\theta}) = \mathbf{Z}_j' \begin{bmatrix} u_{m,j} \\ u_{t,j} \\ u_{v,j} \end{bmatrix}$$

The CF estimator  $\hat{\boldsymbol{\theta}}$  is the value of  $\boldsymbol{\theta}$  that satisfies the sample-moment conditions

$$\mathbf{0} = \frac{1}{N} \sum_i s_j(y_j, t_j, \mathbf{x}_j, \mathbf{w}_j, \boldsymbol{\theta})$$

The Huber/White/robust sandwich estimator is consistent for the VCE. See [Wooldridge \(2010, chap. 14\)](#), [Cameron and Trivedi \(2005, chap. 6\)](#), and [Newey and McFadden \(1994\)](#).

The formula is

$$\hat{\mathbf{V}} = (1/N) \bar{\mathbf{G}} \bar{\mathbf{S}} \bar{\mathbf{G}}'$$

where

$$\bar{\mathbf{G}} = \left\{ (1/N) \sum_i \frac{\partial s_j(y_j, t_j, \mathbf{x}_j, \mathbf{w}_j, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right\}^{-1}$$

and

$$\bar{\mathbf{S}} = (1/N) \sum_i s_j(y_j, t_j, \mathbf{x}_j, \mathbf{w}_j, \boldsymbol{\theta}) s_j(y_j, t_j, \mathbf{x}_j, \mathbf{w}_j, \boldsymbol{\theta})'$$



The matrix  $\overline{\mathbf{G}}$  is not symmetric because our estimator comes from stacking the moment conditions instead of optimizing one objective function. The implication is that the robust formula should always be used because, even under correct specification, the nonsymmetric  $\overline{\mathbf{G}}$  and the symmetric  $\overline{\mathbf{S}}$  converge to different matrices.

## General potential-outcome model

Equation (1) can be generalized to a potential-outcome model with separate variance and correlation parameters for the control and treatment groups.

The generalized model is

$$\begin{aligned} y_{0j} &= \mathbf{x}_j \boldsymbol{\beta}_0 + \epsilon_{0j} \\ y_{1j} &= \mathbf{x}_j \boldsymbol{\beta}_1 + \epsilon_{1j} \\ t_j &= \begin{cases} 1, & \text{if } \mathbf{w}_j \boldsymbol{\gamma} + u_j > 0 \\ 0, & \text{otherwise} \end{cases} \end{aligned}$$

where  $y_{0j}$  is the outcome that person  $j$  obtains if person  $j$  selects treatment 0, and  $y_{1j}$  is the outcome that person  $j$  obtains if person  $j$  selects treatment 1. We never observe both  $y_{0j}$  and  $y_{1j}$ , only one or the other. We observe

$$y_j = t_j y_{1j} + (1 - t_j) y_{0j}$$

In this unconstrained model, the vector of error terms  $(\epsilon_{0j}, \epsilon_{1j}, u_j)'$  comes from a mean zero trivariate normal distribution with covariance matrix

$$\begin{bmatrix} \sigma_0^2 & \sigma_{01} & \sigma_0 \rho_0 \\ \sigma_{01} & \sigma_1^2 & \sigma_1 \rho_1 \\ \sigma_0 \rho_0 & \sigma_1 \rho_1 & 1 \end{bmatrix}$$

The likelihood function for this model is given in [Maddala \(1983, 224\)](#).

$$\begin{aligned} \ln f_j &= \begin{cases} \ln \Phi \left\{ \frac{\mathbf{w}_j \boldsymbol{\gamma} + (y_j - \mathbf{x}_j \boldsymbol{\beta}_1) \rho_1 / \sigma_1}{\sqrt{1 - \rho_1^2}} \right\} - \frac{1}{2} \left( \frac{y_j - \mathbf{x}_j \boldsymbol{\beta}_1}{\sigma_1} \right)^2 - \ln(\sqrt{2\pi} \sigma_1), & t_j = 1 \\ \ln \Phi \left\{ \frac{-\mathbf{w}_j \boldsymbol{\gamma} - (y_j - \mathbf{x}_j \boldsymbol{\beta}_0) \rho_0 / \sigma_0}{\sqrt{1 - \rho_0^2}} \right\} - \frac{1}{2} \left( \frac{y_j - \mathbf{x}_j \boldsymbol{\beta}_0}{\sigma_0} \right)^2 - \ln(\sqrt{2\pi} \sigma_0), & t_j = 0 \end{cases} \\ \ln L &= \sum_{j=1}^n w_j \ln f_j \end{aligned}$$

