> **etpoisson** — Poisson regression with endogenous treatment effects

Syntax   Menu   Description   Options
Remarks and examples   Stored results   Methods and formulas   Reference
Also see

Syntax

etpoisson *depvar* [*indepvars*] [*if*] [*in*] [*weight*] ,

   <u>treat</u>(*depvar_t* = *indepvars_t* [, <u>nocon</u>stant <u>off</u>set(*varname_o*)]) [*options*]

options	Description
[Model]	
* <u>treat</u>()	equation for treatment effects
<u>nocon</u>stant	suppress constant term
<u>exp</u>osure(*varname_e*)	include ln(*varname_e*) in model with coefficient constrained to 1
<u>off</u>set(*varname_o*)	include *varname_o* in model with coefficient constrained to 1
<u>constr</u>aints(*constraints*)	apply specified linear constraints
<u>col</u>linear	keep collinear variables
SE/Robust	
vce(*vcetype*)	*vcetype* may be oim, <u>r</u>obust, <u>cl</u>uster *clustvar*, opg, <u>boot</u>strap, or <u>jack</u>knife
Reporting	
<u>l</u>evel(*#*)	set confidence level; default is level(95)
irr	report incidence-rate ratios
<u>nocns</u>report	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
Integration	
<u>intp</u>oints(*#*)	use *#* Gauss–Hermite quadrature points; default is intpoints(24)
Maximization	
maximize_options	control the maximization process; seldom used
coeflegend	display legend instead of statistics

*treat() is required.

The full specification is <u>treat</u>(*depvar_t* = *indepvars_t* [, <u>nocon</u>stant <u>off</u>set(*varname_o*)]).

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

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

bootstrap, by, jackknife, rolling, statsby, and svy are allowed; see [U] **11.1.10 Prefix commands**.

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

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

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

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

coeflegend does not appear in the dialog box.

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

Menu

Statistics > Treatment effects > Endogenous treatment estimators > Count outcome

Description

etpoisson estimates the parameters of a Poisson regression model in which one of the regressors is an endogenous binary treatment. Both the average treatment effect and the average treatment effect on the treated can be estimated with etpoisson.

Options

> Model

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

The indicator of treatment, *depvar*$_t$, should be coded as 0 or 1.

noconstant, exposure(*varname*$_e$), offset(*varname*$_o$), 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**.

irr reports estimated coefficients transformed to incidence-rate ratios, that is, e^{β_i} rather than β_i. Standard errors and confidence intervals are similarly transformed. This option affects how results are displayed, not how they are estimated or stored. irr may be specified at estimation or when replaying previously estimated results.

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

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

> Integration

`intpoints(#)` specifies the number of integration points to use for integration by quadrature. The default is `intpoints(24)`; the maximum is `intpoints(128)`. Increasing this value improves the accuracy but also increases computation time. Computation time is roughly proportional to its value.

> 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 `etpoisson` 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

`etpoisson` estimates the parameters of a Poisson regression model that includes an endogenous binary-treatment variable. The dependent variable must be a Poisson distributed count. The parameters estimated by `etpoisson` can be used to estimate the average treatment effect (ATE) and average treatment effect on the treated (ATET).

We call the model fit by `etpoisson` 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 fit by `etpoisson` is a specific endogenous treatment-effects model; it uses a nonlinear model for the outcome and a constrained normal distribution to model the deviation from the conditional independence assumption imposed by the estimators implemented by `teffects`; see [TE] **teffects intro**. In treatment-effects jargon, the endogenous binary-variable model fit by `etpoisson` is a nonlinear 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] **etregress** for an estimator that allows for a linear-outcome model and a similar model for the endogeneity of the treatment.

More formally, we have an equation for outcome y_j and an equation for treatment t_j:

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

The $\mathbf{x}_j$ are the covariates used to model the outcome, $\mathbf{w}_j$ are the covariates used to model treatment assignment, and error terms ϵ_j and u_j are bivariate normal with mean 0 and covariance matrix

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

The covariates $\mathbf{x}_j$ and $\mathbf{w}_j$ are unrelated to the error terms; in other words, they are exogenous. Note that y_j may be a count or continuous and nonnegative in this specification.

