



*The World's Largest Open Access Agricultural & Applied Economics Digital Library*

**This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.**

**Help ensure our sustainability.**

Give to AgEcon Search

AgEcon Search

<http://ageconsearch.umn.edu>

[aesearch@umn.edu](mailto:aesearch@umn.edu)

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*

*No endorsement of AgEcon Search or its fundraising activities by the author(s) of the following work or their employer(s) is intended or implied.*

# Endogenous switching regression model and treatment effects of count-data outcome

Takuya Hasebe  
Sophia University  
Tokyo, Japan  
thasebe@sophia.ac.jp

**Abstract.** In this article, I describe the `escount` command, which implements the estimation of an endogenous switching model with count-data outcomes, where a potential outcome differs across two alternate treatment statuses. `escount` allows for either a Poisson or a negative binomial regression model with lognormal latent heterogeneity. After estimating the parameters of the switching regression model, one can estimate various treatment effects with the command `teescount`. I also describe the command `lncount`, which fits the Poisson or negative binomial regression model with lognormal latent heterogeneity.

**Keywords:** `st0612`, `escount`, `lncount`, `teescount`, self-selection, count data, treatment effects

## 1 Introduction

Count data, which are nonnegative and integer, are common in empirical studies of health economics and other applied microeconomics. Another common feature of empirical studies of applied microeconomics is a selection issue. A self-selection issue is particularly relevant in estimating effects of a certain treatment when an individual chooses to receive treatment depending on individual-specific heterogeneity that is unobservable by empirical researchers. Similarly, the problem of external selection may also arise when an external entity, such as a public agency, cherry-picks specific individuals as part of a treatment group. Such selection on unobservables complicates causal inference of treatment effects.

In this article, I present the command `escount`, which implements the maximum likelihood estimation (MLE) of the endogenous-switching regression model with count-data outcomes proposed by Terza (1998). In this model, potential outcomes, which are count data, differ across two alternate treatment statuses, and the treatment status is endogenously determined. Potential outcomes are not independent of treatment even after controlling for observable characteristics. To model the dependence between the potential outcome and the selection process, one relies on the assumption of lognormal latent heterogeneity. As Greene (2009) discusses, the lognormal latent heterogeneity provides a versatile specification. In addition to the command `escount`, I introduce the command `lncount`, which fits the regression model of count-data outcome with the lognormal latent heterogeneity. `lncount` is used to find initial values for the MLE of the switching regression model, but it can be a useful alternative specification of a Poisson regression (`poisson`) and a negative binomial (NB) regression (`nbreg`).

The virtue of a switching regression model is that it enables us to derive various treatment effects straightforwardly. Building on a switching regression model, Heckman, Tobias, and Vytlačil (2003) derive various treatment-effect parameters for continuous outcomes. Hasebe (2018) derives the count-data counterparts of these treatment-effect estimators. The treatment effects can be estimated with the command `teescount` after executing `escount`.

The structure of this article is as follows. The next section outlines the switching regression model with count-data outcomes and the MLE of the model. In section 3, I briefly discuss various treatment effects proposed in the literature and show the expressions of these treatment effects based on the parameters of the switching regression model. In section 4, I describe the commands `escount`, `lncount`, and `teescount`. In section 5, I discuss data applications. In section 6, I conclude the article.

## 2 Model

We consider a model of potential outcomes. There are two potential outcomes,  $y_{0i}$  and  $y_{1i}$ , which are count data, that is, nonnegative integers, for an individual  $i$ ,  $i = 1, \dots, N$ . Depending on a treatment status,  $d_i \in \{0, 1\}$ , we observe one of the two outcomes but not both simultaneously for an individual  $i$ . That is, an observed outcome  $y_i$  is  $y_i = (1 - d_i)y_{0i} + d_i y_{1i}$ . The selection problem arises when  $y_{0i}$  and  $y_{1i}$  are not independent of  $d_i$  even after conditioning on observable characteristics  $\mathbf{x}_i$ .

We consider a model with parametric distributional assumptions originating in Terza (1998). We assume lognormal latent heterogeneity to model the dependence between  $y_{ji}$  and  $d_i$  for  $j = 0, 1$ . Assume that  $y_{ji}$  follows either a Poisson or an NB distribution conditional on  $x_i$  and  $\varepsilon_{ji}$ , which is latent heterogeneity. The selection mechanism depends on observable characteristics  $\mathbf{z}_i$  and an unobservable term  $\nu_i$ :  $d_i = 1(\mathbf{z}_i' \boldsymbol{\gamma} + \nu_i > 0)$ , where  $1(\cdot)$  is an indicator function.  $y_j$  and  $d$  are dependent on each other through the correlation between  $\varepsilon_j$  and  $\nu$ . We assume the joint normality of  $\nu$ ,  $\varepsilon_0$ , and  $\varepsilon_1$ ,

$$\begin{pmatrix} \nu \\ \varepsilon_0 \\ \varepsilon_1 \end{pmatrix} \sim N \left( \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho_0 \sigma_0 & \rho_1 \sigma_1 \\ \rho_0 \sigma_0 & \sigma_0^2 & \rho_{01} \sigma_0 \sigma_1 \\ \rho_1 \sigma_1 & \rho_{01} \sigma_0 \sigma_1 & \sigma_1^2 \end{bmatrix} \right)$$

where  $\sigma_j$  is a standard deviation of  $\varepsilon_j$  and  $\rho_j$  is the coefficient of correlation between  $\nu$  and  $\varepsilon_j$  for  $j = 0, 1$ . The parameter  $\rho_{01}$  is the coefficient of correlation between  $\varepsilon_0$  and  $\varepsilon_1$ , but it is not identified in the model. Let  $\phi(\cdot)$  and  $\Phi(\cdot)$  be the probability density function and cumulative distribution function of a standard normal distribution, respectively. Under the normality assumption, the joint probability  $f_j(y_i, d_i | \mathbf{x}_i, \mathbf{z}_i)$  can be written as follows,

$$f_j(y_i, d_i | \mathbf{x}_i, \mathbf{z}_i) = \int_{-\infty}^{\infty} f_j(y_i | \mathbf{x}_i, \varepsilon_j) \Phi \left\{ \frac{(2d_i - 1)(\mathbf{z}_i' \boldsymbol{\gamma} + \rho_j \sigma_j^{-1} \varepsilon_j)}{\sqrt{1 - \rho_j^2}} \right\} \phi(\varepsilon_j) d\varepsilon_j$$

where  $f_j(y_i|\mathbf{x}_i, \varepsilon_j)$  is

$$f_j(y_i|\mathbf{x}_i, \varepsilon_j) = \frac{\exp(\mathbf{x}_i' \boldsymbol{\beta}_j + \varepsilon_j)^{y_i} \exp\{-\exp(\mathbf{x}_i' \boldsymbol{\beta}_j + \varepsilon_j)\}}{y_i!}$$

for a Poisson distribution and

$$f_j(y_i|\mathbf{x}_i, \varepsilon_j) = \frac{\Gamma(1/\alpha_j + y_i)}{\Gamma(y_i + 1) \Gamma(1/\alpha_j)} \times \{\alpha_j \exp(\mathbf{x}_i' \boldsymbol{\beta}_j + \varepsilon_j)\}^{y_i} \{1 - \alpha_j \exp(\mathbf{x}_i' \boldsymbol{\beta}_j + \varepsilon_j)\}^{-(y_i + 1/\alpha_j)}$$

for an NB distribution with the overdispersion parameter  $\alpha_j$ .