where  $\Phi(\cdot)$  is the cumulative distribution function of the standard normal distribution, and  $w_j$  is an optional weight. The covariance between  $\epsilon_{0j}$  and  $\epsilon_{1j}$ ,  $\sigma_{01}$ , cannot be estimated because the potential outcomes  $y_{0j}$  and  $y_{1j}$  are never observed simultaneously.

As in the constrained model,  $\sigma_0$  and  $\sigma_1$  are not directly estimated in the maximum likelihood estimation; rather,  $\ln \sigma_0$  and  $\ln \sigma_1$  are estimated.

The parameters  $\rho_0$  and  $\rho_1$  are also not directly estimated; rather,  $\text{atanh } \rho_0$  and  $\text{atanh } \rho_1$  are directly estimated.

The new parameter vector is

$$\boldsymbol{\theta} = (\boldsymbol{\beta}'_0, \boldsymbol{\beta}'_1, \boldsymbol{\gamma}', \tanh \rho_0, \ln \sigma_0, \tanh \rho_1, \ln \sigma_1)'$$

The CF estimator for this potential-outcome model uses new error functions for the outcome mean

$$\begin{aligned} u_m(y_j, t_j, \mathbf{x}_j, \mathbf{w}_j, \boldsymbol{\theta}) &= y_j - t_j(\mathbf{x}_j \boldsymbol{\beta}_1 + \rho_1 \sigma_1 u_{t,j}) \\ &\quad - (1 - t_j)(\mathbf{x}_j \boldsymbol{\beta}_0 + \rho_0 \sigma_0 u_{t,j}) \end{aligned}$$

and for the outcome variances

$$\begin{aligned} u_{v,0}(y_j, t_j, \mathbf{x}_j, \mathbf{w}_j, \boldsymbol{\theta}) &= (1 - t_j) (u_{m,j}^2 - \sigma_0^2 [1 - \rho_0^2 \{u_{t,j}(u_{t,j} + \mathbf{w}_j \boldsymbol{\gamma})\}]) \\ u_{v,1}(y_j, t_j, \mathbf{x}_j, \mathbf{w}_j, \boldsymbol{\theta}) &= t_j (u_{m,j}^2 - \sigma_1^2 [1 - \rho_1^2 \{u_{t,j}(u_{t,j} + \mathbf{w}_j \boldsymbol{\gamma})\}]) \end{aligned}$$

These error functions are derived based on the identities

$$\begin{aligned} E(y_j | t_j, \mathbf{x}_j, \mathbf{w}_j) &= t_j(\mathbf{x}_j \boldsymbol{\beta}_1 + \rho_1 \sigma_1 u_{t,j}) + (1 - t_j)(\mathbf{x}_j \boldsymbol{\beta}_0 + \rho_0 \sigma_0 u_{t,j}) \\ \text{Var}(y_j | t_j = 0, \mathbf{x}_j, \mathbf{w}_j) &= \sigma_0^2 [1 - \rho_0^2 \{u_{t,j}(u_{t,j} + \mathbf{w}_j \boldsymbol{\gamma})\}] \\ \text{Var}(y_j | t_j = 1, \mathbf{x}_j, \mathbf{w}_j) &= \sigma_1^2 [1 - \rho_1^2 \{u_{t,j}(u_{t,j} + \mathbf{w}_j \boldsymbol{\gamma})\}] \end{aligned}$$