Terza (1998) describes the maximum likelihood estimator used in etpoisson. Terza (1998) categorized the model fit by etpoisson as an endogenous-switching model. These models involve a binary switch that is endogenous for the outcome. Calculation of the maximum likelihood estimate involves numeric approximation of integrals via Gauss–Hermite quadrature. This is computationally intensive, but the computational costs are reasonable on modern computers.

Basic example

▷ Example 1

In this example, we observe a simulated random sample of 5,000 households. The outcome of interest is the number of trips taken by members of the household in the 24-hour period immediately prior to the interview time.

We have fictional household level data on the following variables: number of trips taken in the past 24 hours (trips), distance to the central business district from the household (cbd), distance from the household to a public transit node (ptn), an indicator of whether there is a full-time worker in the household (worker), an indicator of whether the examined period is on a weekend (weekend), the ratio of the household income to the median income of the census tract (realinc), and an indicator of car ownership (owncar). We suspect that unobservables that affect the number of trips also affect the household's propensity to own a car.

We use etpoisson to estimate the parameters of a Poisson regression model for the number of trips with car ownership as an endogenous treatment. In subsequent examples, we will use margins (see [R] **margins**) to estimate the ATE and the ATET of car ownership on the number of trips taken by the household. In the etpoisson command below, we specify the vce(robust) option because we need to specify vce(unconditional) when we use margins later.

```
. use http://www.stata-press.com/data/r13/trip1
(Household trips, car ownership)

. etpoisson trips cbd ptn worker weekend,
> treat(owncar = cbd ptn worker realinc) vce(robust)

Iteration 0:   log pseudolikelihood = -14845.147  (not concave)
Iteration 1:   log pseudolikelihood = -14562.997  (not concave)
Iteration 2:   log pseudolikelihood = -13655.592  (not concave)
Iteration 3:   log pseudolikelihood = -12847.219  (not concave)
Iteration 4:   log pseudolikelihood = -12566.037
Iteration 5:   log pseudolikelihood = -12440.974
Iteration 6:   log pseudolikelihood = -12413.485
Iteration 7:   log pseudolikelihood = -12412.699
Iteration 8:   log pseudolikelihood = -12412.696
Iteration 9:   log pseudolikelihood = -12412.696
```

Poisson regression with endogenous treatment	Number of obs	=	5000
(24 quadrature points)	Wald chi2(5)	=	397.94
Log pseudolikelihood = -12412.696	Prob > chi2	=	0.0000

	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf. Interval]	
trips						
cbd	-.0100919	.0020071	-5.03	0.000	-.0140258	-.006158
ptn	-.0204038	.0020289	-10.06	0.000	-.0243805	-.0164272
worker	.692301	.0548559	12.62	0.000	.5847854	.7998166
weekend	.0930517	.034538	2.69	0.007	.0253585	.160745
1.owncar	.5264713	.1124157	4.68	0.000	.3061406	.746802
_cons	-.2340772	.0810812	-2.89	0.004	-.3929934	-.0751609
owncar						
cbd	.007218	.00239	3.02	0.003	.0025337	.0119023
ptn	.0084769	.0024518	3.46	0.001	.0036714	.0132824
worker	.543643	.0504267	10.78	0.000	.4448085	.6424774
realinc	.176479	.0108746	16.23	0.000	.1551652	.1977928
_cons	-.4611246	.0592161	-7.79	0.000	-.5771859	-.3450633
/athrho	.5741169	.0957832	5.99	0.000	.3863852	.7618486
/lnsigma	-.2182037	.0256281	-8.51	0.000	-.2684338	-.1679735
rho	.5183763	.0700449			.3682398	.6421645
sigma	.8039617	.020604			.764576	.8453762

Wald test of indep. eqns. (rho = 0): chi2(1) =   35.93   Prob > chi2 = 0.0000

The Wald test in the header is highly significant, indicating a good model fit. All the covariates are statistically significant, and the Wald test in the footer indicates that we can reject the null hypothesis of no correlation between the treatment errors and the outcome errors.

We can interpret the coefficient on 1.owncar as the logarithm of the ratio of the treatment potential-outcome mean to the control potential-outcome mean. The treatment variable did not interact with any of the outcome covariates, so the effect of each regressor is the same in the two regimes and will cancel from the ratio of potential-outcome means. This means the ratio is equivalent to the exponentiated coefficient on 1.owncar. After discussing the other parameters, we will use lincom to obtain this ratio. See [R] **lincom** for more information.