The command **escount** estimates the parameters  $\boldsymbol{\theta} = (\boldsymbol{\beta}_0', \boldsymbol{\beta}_1', \boldsymbol{\gamma}', \sigma_0, \sigma_1, \rho_0, \rho_1)'$  using the MLE method. In the case of an NB distribution,  $\boldsymbol{\theta}$  additionally includes  $\alpha_0$  and  $\alpha_1$ . The log-likelihood function has the following form:

$$\ln L = \sum_{i=1}^N (1 - d_i) \times \ln f_0(y_i, d_i|\mathbf{x}_i, \mathbf{z}_i) + d_i \times \ln f_1(y_i, d_i|\mathbf{x}_i, \mathbf{z}_i)$$

The integration of  $f_j(y_i, d_i|\mathbf{x}_i, \mathbf{z}_i)$  is evaluated numerically via Gauss–Hermite quadrature. To obtain initial values for  $\boldsymbol{\gamma}$ , we fit the probit model of the treatment status  $d_i$ . Initial values for  $\boldsymbol{\beta}_j$  and  $\sigma_j$  (and  $\alpha_j$  for the NB distribution) are obtained from the model without the selection process, of which the log-likelihood function is  $\ln L = \sum_{i=1}^N \int_{-\infty}^{\infty} f(y_i|\mathbf{x}_i, \varepsilon) d\varepsilon$ , using subsamples with  $d_i = 1$  and  $d_i = 0$  separately. The MLE routine for this model is named **lncount**.

This switching model is a generalization of the Poisson regression model with an endogenous treatment dummy variable that the command **etpoisson** estimates. While the endogenous dummy variable model restricts  $\sigma_0 = \sigma_1$  and  $\rho_0 = \rho_1$ , the switching model allows for more flexible behaviors of unobservable heterogeneity. Because the endogenous treatment dummy variable is more parsimonious, it usually has more precise estimates than the switching model. However, if the restrictions on  $\rho$  and  $\sigma$  do not hold, the endogenous treatment dummy variable model is misspecified and suffers from biases. In such cases, the switching model is preferable. In practice, applied researchers do not usually know the behaviors of unobservable heterogeneity in advance. One of the advantages of the switching model is that because it nests the endogenous dummy variable model, we can test the restrictions easily with Wald and likelihood-ratio tests after fitting the switching model. In this sense, the command **escount** provides applied researchers with not only an additional model option but also a model-selection tool. Furthermore, **escount** allows for an NB specification, while **etpoisson** implements only a Poisson specification.

### 3 Treatment effects

Several treatment effects have been defined to allow for heterogeneous effects of the treatment among different populations in the literature. Estimating heterogeneity of

treatment effects is useful and important to conduct policy analysis. See, for example, Heckman (2010). Heckman, Tobias, and Vytlačil (2003) derive the estimators of treatment effects for a continuous outcome based on a switching regression model under the assumption of joint normality. Building on the switching regression model with count-data outcomes, Hasebe (2018) derives the estimators of the average treatment effect (ATE), average treatment effect on the treated (ATT), average treatment effect on the untreated (ATU), local average treatment effect (LATE), and marginal treatment effect (MTE). This article shows the derived expressions. The following expressions of the treatment effects are the same regardless of whether the conditional distribution of  $y_j$  is either Poisson or NB. After one executes the command `escount`, the command `teescount` estimates these treatment effects.

ATE measures the expected effect of the treatment for an individual randomly chosen from the entire population. Conditional on  $\mathbf{x}$ ,  $\mu_{\text{ATE}}(\mathbf{x}) = E(y_1 - y_0 | \mathbf{x})$ . In the lognormal latent heterogeneity model, the expectation of each potential outcome  $y_j$  conditional on  $\mathbf{x}$  can be expressed as  $E(y_j | \mathbf{x}) = \exp(\mathbf{x}'\boldsymbol{\beta}_j + \sigma_j^2/2)$ . That is,  $\mu_{\text{ATE}}(\mathbf{x}) = \exp(\mathbf{x}'\boldsymbol{\beta}_1 + \sigma_1^2/2) - \exp(\mathbf{x}'\boldsymbol{\beta}_0 + \sigma_0^2/2)$ . The estimated effect is obtained by replacing the population parameters with  $\hat{\boldsymbol{\beta}}_0, \hat{\boldsymbol{\beta}}_1, \hat{\sigma}_0$ , and  $\hat{\sigma}_1$ . The unconditional ATE is estimated by averaging the whole sample:  $\hat{\mu}_{\text{ATE}} = N^{-1} \sum_{i=1}^N \hat{\mu}_{\text{ATE}}(\mathbf{x}_i)$ .

The treatment effects may be different between those who select the treatment and those who do not. Let  $\mathbf{w}$  be a union of  $\mathbf{x}$  and  $\mathbf{z}$ :  $\mathbf{w} = \mathbf{x} \cup \mathbf{z}$ . ATT is the average effect of the treatment for those who actually select the treatment:  $\mu_{\text{ATT}}(\mathbf{w}) = E(y_1 - y_0 | \mathbf{x}, d=1) = E(y_1 - y_0 | \mathbf{x}, \nu > -\mathbf{z}'\boldsymbol{\gamma})$ . In contrast, ATU measures the average effect for those who do not receive the treatment:  $\mu_{\text{ATU}}(\mathbf{w}) = E(y_1 - y_0 | \mathbf{x}, d=0) = E(y_1 - y_0 | \mathbf{x}, \nu \leq -\mathbf{z}'\boldsymbol{\gamma})$ . The expectation conditional on the treatment status is

$$E(y_j | \mathbf{x}, d) = \exp(\mathbf{x}'\boldsymbol{\beta}_j + \sigma_j^2/2) \frac{\Phi\{(2d-1)(\rho_j\sigma_j + \mathbf{z}'\boldsymbol{\gamma})\}}{\Phi\{(2d-1)(\mathbf{z}'\boldsymbol{\gamma})\}}$$