We calculate the hazard,  $h_j$ , prior to estimation from a probit regression of the treatment,  $t_j$ , on the treatment covariates,  $\mathbf{w}_j$ . Let  $\tilde{\mathbf{Z}}_j = \{\mathbf{x}_j, t_j h_j, (1 - t_j) h_j\}$ . Now we define

$$\mathbf{Z}_j = \begin{bmatrix} \tilde{\mathbf{Z}}_j & \mathbf{0} & 0 & 0 \\ 0 & \mathbf{w}_j & 0 & 0 \\ 0 & \mathbf{0} & 1 & 0 \\ 0 & \mathbf{0} & 0 & 1 \end{bmatrix}$$

and

$$s_j(y_j, t_j, \mathbf{x}_j, \mathbf{w}_j, \boldsymbol{\theta}) = \mathbf{Z}'_j \begin{bmatrix} u_{m,j} \\ u_{t,j} \\ u_{v,0,j} \\ u_{v,1,j} \end{bmatrix}$$

The CF estimator  $\hat{\boldsymbol{\theta}}$  is the value of  $\boldsymbol{\theta}$  that satisfies the sample-moment conditions

$$\mathbf{0} = \frac{1}{N} \sum_i \mathbf{s}_j(y_j, t_j, \mathbf{x}_j, \mathbf{w}_j, \boldsymbol{\theta})$$

The Huber/White/robust sandwich estimator is consistent for the VCE. See Wooldridge (2010, chap. 14), Cameron and Trivedi (2005, chap. 6), and Newey and McFadden (1994).

The formula is

$$\widehat{\mathbf{V}} = (1/N) \overline{\mathbf{G}} \overline{\mathbf{S}} \overline{\mathbf{G}}'$$

where

$$\overline{\mathbf{G}} = \left\{ (1/N) \sum_i \frac{\partial s_j(y_j, t_j, \mathbf{x}_j, \mathbf{w}_j, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right\}^{-1}$$

and

$$\overline{\mathbf{S}} = (1/N) \sum_i s_j(y_j, t_j, \mathbf{x}_j, \mathbf{w}_j, \boldsymbol{\theta}) s_j(y_j, t_j, \mathbf{x}_j, \mathbf{w}_j, \boldsymbol{\theta})'$$

The matrix  $\overline{\mathbf{G}}$  is not symmetric because our estimator comes from stacking the moment conditions instead of optimizing one objective function. The implication is that the robust formula should always be used because, even under correct specification, the nonsymmetric  $\overline{\mathbf{G}}$  and the symmetric  $\overline{\mathbf{S}}$  converge to different matrices.

## Average treatment effect

The ATE is the average difference of the treated potential outcomes and the control potential outcomes.

By the law of iterated expectations, the ATE is

$$\begin{aligned} E(y_{1j} - y_{0j}) &= E\{E(y_{1j} - y_{0j} | \mathbf{x}_j, \epsilon_{0j}, \epsilon_{1j})\} \\ &= E(\mathbf{x}_j \boldsymbol{\beta}_1 + \epsilon_1 - \mathbf{x}_j \boldsymbol{\beta}_0 - \epsilon_0) \\ &= E\{\mathbf{x}_j (\boldsymbol{\beta}_1 - \boldsymbol{\beta}_0)\} \end{aligned}$$

This expectation can be estimated as a predictive margin when  $\mathbf{x}_j (\boldsymbol{\beta}_1 - \boldsymbol{\beta}_0)$  varies in  $\mathbf{x}_j$ . Otherwise, the ATE is estimated as the coefficient of  $t_j$  in the model.

## Average treatment effect on the treated

The ATE is the average difference of the treated potential outcomes and the control potential outcomes on the treated population.

The conditional means of the potential outcomes  $y_{tj}$ ,  $t \in (0, 1)$  for exogenous covariates  $\mathbf{x}_j$  and treatment covariates  $\mathbf{w}_j$  at treatment  $t_j = 1$  are

$$E(y_{tj} | \mathbf{x}_j, \mathbf{w}_j, t_j = 1) = \mathbf{x}_j \boldsymbol{\beta}_t + \rho_t \sigma_t \phi(\mathbf{w}_j \boldsymbol{\gamma}) / \Phi(\mathbf{w}_j \boldsymbol{\gamma})$$