The estimated correlation between the treatment-assignment errors and the outcome errors is 0.518, indicating that unobservables that increase the number of trips tend to occur with unobservables that increase the chance of car ownership.

The results for the two ancillary parameters require explanation. etpoisson estimates the inverse hyperbolic tangent of ρ,

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

and $\ln\sigma$ rather than ρ and σ. For numerical stability during optimization, etpoisson does not directly estimate ρ and σ.

Now we use lincom and the eform option to estimate the exponentiated coefficient for 1.owncar. This corresponds to the ratio of the treatment regime potential-outcome mean to the control regime potential-outcome mean.

```
. lincom [trips]_b[1.owncar], eform
( 1)  [trips]1.owncar = 0
```

	exp(b)	Std. Err.	z	P>|z|	[95% Conf. Interval]	
(1)	1.692948	.1903139	4.68	0.000	1.358173	2.110241

The potential-outcome mean for the treatment regime is 1.69 times the potential-outcome mean for the control regime. So the average number of trips in the treatment regime is over one and a half times the average number of trips in the control regime.

By interacting the treatment, owncar, with the other regressors, we could estimate different coefficients for the regressors in the treatment and control regimes. In the current model, there are no treatment interactions, so the coefficients are the same in each regime.

◁

Average treatment effect (ATE)

The parameter estimates from etpoisson can be used by margins to estimate the ATE, the average difference of the treatment and control potential outcomes.

▷ Example 2

Continuing with example 1, we use margins to estimate the ATE of car ownership on the number of trips taken in a 24-hour period.

We can estimate the ATE of car ownership by using the potential-outcome means obtained through the predict, pomean command and the margins command; see *Methods and formulas* below and [TE] **etpoisson postestimation** for more details about the use of predict after etpoisson.

The r. notation indicates that the potential-outcome means for treatment and control will be contrasted. We specify the contrast(nowald) option to suppress the Wald tests that margins displays by default for contrasts.

```
. margins r.owncar, vce(unconditional) contrast(nowald)
Contrasts of predictive margins
Expression   : Potential-outcome mean, predict()
```

		Unconditional		
	Contrast	Std. Err.	[95% Conf. Interval]	
owncar (1 vs 0)	1.058914	.1922909	.6820309	1.435797

The estimated ATE of car ownership on the number of trips taken is 1.06. The average household will take 1.06 more trips when it owns a car.

◁

Average treatment effect on the treated (ATET)

The parameter estimates from etpoisson can be used by margins to estimate the ATET, the average difference of the treatment and control potential outcomes in the treated population.

▷ Example 3

Continuing with the previous example, we use margins to estimate the ATET of car ownership on the number of trips taken in a 24-hour period.

We can estimate the ATET of car ownership by using the conditional treatment effect (conditional on exogenous covariates and treatment level) obtained through the predict, cte command and the margins command; see *Methods and formulas* below and [TE] **etpoisson postestimation** for more details about the use of predict after etpoisson.

We estimate the ATET with margins. We specify cte in the predict() option. Estimation is restricted to the treated subpopulation by specifying owncar in the subpop() option.

```
. margins, predict(cte) vce(unconditional) subpop(owncar)
Predictive margins                              Number of obs   =       5000
                                                Subpop. no. obs =       3504
Expression   : Conditional treatment effect, predict(cte)
```

	Margin	Unconditional Std. Err.	z	P>|z|	[95% Conf. Interval]	
_cons	1.251971	.2059201	6.08	0.000	.8483747	1.655567

The estimated ATET of car ownership on the number of trips taken is 1.25. Thus the average household in the treated population will take 1.25 more trips than it would if it did not own a car. This number is higher than the ATE. In this model, the ATE and ATET will only coincide when there is no correlation between the treatment errors and outcome errors and the exogenous covariates x have the same distribution in the general population and treated subpopulation. See *Methods and formulas* for more details.