The unconditional versions of ATT and ATU are estimated by averaging over relevant subsamples:

$$\hat{\mu}_{\text{ATT}} = \frac{\sum_i^N d_i \hat{\mu}_{\text{ATT}}(\mathbf{w}_i)}{\sum_i^N d_i}, \quad \hat{\mu}_{\text{ATU}} = \frac{\sum_i^N (1 - d_i) \hat{\mu}_{\text{ATU}}(\mathbf{w}_i)}{\sum_i^N (1 - d_i)}$$

LATE is the average effect of the treatment for those who are induced to switch their treatment status in response to changes in an instrument variable. Suppose that a value of the  $k$ th variable of  $\mathbf{z}$ ,  $z_k$  changes from the lower value  $z_{k,L}$  to the upper value  $z_{k,U}$ . Accordingly, the treatment status changes;  $d = 0$  with  $\mathbf{z}_L'\boldsymbol{\gamma}$  while  $d = 1$  with  $\mathbf{z}_U'\boldsymbol{\gamma}$ . Then, LATE can be defined as  $\mu_{\text{LATE}}(\mathbf{w}, z_k) = E(y_1 - y_0 | x, -\mathbf{z}_U'\boldsymbol{\gamma} \leq \nu \leq -\mathbf{z}_L'\boldsymbol{\gamma})$ . The relevant conditional expectation of  $y_j$  is

$$E(y_j | x, -\mathbf{z}_U'\boldsymbol{\gamma} \leq \nu \leq -\mathbf{z}_L'\boldsymbol{\gamma}) = \exp(\mathbf{x}'\boldsymbol{\beta}_j + \sigma_j^2/2) \frac{\Phi(\rho_j\sigma_j + \mathbf{z}_U'\boldsymbol{\gamma}) - \Phi(\rho_j\sigma_j + \mathbf{z}_L'\boldsymbol{\gamma})}{\Phi(\mathbf{z}_U'\boldsymbol{\gamma}) - \Phi(\mathbf{z}_L'\boldsymbol{\gamma})}$$

Note that we cannot identify exactly which observations are induced to change the treatment status in response to the change in an instrument variable. Thus, it is

standard to estimate the unconditional LATE as the average of the whole sample:  $\hat{\mu}_{\text{LATE}}(z_k) = N^{-1} \sum_{i=1}^N \hat{\mu}_{\text{LATE}}(\mathbf{w}_i, z_k)$ . When there are multiple instrument variables, LATE can be defined for each instrument.

Finally, MTE measures the average effect of the treatment for those who have a particular value of  $\nu$ . MTE at  $\nu = \tilde{\nu}$  is  $\mu_{\text{MTE}}(\mathbf{x}, \tilde{\nu}) = E(y_1 - y_0 | \mathbf{x}, \nu = \tilde{\nu})$ . MTE evaluated at low values of  $\tilde{\nu}$  measures the treatment effects for those less likely to receive the treatment, while MTE at high values of  $\tilde{\nu}$  measures the treatment effects for those more likely to receive the treatment. The expected potential outcome conditional on  $\nu = \tilde{\nu}$  is

$$E(y_j | \mathbf{x}, \nu = \tilde{\nu}) = \exp \{ \mathbf{x}' \boldsymbol{\beta}_j + \rho_j \sigma_j \tilde{\nu} + \sigma_j^2 (1 - \rho_j^2) / 2 \}$$

Like ATE and LATE, the unconditional version of MTE is estimated by averaging the whole sample:  $\hat{\mu}_{\text{MTE}}(\tilde{\nu}) = N^{-1} \sum_{i=1}^N \hat{\mu}_{\text{MTE}}(\mathbf{x}_i, \tilde{\nu})$ .

MTE is a unifying parameter in policy evaluation. ATE, ATT, and ATU can be expressed as weighted averages of MTE (Heckman and Vytlacil 2005). ATE is a simple average of MTE over the entire range of  $\nu$ . ATT puts more weight on MTE at larger values of  $\nu$ , that is, those more likely to select into the treatment. In contrast, ATU more heavily weighs those less likely to receive the treatment. MTE can also be interpreted as the limiting case of LATE as  $\mathbf{z}'_U \boldsymbol{\gamma} \rightarrow \mathbf{z}'_L \boldsymbol{\gamma}$ . Thus, it measures the ATE for those who are indifferent between receiving treatment and not at a given value of  $\nu$ . By estimating MTE over a range of  $\nu$ , we can understand how treatment effects are heterogeneous, and we can also tell who is more likely to benefit from marginal policy expansion. For example, constant MTE over  $\nu$  indicates homogeneous treatment effects across different populations. Treatment effects can be heterogeneous. For example, MTE may be large among those unlikely to receive treatment. In this case, a policy maker may design an intervention targeting such a population to make the intervention more effective. Furthermore, MTE can be used to construct not only conventional treatment effects such as ATE and ATT but also treatment effects of hypothetical policy changes with different weights from ATE and ATT. See, for example, Heckman and Vytlacil (2005, 2007) for further discussion.

Note that the treatment-effect estimators in this article heavily rely on the assumption of joint normality discussed in section 2. Such parametric assumptions are often considered as restrictive in the recent literature of program evaluation, and the literature has moved toward less stringent assumptions in modeling and estimating causal effects. See, for example, Imbens and Wooldridge (2009). Indeed, several semiparametric estimators for MTE have been proposed. The package `mtefe` by Andresen (2018) implements several estimation methods for MTE. However, these estimators are not developed specifically for count-data outcome. Our treatment-effect estimators can serve as a benchmark estimation result in empirical studies with count-data outcome, although it is necessary to keep rather stringent distributional assumptions in mind.

As for the derivation of the asymptotic variance of these treatment-effect estimators, note that the estimation of the treatment effect involves two steps sequentially. The first step estimates the parameters of the switching regression model. Then, in the second step, the treatment effects are estimated using these estimated parameters. The

asymptotic variance of the conditional treatment effects can be obtained by applying the delta method, which accounts for the fact that the parameters of the switching model are fit. To consistently estimate the asymptotic variances of the unconditional treatment effects, we must additionally consider the sampling variability from the randomness of  $\mathbf{w}_i$ . Our derivation of the asymptotic variance is based on Newey and McFadden (1994). See also Terza (2016) for the correct asymptotic variance of sample means of nonlinear transformations such as our proposed treatment-effect estimators.