By the law of iterated expectations, the ATET is

$$\begin{aligned} E(y_{1j} - y_{0j} | t_j = 1) &= E\{E(y_{1j} - y_{0j} | \mathbf{x}_j, \mathbf{w}_j, t_j = 1)\} \\ &= E\{\mathbf{x}_j (\boldsymbol{\beta}_1 - \boldsymbol{\beta}_0) + (\rho_1 \sigma_1 - \rho_0 \sigma_0) \phi(\mathbf{w}_j \boldsymbol{\gamma}) / \Phi(\mathbf{w}_j \boldsymbol{\gamma}) | t_j = 1\} \end{aligned}$$

This expectation can be estimated as a predictive margin on the treated population when  $\mathbf{x}_j (\boldsymbol{\beta}_1 - \boldsymbol{\beta}_0)$  varies in  $\mathbf{x}_j$  or when the variance and correlation parameters differ by treatment group. Otherwise, the ATET is estimated as the coefficient of  $t_j$  in the model.

## References

- Barnow, B. S., G. G. Cain, and A. S. Goldberger. 1981. “Issues in the analysis of selectivity bias”. In *Evaluation Studies Review Annual*, edited by E. W. Stromsdorfer and G. Farkas, vol. 5: 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.
- Cerulli, G. 2015. [ctreatreg: Command for fitting dose-response models under exogenous and endogenous treatment](#). *Stata Journal* 15: 1019–1045.
- . 2022. *Econometric Evaluation of Socio-Economic Programs: Theory and Applications*. 2nd ed. Berlin: Springer.
- 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. J. 1976. “The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models”. In *Annals of Economic and Social Measurement*, edited by S. V. Berg, vol. 5: 475–492. Cambridge, MA: National Bureau of Economic Research.
- . 1978. Dummy endogenous variables in a simultaneous equation system. *Econometrica* 46: 931–959. <https://doi.org/10.2307/1909757>.
- Maddala, G. S. 1983. *Limited-Dependent and Qualitative Variables in Econometrics*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511810176>.
- Nannicini, T. 2007. [Simulation-based sensitivity analysis for matching estimators](#). *Stata Journal* 7: 334–350.
- Newey, W. K. 1984. A method of moments interpretation of sequential estimators. *Economics Letters* 14: 201–206. [https://doi.org/10.1016/0165-1765\(84\)90083-1](https://doi.org/10.1016/0165-1765(84)90083-1).
- Newey, W. K., and D. L. McFadden. 1994. “Large sample estimation and hypothesis testing”. In *Handbook of Econometrics*, edited by R. F. Engle and D. L. McFadden, vol. 4: 2111–2245. Amsterdam: Elsevier. [https://doi.org/10.1016/S1573-4412\(05\)80005-4](https://doi.org/10.1016/S1573-4412(05)80005-4).
- Nichols, A. 2007. [Causal inference with observational data](#). *Stata Journal* 7: 507–541.
- Vella, F., and M. Verbeek. 1998. Whose wages do unions raise? A dynamic model of unionism and wage rate determination for young men. *Journal of Applied Econometrics* 13: 163–183. [https://doi.org/10.1002/\(SICI\)1099-1255\(199803/04\)13:2<163::AID-JAE460>3.0.CO;2-Y](https://doi.org/10.1002/(SICI)1099-1255(199803/04)13:2<163::AID-JAE460>3.0.CO;2-Y).
- Wooldridge, J. M. 2010. *Econometric Analysis of Cross Section and Panel Data*. 2nd ed. Cambridge, MA: MIT Press.

## Also see

- [CAUSAL] [etregress postestimation](#) — Postestimation tools for etregress
- [CAUSAL] [etpoisson](#) — Poisson regression with endogenous treatment effects
- [ERM] [eregress](#) — Extended linear regression
- [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](#)

Stata, Stata Press, and Mata are registered trademarks of StataCorp LLC. Stata and Stata Press are registered trademarks with the World Intellectual Property Organization of the United Nations. StataNow and NetCourseNow are trademarks of StataCorp LLC. Other brand and product names are registered trademarks or trademarks of their respective companies. Copyright © 1985–2025 StataCorp LLC, College Station, TX, USA. All rights reserved.



For suggested citations, see the FAQ on [citing Stata documentation](#).