◁

Stored results

etpoisson 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(N_clust)	number of clusters
e(chi2)	χ^2
e(chi2_c)	χ^2 for comparison, $\rho=0$ test
e(n_quad)	number of quadrature points
e(p)	significance
e(p_c)	significance of comparison test
e(rank)	rank of e(V)
e(ic)	number of iterations
e(rc)	return code
e(converged)	1 if converged, 0 otherwise

Macros
e(cmd)	etpoisson
e(cmdline)	command as typed
e(depvar)	name of dependent variable
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(offset1)	offset for regression equation
e(offset2)	offset for treatment equation
e(chi2type)	Wald; type of model χ^2 test
e(chi2_ct)	Wald; type of comparison χ^2 test
e(vce)	*vcetype* specified in vce()
e(vcetype)	title used to label Std. Err.
e(opt)	type of optimization
e(which)	max or min; whether optimizer is to perform maximization or minimization
e(ml_method)	type of ml method
e(user)	name of likelihood-evaluator program
e(technique)	maximization technique
e(properties)	b V
e(predict)	program used to implement predict
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

Methods and formulas

Terza (1998) derives the maximum likelihood estimator implemented here. We provide some details of the derivation and then explain how the model is nested in a more general potential-outcomes model. Then the ATE and ATET are derived.

Let $\mathbf{x}_j$ be the covariates used to model the outcome, and let $\mathbf{w}_j$ be the covariates used to model treatment assignment. Define $\mathbf{z}_j = (\mathbf{w}_j, \mathbf{x}_j)$. The vector $\mathbf{z}_j$ contains all the exogenous covariates in the model. When offsets o_j^β are used in the outcome variable equation, the following formulas apply with $\mathbf{x}_j\boldsymbol{\beta}$ changed to $\mathbf{x}_j\boldsymbol{\beta} + o_j^\beta$. Similarly, when offsets o_j^γ are used in the endogenous treatment equation, the following formulas apply with $\mathbf{w}_j\boldsymbol{\gamma}$ changed to $\mathbf{w}_j\boldsymbol{\gamma} + o_j^\gamma$. If offsets are used in either equation, they are included in the vector of exogenous covariates $\mathbf{z}_j$.

For treatment t_j, $\mathbf{z}_j$, and ϵ_j, outcome y_j of this model has conditional mean

$$E(y_j|\mathbf{x}_j, t_j, \epsilon_j) = \exp(\mathbf{x}_j\boldsymbol{\beta} + \delta t_j + \epsilon_j) \tag{1}$$

The probability density function of y_j for this model, conditioned on treatment t_j, $\mathbf{z}_j$, and ϵ_j, is given by

$$f(y_j|\mathbf{z}_j, t_j, \epsilon_j) = \frac{\exp\{-\exp(\mathbf{x}_j\boldsymbol{\beta} + \delta t_j + \epsilon_j)\}\{\exp(\mathbf{x}_j\boldsymbol{\beta} + \delta t_j + \epsilon_j)\}^{y_j}}{y_j!}$$

The treatment t_j is determined by

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

The error terms ϵ_j and u_j are bivariate normal with mean zero and covariance matrix

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

Conditional on ϵ_j, u_j is normal with mean $\epsilon_j\rho/\sigma$ and variance $(1 - \rho^2)$; thus we obtain the following conditional probability density for t_j:

$$\Pr(t_j|\mathbf{z}_j, \epsilon_j) = t_j\Phi\left\{\frac{\mathbf{w}_j\boldsymbol{\gamma} + (\rho/\sigma)\epsilon_j}{\sqrt{1-\rho^2}}\right\} + (1 - t_j)\left[1 - \Phi\left\{\frac{\mathbf{w}_j\boldsymbol{\gamma} + (\rho/\sigma)\epsilon_j}{\sqrt{1-\rho^2}}\right\}\right]$$

Φ denotes the standard normal cumulative distribution function. This leads to the following joint density of y_j, t_j, and ϵ_j:

$$f(y_j, t_j, \epsilon_j|\mathbf{z}_j) = f(y_j|\mathbf{z}_j, t_j, \epsilon_j)P(t_j|\mathbf{z}_j, \epsilon_j)f(\epsilon_j)$$

The density of y_j and t_j, conditioned on $\mathbf{z}_j$, is obtained by integrating the above with respect to ϵ_j. Recall that ϵ_j is normal with mean 0 and variance σ^2.

$$f(y_j, t_j|\mathbf{z}_j) = \int_{-\infty}^{\infty} f(y_j|\mathbf{z}_j, t_j, \epsilon_j)P(t_j|\mathbf{z}_j, \epsilon_j)\frac{1}{\sigma\sqrt{2\pi}}\exp\left\{-\left(\frac{\epsilon_j}{\sigma\sqrt{2}}\right)^2\right\}d\epsilon_j$$