Assuming independence over observations, we estimate the asymptotic variance of the treatment-effect estimator by

$$V(\hat{\mu}_E) = N^{-1} \sum_{i=1}^N \{\hat{\mu}_E(\mathbf{w}_i) - \hat{\mu}_E\}^2 \\ + \left\{ N^{-1} \sum_{i=1}^N \frac{\partial \hat{\mu}_E(\mathbf{w}_i)}{\partial \boldsymbol{\theta}} \right\}' \hat{V}(\hat{\boldsymbol{\theta}}) \left\{ N^{-1} \sum_{i=1}^N \frac{\partial \hat{\mu}_E(\mathbf{w}_i)}{\partial \boldsymbol{\theta}} \right\}$$

where  $\hat{V}(\hat{\boldsymbol{\theta}})$  is the estimated asymptotic variance of  $\hat{\boldsymbol{\theta}}$  for  $E = \text{ATE, LATE, and MTE}$ . The expressions for ATT and ATU are slightly different because we average over subsamples, but the essential structure is the same. The second term of this expression corresponds to the delta-method approach, and the first term reflects the variability arising from the randomness of  $\mathbf{w}_i$ . Furthermore, one can estimate a cluster-robust variance estimator. The first term of the expression above becomes  $N^{-1} \sum_{g=1}^G [\sum_{i=1}^{N_g} \{\hat{\mu}_E(\mathbf{w}_i) - \hat{\mu}_E\}]^2$ , where  $G$  denotes the number of clusters and  $N_g$  is the number of observations with each cluster such that  $N = \sum_{g=1}^G N_g$ .  $\hat{V}(\hat{\boldsymbol{\theta}})$  is replaced with the cluster-robust variance of  $\boldsymbol{\theta}$  in the second term.

## 4 The commands

This section describes the commands `escount` and `lncount` to implement the MLE of the models with lognormal latent heterogeneity. These commands use the Stata command `ml`, and likelihood-evaluator programs are written in Mata. We also describe the command `teescount`, which estimates the treatment effects after the execution of `escount`.

## 4.1 **escount**

### Syntax

The syntax for the command is as follows:

```
escount depvar [=] indepvars [if] [in] [weight],
      select(depvars = varlists [, noconstant offset(varnameo)]) [model(model)
      noconstant exposure(varnamee) offset(varnameo) constraints(constraints)
      intpoints(#) vce(vcetype) maximize_options]
```

**aweight**s, **fweight**s, **iweight**s, and **pweight**s can be used depending on the methods chosen; see [U] **11.1.6 weight**.

### Options

**select**(*depvars* = *varlist<sub>s</sub>* [, **noconstant** **offset**(*varname<sub>o</sub>*)]) specifies the variables and options for the selection equation. **select**() is required.

**model**(*model*) specifies the distribution for count-data outcomes. The default is **model**(**poisson**). The alternative specification is **nb**.

**noconstant** suppresses a constant term of the outcome equation.

**exposure**(*varname<sub>e</sub>*) includes  $\ln(\text{varname}_e)$  in the model with the coefficient constrained to 1.

**offset**(*varname<sub>o</sub>*) includes *varname<sub>o</sub>* in the model with the coefficient constrained to 1.

**constraints**(*constraints*); see [R] **Estimation options**.

**intpoints**(#) specifies the number of integration points to use for integration by quadrature. The default is **intpoints**(24); the maximum is **intpoints**(512).

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

*maximize\_options*: **difficult**, **technique**(*algorithm\_spec*), **iterate**(#), [**no**] **log**, **trace**, **gradient**, **showstep**, **hessian**, **showtolerance**, **tolerance**(#), **ltolerance**(#), **nrtolerance**(#), **nonnrtolerance**, and **from**(*init\_specs*); see [R] **Maximize**. These options are seldom used.



## Stored results

**escount** stores the following in **e()**:

### Scalars

<b>e(N)</b>	number of observations	<b>e(ll)</b>	log likelihood
<b>e(k)</b>	number of parameters	<b>e(p)</b>	significance of model test
<b>e(k.eq)</b>	number of equations in <b>e(b)</b>	<b>e(rank)</b>	rank of <b>e(V)</b>
<b>e(k.eq_model)</b>	number of equations in overall model test	<b>e(chi2)</b>	$\chi^2$
<b>e(k_aux)</b>	number of auxiliary parameters	<b>e(ic)</b>	number of iterations
<b>e(k_dv)</b>	number of dependent variables	<b>e(rc)</b>	return code
<b>e(df_m)</b>	model degrees of freedom	<b>e(converged)</b>	1 if converged, 0 otherwise
		<b>e(n_quad)</b>	number of quadrature points

### Macros

<b>e(cmd)</b>	<b>escount</b>	<b>e(opt)</b>	type of optimization
<b>e(depvar)</b>	name of dependent variable	<b>e(which)</b>	<b>max</b> or <b>min</b> ; whether optimizer is to perform maximization or minimization
<b>e(wtype)</b>	weight type		
<b>e(wexp)</b>	weight expression		
<b>e(title)</b>	title in estimation output	<b>e(ml_method)</b>	type of <b>ml</b> method
<b>e(clustvar)</b>	name of cluster variable	<b>e(user)</b>	name of likelihood-evaluator program
<b>e(offset1)</b>	offset for outcome equation	<b>e(technique)</b>	maximization technique
<b>e(offset2)</b>	offset for selection equation	<b>e(crittype)</b>	optimization criterion
<b>e(chi2type)</b>	Wald or LR; type of model $\chi^2$ test	<b>e(properties)</b>	<b>b V</b>
<b>e(vce)</b>	<i>vcetype</i> specified in <b>vce()</b>	<b>e(predict)</b>	program used to implement <b>predict</b>
<b>e(vcetype)</b>	title used to label Std. Err.		

### Matrices

<b>e(b)</b>	coefficient vector	<b>e(gradient)</b>	gradient vector
<b>e(ilog)</b>	iteration log (up to 20 iterations)	<b>e(V)</b>	variance-covariance matrix of the estimators

### Functions

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

## Prediction

After an execution of **escount**, the **predict** command is available to compute several statistics with the following syntax:

```
predict [type] newvar [if] [in] [, psel xb0 xb1 xbsel mu0 mu1 mu0d0
      mu0d1 mu1d0 mu1d1]
```

The options for **predict** are the following:

**psel**, the default, calculates the probability of receiving the treatment status  $d = 1$  for each observation:  $\Phi(\mathbf{z}_i'\boldsymbol{\gamma})$ .

**xb0** calculates the linear prediction of the outcome variable of the treatment status  $d = 0$  for each observation:  $\mathbf{x}_i'\boldsymbol{\beta}_0$ .

**xb1** calculates the linear prediction of the outcome variable of the treatment status  $d = 1$  for each observation:  $\mathbf{x}_i'\boldsymbol{\beta}_1$ .