$f(y_j, t_j | \mathbf{z}_j)$ cannot be evaluated in a closed form. We change the variable of integration from ϵ_j to $\eta_j = \epsilon_j / (\sigma\sqrt{2})$, which yields

$$f(y_j, t_j | \mathbf{z}_j) = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} f(y_j | \mathbf{z}_j, t_j, \sqrt{2}\sigma\eta_j) P(t_j | \mathbf{z}_j, \sqrt{2}\sigma\eta_j) \exp(-\eta_j^2) \, d\eta_j$$

We approximate this integral by Gauss–Hermite quadrature. Observing a sample of t_j, y_j, and $\mathbf{z}_j$, we calculate the log likelihood as the following:

$$\ln L = \sum_{j=1}^{n} w_j \ln \{ f(y_j, t_j | \mathbf{z}_j) \}$$

The w_j terms denote optional weights.

In the maximum likelihood estimation, σ and ρ are not directly estimated. Directly estimated are $\ln \sigma$ and $\mathrm{atanh}\, \rho$:

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

Now we present formulas for the ATE and ATET. First, we nest the endogenous-treatment Poisson regression model in a potential-outcome model. A potential-outcome model specifies what each individual would obtain in each treatment level.

A potential-outcome model that nests the endogenous-treatment Poisson regression fit by `etpoisson` is

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

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. This formulation allows differing coefficients for the control ($\boldsymbol{\beta}_0$) and treatment ($\boldsymbol{\beta}_1$) regimes. The constant intercept for the control group is β_{00}. The constant intercept for the treatment group is $\beta_{11} = \beta_{00} + \delta$, where δ is the coefficient for treatment t_j in the outcome (1). The remaining notation was defined above.

We may allow other coefficients to differ across regimes in the outcome (1) by adding interactions between the treatment t_j and covariates $\mathbf{x}_j$ to the model. To be concise, we use two coefficient vectors $\boldsymbol{\beta}_0$ and $\boldsymbol{\beta}_1$ here rather than a single coefficient vector with interactions between the treatment t_j and covariates $\mathbf{x}_j$. The two formulations are equivalent.

We never observe both y_{0j} and y_{1j}, only one or the other. We observe

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

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

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

The parameters σ and ρ were discussed earlier. The parameter θ is the covariance between the two potential outcomes. We cannot estimate θ because we have no observations in which an individual is observed in both potential outcomes. Fortunately, θ is not required for the calculations that we present.

The ATE is the difference in means of the potential outcomes. The mean of each potential outcome accounts for each individual's contribution, regardless of whether that individual selects that treatment level.

The conditional means of the potential outcomes y_{tj}, $t \in (0, 1)$ for exogenous covariates $\mathbf{z}_j$ are

$$
E(y_{tj}|\mathbf{z}_j) = \exp\left(\mathbf{x}_j\boldsymbol{\beta}_t + \frac{\sigma^2}{2}\right)
$$

We can see that when the coefficients are the same across the regimes, the ratio of potential-outcome means will be equal to $\exp(\delta)$; this is true of the conditional and marginal potential-outcome means.

The difference in potential-outcome means or treatment effect at exogenous covariates $\mathbf{z}_j$ is

$$
E(y_{1j} - y_{0j}|\mathbf{z}_j) = \{\exp(\mathbf{x}_j\boldsymbol{\beta}_1) - \exp(\mathbf{x}_j\boldsymbol{\beta}_0)\}\exp\left(\frac{\sigma^2}{2}\right)
$$

By the law of iterated expectations, the ATE is

$$
\begin{aligned}
E(y_{1j} - y_{0j}) &= E\{E(y_{1j} - y_{0j}|\mathbf{z}_j)\} \\
&= E\left[\{\exp(\mathbf{x}_j\boldsymbol{\beta}_1) - \exp(\mathbf{x}_j\boldsymbol{\beta}_0)\}\exp\left(\frac{\sigma^2}{2}\right)\right]
\end{aligned}
$$

This expectation can be estimated as a predictive margin.

Now we will derive an expression for the ATET.

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

$$
E(y_{tj}|\mathbf{z}_j, t_j) = \exp\left(\mathbf{x}_j\boldsymbol{\beta}_t + \frac{\sigma^2}{2}\right)\left\{\frac{\Phi(\rho\sigma + \mathbf{w}_j\boldsymbol{\gamma})}{\Phi(\mathbf{w}_j\boldsymbol{\gamma})}\right\}^{t_j}\left\{\frac{1 - \Phi(\rho\sigma + \mathbf{w}_j\boldsymbol{\gamma})}{1 - \Phi(\mathbf{w}_j\boldsymbol{\gamma})}\right\}^{1-t_j}
$$

Rather than the conditional potential-outcome means, the conditional mean of the observed outcome may be of interest. The conditional mean of the observed outcome y_j for endogenous treatment indicator t_j and exogenous covariates $\mathbf{z}_j$ is given by

$$E(y_j|\mathbf{z}_j, t_j) = t_j \exp\left(\mathbf{x}_j\boldsymbol{\beta}_1 + \frac{\sigma^2}{2}\right) \frac{\Phi\left(\rho\sigma + \mathbf{w}_j\boldsymbol{\gamma}\right)}{\Phi\left(\mathbf{w}_j\boldsymbol{\gamma}\right)}$$
$$+ (1 - t_j) \exp\left(\mathbf{x}_j\boldsymbol{\beta}_0 + \frac{\sigma^2}{2}\right) \frac{1 - \Phi\left(\rho\sigma + \mathbf{w}_j\boldsymbol{\gamma}\right)}{1 - \Phi\left(\mathbf{w}_j\boldsymbol{\gamma}\right)}$$

The treatment effect at exogenous covariates $\mathbf{z}_j$ and treatment t_j is

$$E(y_{1j} - y_{0j}|\mathbf{z}_j, t_j) =$$
$$\{\exp\left(\mathbf{x}_j\boldsymbol{\beta}_1\right) - \exp\left(\mathbf{x}_j\boldsymbol{\beta}_0\right)\} \exp\left(\frac{\sigma^2}{2}\right) \left\{\frac{\Phi\left(\rho\sigma + \mathbf{w}_j'\boldsymbol{\gamma}\right)}{\Phi\left(\mathbf{w}_j\boldsymbol{\gamma}\right)}\right\}^{t_j} \left\{\frac{1 - \Phi\left(\rho\sigma + \mathbf{w}_j\boldsymbol{\gamma}\right)}{1 - \Phi\left(\mathbf{w}_j'\boldsymbol{\gamma}\right)}\right\}^{1 - t_j}$$

By the law of iterated expectations, the ATET is

$$E(y_{1j} - y_{0j}|t_j = 1) = E\{E(y_{1j} - y_{0j}|\mathbf{z}_j, t_j = 1)|t_j = 1\}$$
$$= E\left[\{\exp\left(\mathbf{x}_j\boldsymbol{\beta}_1\right) - \exp\left(\mathbf{x}_j\boldsymbol{\beta}_0\right)\} \exp\left(\frac{\sigma^2}{2}\right) \frac{\Phi\left(\rho\sigma + \mathbf{w}_j\boldsymbol{\gamma}\right)}{\Phi\left(\mathbf{w}_j\boldsymbol{\gamma}\right)}\bigg|t_j = 1\right]$$

This can be estimated as a predictive margin on the treated subpopulation.

We note that when $\rho = 0$, the correction factor involving Φ will disappear from the ATET. Then the ATE and ATET will be equivalent if the distribution of $\mathbf{x}_j$ under the treated population is identical to the distribution over the entire population.

The probability of y_j conditional on t_j and $\mathbf{z}_j$ is

$$\Pr(y_j = n|\mathbf{z}_j, t_j) = \frac{f(y_j = n, t_j|\mathbf{z}_j)}{\Phi\left(\mathbf{w}_j\boldsymbol{\gamma}\right)^{t_j} \Phi\left(-\mathbf{w}_j\boldsymbol{\gamma}\right)^{1 - t_j}}$$

As discussed earlier, we approximate $f(y_j, t_j|\mathbf{z}_j)$ using Gauss–Hermite quadrature.

Reference

Terza, J. V. 1998. Estimating count data models with endogenous switching: Sample selection and endogenous treatment effects. *Journal of Econometrics* 84: 129–154.

Also see

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

[TE] **etregress** — Linear regression with endogenous treatment effects

[R] **ivpoisson** — Poisson regression with endogenous regressors

[R] **ivprobit** — Probit model with continuous endogenous regressors

[R] **ivregress** — Single-equation instrumental-variables regression

[R] **ivtobit** — Tobit model with continuous endogenous regressors

[R] **poisson** — Poisson regression

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

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