**xbse1** calculates the linear prediction of the selection equation for each observation:  
 $\mathbf{z}_i' \boldsymbol{\gamma}$ .

**mu0** calculates the expected value of the potential-outcome variable of the treatment status  $d = 0$  for each observation:  $E(y_0|\mathbf{x}_i) = \exp(\mathbf{x}_i' \boldsymbol{\beta}_0 + \sigma_0^2/2)$ .

**mu1** calculates the expected value of the potential-outcome variable of the treatment status  $d = 1$  for each observation:  $E(y_1|\mathbf{x}_i) = \exp(\mathbf{x}_i' \boldsymbol{\beta}_1 + \sigma_1^2/2)$ .

**mu0d0** calculates the expected value of the potential-outcome variable of the treatment status  $d = 0$  conditional on receiving the treatment status  $d = 0$  for each observation:  
 $E(y_0|\mathbf{x}_i, d = 0) = \exp(\mathbf{x}_i' \boldsymbol{\beta}_0 + \sigma_0^2/2) \{ \Phi(-\rho_0 \sigma_0 - \mathbf{z}' \boldsymbol{\gamma}) / \Phi(-\mathbf{z}' \boldsymbol{\gamma}) \}$ .

**mu0d1** calculates the expected value of the potential-outcome variable of the treatment status  $d = 0$  conditional on receiving the treatment status  $d = 1$  for each observation:  
 $E(y_0|\mathbf{x}_i, d = 1) = \exp(\mathbf{x}_i' \boldsymbol{\beta}_0 + \sigma_0^2/2) \{ \Phi(\rho_0 \sigma_0 + \mathbf{z}' \boldsymbol{\gamma}) / \Phi(\mathbf{z}' \boldsymbol{\gamma}) \}$ .

**mu1d0** calculates the expected value of the potential-outcome variable of the treatment status  $d = 1$  conditional on receiving the treatment status  $d = 0$  for each observation:  
 $E(y_1|\mathbf{x}_i, d = 0) = \exp(\mathbf{x}_i' \boldsymbol{\beta}_1 + \sigma_1^2/2) \{ \Phi(-\rho_1 \sigma_1 - \mathbf{z}' \boldsymbol{\gamma}) / \Phi(-\mathbf{z}' \boldsymbol{\gamma}) \}$ .

**mu1d1** calculates the expected value of the potential-outcome variable of the treatment status  $d = 1$  conditional on receiving the treatment status  $d = 1$  for each observation:  
 $E(y_1|\mathbf{x}_i, d = 1) = \exp(\mathbf{x}_i' \boldsymbol{\beta}_1 + \sigma_1^2/2) \{ \Phi(\rho_1 \sigma_1 + \mathbf{z}' \boldsymbol{\gamma}) / \Phi(\mathbf{z}' \boldsymbol{\gamma}) \}$ .

## 4.2 teescount

### Syntax

The syntax for the command is as follows:

```
teescount [ , ate att atu late(varname lower upper) mte(nu) ]
```

This command estimates the treatment effects and their standard errors using the estimated parameters of the switching regression model with count-data outcomes. It can be executed only after the execution of **escount**.

### Options

**ate** estimates ATE. This is the default.

**att** estimates ATT.

**atu** estimates ATU.

**late(*varname lower upper*)** estimates LATE when the value of *varname* ( $z_k$ ) changes from *lower* ( $z_{k,L}$ ) to *upper* ( $z_{k,U}$ ). *lower* and *upper* should be numerical.

**mte(*nu*)** estimates MTE evaluated at *nu* ( $\nu$ ). *nu* should be numerical.

## Stored results

`teescount` stores the following in `r()`:

Matrices  
`r(table)`      table of result

The matrix `r(table)` contains the estimated treatment effect in the first column, its standard error in the second column, the  $z$ -value in the third column, and the  $p$ -value in the fourth column. When only one option is specified, this matrix is a row vector. When more than one option is specified, the rows of this matrix correspond to the treatment effects that are specified as options.

## 4.3 Incount

### Syntax

The syntax for the command is as follows:

```
lncount devar [=] indepvars [if] [in] [weight] [, model(model) noconstant
exposure(varnamee) offset(varnameo) constraints(constraints)
intpoints(#) vce(vcetype) maximize_options]
```

`aweight`s, `fweight`s, `iweight`s, and `pweight`s can be used depending on the methods chosen; see [U] **11.1.6 weight**.

### Options

`model(model)` specifies the distribution for count-data outcomes. The default is `model(poisson)`. The alternative specification is `nb`.

`noconstant` suppresses a constant term of the outcome equation.

`exposure(varnamee)` includes  $\ln(\text{varname}_e)$  in the model with the coefficient constrained to 1.

`offset(varnameo)` includes  $\text{varname}_o$  in the model with the coefficient constrained to 1.

`constraints(constraints)`; see [R] **Estimation options**.

`intpoints(#)` specifies the number of integration points to use for integration by quadrature. The default is `intpoints(24)`; the maximum is `intpoints(512)`.

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

*maximize\_options*: **difficult**, **technique**(*algorithm\_spec*), **iterate**(#), [**no**] **log**, **trace**, **gradient**, **showstep**, **hessian**, **showtolerance**, **tolerance**(#), **ltolerance**(#), **nrtolerance**(#), **nonnrtolerance**, and **from**(*init\_specs*); see [R] **Maximize**. These options are seldom used.

## Stored results

**lncount** stores the following in **e()**:

### Scalars

<b>e(N)</b>	number of observations	<b>e(ll)</b>	log likelihood
<b>e(k)</b>	number of parameters	<b>e(chi2)</b>	$\chi^2$
<b>e(k.eq)</b>	number of equations in <b>e(b)</b>	<b>e(rank)</b>	rank of <b>e(V)</b>
<b>e(k.eq_model)</b>	number of equations in overall model test	<b>e(ic)</b>	number of iterations
<b>e(k.aux)</b>	number of auxiliary parameters	<b>e(rc)</b>	return code
<b>e(k.dv)</b>	number of dependent variables	<b>e(converged)</b>	1 if converged, 0 otherwise
<b>e(df_m)</b>	model degrees of freedom	<b>e(p)</b>	significance of model test
		<b>e(n.quad)</b>	number of quadrature points

### Macros

<b>e(cmd)</b>	<b>lncount</b>	<b>e(which)</b>	<b>max</b> or <b>min</b> ; whether optimizer is to perform maximization or minimization
<b>e(depvar)</b>	name of dependent variable	<b>e(ml_method)</b>	type of ml method
<b>e(wtype)</b>	weight type	<b>e(user)</b>	name of likelihood-evaluator program
<b>e(wexp)</b>	weight expression	<b>e(technique)</b>	maximization technique
<b>e(title)</b>	title in estimation	<b>e(crittype)</b>	optimization criterion
<b>e(clustvar)</b>	name of cluster variable	<b>e(properties)</b>	<b>b</b> <b>V</b>
<b>e(offset)</b>	offset for outcome equation	<b>e(predict)</b>	program used to implement <b>predict</b>
<b>e(chi2type)</b>	Wald or LR; type of model $\chi^2$ test		
<b>e(vce)</b>	<i>vcetype</i> specified in <b>vce()</b>		
<b>e(vcetype)</b>	title used to label Std. Err.		
<b>e(opt)</b>	type of optimization		

### Matrices

<b>e(b)</b>	coefficient vector	<b>e(gradient)</b>	gradient vector
<b>e(ilog)</b>	iteration log (up to 20 iterations)	<b>e(V)</b>	variance-covariance matrix of the estimators

### Functions

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

## Prediction

After an execution of **lncount**, the **predict** command is available to compute several statistics with the following syntax:

```
predict [type] newvar [if] [in] [, n xb]
```

The options for **predict** are the following:

**n** computes the predicted number of events for each observation:  $\exp(\mathbf{x}_i'\boldsymbol{\beta} + \sigma^2/2)$ . This is the default.

**xb** computes the linear prediction of the dependent variable for each observation:  $\mathbf{x}_i'\boldsymbol{\beta}$ .

## 4.4 Notes

In the MLE,  $\sigma$  and  $\rho$  are not directly estimated to guarantee that  $\sigma > 0$  and  $-1 < \rho < 1$ . Instead, ancillary parameters  $\ln \sigma$  and  $\rho^*$  are estimated. These ancillary parameters are transformed as  $\sigma_j = \exp(\ln \sigma_j)$  and  $\rho_j = \{\exp(\rho_j^*) - \exp(-\rho_j^*)\} / \{\exp(\rho_j^*) + \exp(-\rho_j^*)\}$ . The estimates of these ancillary parameters are respectively reported as **lnsigma** and **athrho**, while the estimates of these transformed parameters are reported as **sigma** and **rho**. For the switching model, the estimates of the ancillary and transformed parameters are reported with 0 or 1 indicating the treatment status. Similarly, when **model** is **nb**, the ancillary parameter  $\ln \alpha$  is estimated and then transformed to  $\alpha$ . The estimates of these parameters are reported as **lnalpha** and **alpha**, respectively.

## 5 Example

To illustrate the use of **escount**, we show some examples using a dataset from the Stata website. This is the same dataset used for the example of **etpoisson** and a simulated random sample of households. The outcome of interest is the number of trips (**trips**), and a possible endogenous treatment indicator is car ownership (**owncar**). There are some other controlling variables of household characteristics. **realinc**, which is the ratio of household income to the median income of the census tract, serves as an instrumental variable. See the example in [TE] **etpoisson** for the details.

First, we execute **escount** to fit the switching regression model with the Poisson specification. It is followed by the command that tests whether the self-selection matters, that is, tests the null hypothesis  $H_0 : \rho_0 = \rho_1 = 0$ . It is equivalent to testing  $H_0 : \rho_0^* = \rho_1^* = 0$ .<sup>1</sup>

---

1. Technically speaking, the Wald test is not invariant to the transformation. However, it does not matter so much in practice.

```

. webuse tripl
(Household trips, car ownership)
. escount trips cbd ptn worker weekend, select(owncar = cbd ptn worker realinc)
> vce(robust)
(output omitted)
Iteration 5:   log pseudolikelihood = -12409.546
endogenously switching poisson regression           Number of obs   =       5,000
                                                    Wald chi2(4)    =       61.18
Log pseudolikelihood = -12409.546                 Prob > chi2     =       0.0000

```

	Coef.	Robust Std. Err.	z	P> z	[95% Conf. Interval]	
<b>status0</b>						
cbd	-.0058734	.0045937	-1.28	0.201	-.0148768	.00313
ptn	-.0222624	.0048025	-4.64	0.000	-.0316751	-.0128497
worker	.6316955	.1184363	5.33	0.000	.3995647	.8638263
weekend	-.0141076	.0781904	-0.18	0.857	-.1673581	.1391428
_cons	-.1001897	.1386091	-0.72	0.470	-.3718585	.1714791
<b>status1</b>						
cbd	-.0108746	.0022049	-4.93	0.000	-.0151961	-.0065532
ptn	-.0198676	.0022123	-8.98	0.000	-.0242036	-.0155316
worker	.736345	.0655705	11.23	0.000	.6078292	.8648609
weekend	.1187595	.0384917	3.09	0.002	.0433172	.1942019
_cons	.2508071	.0889429	2.82	0.005	.0764822	.425132
<b>selection</b>						
cbd	.0073465	.0023976	3.06	0.002	.0026473	.0120458
ptn	.0083905	.0024525	3.42	0.001	.0035837	.0131973
worker	.5422459	.0506216	10.71	0.000	.4430295	.6414624
realinc	.1759857	.0110051	15.99	0.000	.1544161	.1975553
_cons	-.4597375	.0594462	-7.73	0.000	-.57625	-.3432251
/lnsigma0	-.1432683	.0941404	-1.52	0.128	-.3277801	.0412436
/lnsigma1	-.2248189	.0274594	-8.19	0.000	-.2786383	-.1709995
/athrho0	.6656744	.2175962	3.06	0.002	.2391937	1.092155
/athrho1	.5711723	.1225211	4.66	0.000	.3310354	.8113093
sigma0	.8665216	.0815747			.7205214	1.042106
sigma1	.7986608	.0219307			.7568136	.842822
rho0	.5821273	.1438589			.234734	.7976634
rho1	.5162197	.0898713			.3194508	.6703119

```

. test ([/athrho0]=0) ([/athrho1]=0)
( 1)  [/athrho0 = 0
( 2)  [/athrho1 = 0
      chi2( 2) =    30.29
      Prob > chi2 =    0.0000

```

This is the result of MLE of the switching regression model with count-data outcomes. Each of the coefficients of correlation between  $\nu$  and  $\varepsilon_j$  for  $j = 0, 1$  is statistically significant at any meaningful level. The result of the `test` command also shows that these parameters are jointly significant. These results indicate the presence of the self-selection problem. Households that tend to own a car are more likely to make a trip,

even controlling for observable characteristics. Without considering the self-selection, the estimation of the treatment effects is biased.

Next, we estimate the treatment effects: ATE, ATT, and ATU. We also report the estimated LATE when the value of `realinc` changes from 1.0 to 1.1, which may not necessarily be an interesting quantity but is shown for illustration. Besides, we also show the estimated treatment effects ignoring the self-selection problem, that is, restricting  $\rho_0 = \rho_1 = 0$ .

```
. teescount, ate att atu late(realinc 1.0 1.1)
Treatment Effects:
      Estimate      SE      z      P>|z|
ATE      .86494    .43936    1.9686    .049
ATT      .97261    .6165     1.5776    .11465
ATU      .61178    .20322    3.0105    .00261
LATE      .7802     .22721    3.4338    .0006

. constraint 1 [/athrho0] = 0
. constraint 2 [/athrho1] = 0
. escount trips cbd ptn worker weekend, select(owncar = cbd ptn worker realinc)
> vce(robust) constraints(1/2)
(output omitted)

. teescount, ate att atu late(realinc 1.0 1.1)
Treatment Effects:
      Estimate      SE      z      P>|z|
ATE      2.119     .06487    32.667      0
ATT      2.155     .06913    31.173      0
ATU      2.0347    .07966    25.543      0
LATE      2.119     .06487    32.667      0
```

The estimated treatment effects from the model without considering the self-selection problem are more than twice as large as those from the switching regression model. This result is consistent with the positive values of the estimated  $\rho_0$  and  $\rho_1$ . Ignoring the self-selection problem results in the upward biases of the treatment effects. Note that when  $\rho_0 = 0$  and  $\rho_1 = 0$ , the treatment effects are not heterogeneous so that ATE and LATE are the same. Although not reported here, MTE is also constant at the same value as ATE and LATE over the entire range of  $\nu$ . The reason why ATT and ATU are different is that we average over different subsamples to obtain the unconditional effects.

When considering the self-selection issue, we find that ATE is marginally significant at the 5% level. Although ATU is highly significant, ATT is not significant even at the 10% level. The insignificance of ATT may be due to the efficiency loss from the estimation of the switching regression model. An endogenous dummy variable model may be parsimonious and lead to a more efficient estimator when the restrictions that  $\sigma_0 = \sigma_1$ ,  $\rho_0 = \rho_1$ , and  $\beta_0 = \beta_1$ , except for constant terms being correct. As shown below, when we conduct the Wald test of the restrictions, we fail to reject the null hypothesis with the  $p$ -value of 0.3472.

```
. test ([status0 = status1]) ([/athrho0] = [/athrho1]) ([/lnsigma0] =
> [/lnsigma1])
( 1)  [status0]cbd - [status1]cbd = 0
( 2)  [status0]ptn - [status1]ptn = 0
( 3)  [status0]worker - [status1]worker = 0
( 4)  [status0]weekend - [status1]weekend = 0
( 5)  [/]athrho0 - [/]athrho1 = 0
( 6)  [/]lnsigma0 - [/]lnsigma1 = 0
      chi2( 6) =    6.72
      Prob > chi2 =    0.3472
```

Given this result, we next fit the model and the treatment effects under the restrictions. For comparison and as an example of `lncount`, we also fit the dummy variable model without considering the self-selection issue.

```
. lncount trips owncar cbd ptn worker weekend, vce(robust)
(output omitted)
Iteration 4:   log pseudolikelihood = -9641.2706
log-normal poisson regression               Number of obs   =      5,000
                                              Wald chi2(5)     =     1244.32
Log pseudolikelihood = -9641.2706          Prob > chi2     =      0.0000
```

trips	Coef.	Robust Std. Err.	z	P> z	[95% Conf. Interval]	
owncar	1.199065	.0400493	29.94	0.000	1.120569	1.27756
cbd	-.0114598	.0019539	-5.86	0.000	-.0152894	-.0076301
ptn	-.0220301	.0019578	-11.25	0.000	-.0258672	-.0181929
worker	.5456353	.0494759	11.03	0.000	.4486643	.6426063
weekend	.0962581	.0341605	2.82	0.005	.0293047	.1632115
_cons	-.5587373	.0608177	-9.19	0.000	-.6779378	-.4395368
/lnsigma	-.2756947	.0203599	-13.54	0.000	-.3155994	-.2357899
sigma	.7590447	.0154541			.7293516	.7899466

```
. constraint 3 [/athrho0] = [/athrho1]
. constraint 4 [/lnsigma0] = [/lnsigma1]
. constraint 5 [status0 = status1]
```



```
. escount trips cbd ptn worker weekend, select(owncar = cbd ptn worker realinc)
> vce(robust) constraints(3/5)
(output omitted)
```

```
Iteration 4: log pseudolikelihood = -12412.696
```

```
endogenously switching poisson regression      Number of obs      =      5,000
                                                Wald chi2(0)        =      .
Log pseudolikelihood = -12412.696             Prob > chi2         =      .
```

```
( 1)  [/]athrho0 - [/]athrho1 = 0
( 2)  [/]lnsigma0 - [/]lnsigma1 = 0
( 3)  [status0]cbd - [status1]cbd = 0
( 4)  [status0]ptn - [status1]ptn = 0
( 5)  [status0]worker - [status1]worker = 0
( 6)  [status0]weekend - [status1]weekend = 0
```

	Coef.	Robust Std. Err.	z	P> z	[95% Conf. Interval]	
status0						
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
_cons	-.2340771	.0810812	-2.89	0.004	-.3929933	-.0751609
status1						
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
_cons	.2923941	.0766251	3.82	0.000	.1422117	.4425766
selection						
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
/lnsigma0	-.2182037	.0256281	-8.51	0.000	-.2684338	-.1679735
/lnsigma1	-.2182037	.0256281	-8.51	0.000	-.2684338	-.1679735
/athrho0	.5741169	.0957832	5.99	0.000	.3863852	.7618486
/athrho1	.5741169	.0957832	5.99	0.000	.3863852	.7618486
sigma0	.8039617	.020604			.764576	.8453762
sigma1	.8039617	.020604			.764576	.8453762
rho0	.5183764	.0700449			.3682398	.6421646
rho1	.5183764	.0700449			.3682398	.6421646

```
. lincom [status1]_cons - [status0]_cons
```

```
( 1) - [status0]_cons + [status1]_cons = 0
```

	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
(1)	.5264713	.1124157	4.68	0.000	.3061406	.746802

```
. teescount, ate att atu late(realinc 1.0 1.1)
Treatment Effects:
      Estimate      SE      z      P>|z|
ATE      1.0589    .19233    5.5056    3.7e-08
ATT       1.252    .20636    6.0668    1.3e-09
ATU       .60765   .16267    3.7355    .00019
LATE      .83589   .19122    4.3713    1.2e-05
```

The estimation result of the restricted model, that is, the endogenous dummy variable model, is identical to the result from the built-in command `etpoisson`, which is not reported here. The coefficient on the endogenous dummy variable corresponds to the difference in the constant terms between the treatment statuses. The command `lincom` returns the estimated difference. This estimate is much lower than the estimated coefficient on `owncar` from `lncount`, which ignores the endogeneity of car ownership. The estimates of the treatment effects based on the endogenous dummy variable model are similar to those from the switching regression model, but with smaller standard errors.

Based on the endogenous dummy variable model, we estimate and plot MTEs over normalized values of  $\nu$  with 95% confidence intervals. The normalized value of  $\nu$  is  $\Phi(\nu)$ , and thus it takes values between 0 and 1. It represents the propensity for treatment. MTE and its standard error are estimated at each point of the normalized value using a loop.

```
. capture generate P = .
. quietly generate MTE = .
. quietly generate MTEse = .
. forvalue j = 1/99 {
2.     local p = `j'/100
3.     local v = invnormal(`p')
4.     quietly teescount, mte(`v')
5.     quietly matrix b = r(table)
6.     quietly replace MTE = b[1,1] if _n == `j'
7.     quietly replace MTEse = b[1,2] if _n == `j'
8.     quietly replace P = `p' if _n==`j'
9. }
```

Figure 1 presents the estimated MTEs. The figure shows that MTE increases with higher propensity for treatment. The effect of car ownership on the number of trips is larger for a household that is more likely to own a car. The result, that ATT is larger than ATU, reflects this upward slope of MTE. At higher values of treatment propensity, where ATT puts more weight on MTE, the treatment effect is larger.

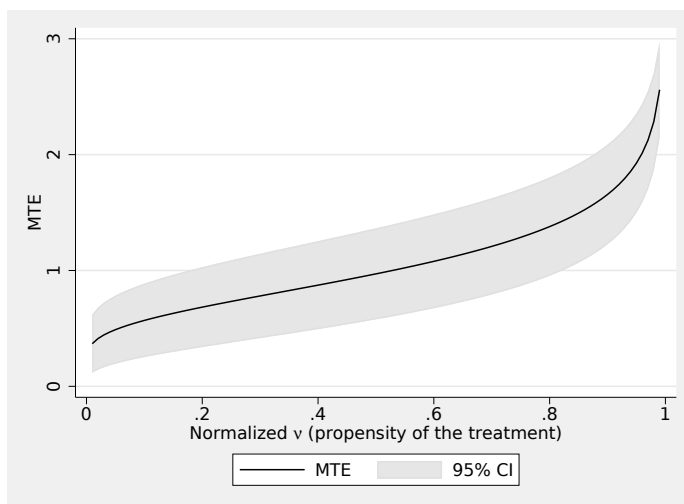


Figure 1. MTEs

## 6 Conclusion

In estimating the effects of treatment, endogeneity is a common problem that researchers face. Researchers in health economics and other applied microeconomics also come across count-data outcomes frequently. It is important to equip researchers with a statistical tool to address endogenous treatment with a count-data outcome. In this article, I introduced the command `escount`, which implements the MLE of the endogenously switching regression model with count-data outcomes, where a possible outcome differs between different treatment statuses and an individual selects into one of the statuses endogenously. Based on lognormal latent heterogeneity, the model allows for the dependence between the selection process and potential outcomes. After one estimates the parameters of the switching regression model, various treatment effects can be estimated with the command `teescount`.

## 7 Acknowledgments

I thank an anonymous referee for helpful suggestions. This study was partially funded by Grant-in-Aid for Young Scientists from the Japan Society for the Promotion of Science: JSPS KAKENHI grant number JP18K12801.

## 8 Programs and supplemental materials

To install a snapshot of the corresponding software files as they existed at the time of publication of this article, type

```
. net sj 20-3
. net install st0612      (to install program files, if available)
. net get st0612          (to install ancillary files, if available)
```

## 9 References

- Andresen, M. E. 2018. Exploring marginal treatment effects: Flexible estimation using Stata. *Stata Journal* 18: 118–158. <https://doi.org/10.1177/1536867X1801800108>.
- Greene, W. 2009. Models for count data with endogenous participation. *Empirical Economics* 36: 133–173. <https://doi.org/10.1007/s00181-008-0190-5>.
- Hasebe, T. 2018. Treatment effect estimators for count data models. *Health Economics* 27: 1868–1873. <https://doi.org/10.1002/hec.3790>.
- Heckman, J. J. 2010. Building bridges between structural and program evaluation approaches to evaluating policy. *Journal of Economic Literature* 48: 356–398. <https://doi.org/10.1257/jel.48.2.356>.
- Heckman, J. J., J. L. Tobias, and E. Vytlacil. 2003. Simple estimators for treatment parameters in a latent-variable framework. *Review of Economics and Statistics* 85: 748–755. <https://doi.org/10.1162/003465303322369867>.
- Heckman, J. J., and E. J. Vytlacil. 2005. Structural equations, treatment effects, and econometric policy evaluation. *Econometrica* 73: 669–738. <https://doi.org/10.1111/j.1468-0262.2005.00594.x>.
- . 2007. Econometric evaluation of social programs, part II: Using the marginal treatment effect to organize alternative econometric estimators to evaluate social programs, and to forecast their effects in new environments. In *Handbook of Econometrics*, vol. 6B, ed. J. J. Heckman and E. E. Leamer, 4875–5143. Amsterdam: Elsevier. [https://doi.org/10.1016/S1573-4412\(07\)06071-0](https://doi.org/10.1016/S1573-4412(07)06071-0).
- Imbens, G. W., and J. M. Wooldridge. 2009. Recent developments in the econometrics of program evaluation. *Journal of Economic Literature* 47: 5–86. <https://doi.org/10.1257/jel.47.1.5>.
- Newey, W. K., and D. McFadden. 1994. Large sample estimation and hypothesis testing. In *Handbook of Econometrics*, vol. 4, ed. R. F. Engle and D. L. McFadden, 2111–2245. Amsterdam: Elsevier. [https://doi.org/10.1016/S1573-4412\(05\)80005-4](https://doi.org/10.1016/S1573-4412(05)80005-4).
- Terza, J. V. 1998. Estimating count data models with endogenous switching: Sample selection and endogenous treatment effects. *Journal of Econometrics* 84: 129–154. [https://doi.org/10.1016/S0304-4076\(97\)00082-1](https://doi.org/10.1016/S0304-4076(97)00082-1).

- . 2016. Inference using sample means of parametric nonlinear data transformations. *Health Services Research* 51: 1109–1113. <https://doi.org/10.1111/1475-6773.12494>.

**About the author**

Takuya Hasebe is an associate professor in the Faculty of Liberal Arts at Sophia University, Tokyo, Japan.