The Stata Journal

The *Stata Journal* publishes reviewed papers together with shorter notes or comments, regular columns, book reviews, and other material of interest to Stata users. Examples of the types of papers include 1) expository papers that link the use of Stata commands or programs to associated principles, such as those that will serve as tutorials for users first encountering a new field of statistics or a major new technique; 2) papers that go "beyond the Stata manual" in explaining key features or uses of Stata that are of interest to intermediate or advanced users of Stata; 3) papers that discuss new commands or Stata programs of interest either to a wide spectrum of users (e.g., in data management or graphics) or to some large segment of Stata users (e.g., in survey statistics, survival analysis, panel analysis, or limited dependent variable modeling); 4) papers analyzing the statistical properties of new or existing estimators and tests in Stata; 5) papers that could be of interest or usefulness to researchers, especially in fields that are of practical importance but are not often included in texts or other journals, such as the use of Stata in managing datasets, especially large datasets, with advice from hard-won experience; and 6) papers of interest to those who teach, including Stata with topics such as extended examples of techniques and interpretation of results, simulations of statistical concepts, and overviews of subject areas.

The *Stata Journal* is indexed and abstracted by *CompuMath Citation Index*, *Current Contents/Social and Behavioral Sciences*, *RePEc: Research Papers in Economics*, *Science Citation Index Expanded* (also known as *SciSearch*), *Scopus*, and *Social Sciences Citation Index*.

For more information on the *Stata Journal*, including information for authors, see the webpage

http://www.stata-journal.com

Subscriptions are available from StataCorp, 4905 Lakeway Drive, College Station, Texas 77845, telephone 979-696-4600 or 800-STATA-PC, fax 979-696-4601, or online at

http://www.stata.com/bookstore/sj.html

Subscription rates listed below include both a printed and an electronic copy unless otherwise mentioned.

U.S. and Canada		Elsewhere	
Printed & electronic		**Printed & electronic**	
1-year subscription	$115	1-year subscription	$145
2-year subscription	$210	2-year subscription	$270
3-year subscription	$285	3-year subscription	$375
1-year student subscription	$ 85	1-year student subscription	$115
1-year institutional subscription	$345	1-year institutional subscription	$375
2-year institutional subscription	$625	2-year institutional subscription	$685
3-year institutional subscription	$875	3-year institutional subscription	$965
Electronic only		**Electronic only**	
1-year subscription	$ 85	1-year subscription	$ 85
2-year subscription	$155	2-year subscription	$155
3-year subscription	$215	3-year subscription	$215
1-year student subscription	$ 55	1-year student subscription	$ 55

Back issues of the *Stata Journal* may be ordered online at

http://www.stata.com/bookstore/sjj.html

Individual articles three or more years old may be accessed online without charge. More recent articles may be ordered online.

http://www.stata-journal.com/archives.html

The *Stata Journal* is published quarterly by the Stata Press, College Station, Texas, USA.

Address changes should be sent to the *Stata Journal*, StataCorp, 4905 Lakeway Drive, College Station, TX 77845, USA, or emailed to sj@stata.com.

The *Stata Journal* (ISSN 1536-867X) is a publication of Stata Press. Stata, Stata, Stata Press, Mata, Mata, and NetCourse are registered trademarks of StataCorp LP.

The Stata Journal (2015)
15, Number 3, pp. 756–774

Estimating treatment effects for ordered outcomes using maximum simulated likelihood

Christian A. Gregory
Economic Research Service, USDA
Washington, DC
cgregory@ers.usda.gov

Abstract. I present four new commands to estimate the effect of a binary endogenous treatment on an ordered outcome. Such models conventionally rely upon joint normality of the unobservables in treatment and outcome processes, as do `treatoprobit` and `switchoprobit`. In this article, I highlight the capabilities of `treatoprobitsim` and `switchoprobitsim`, which both use a latent-factor structure to model the joint distribution of the treatment and outcome and allow the researcher to relax the assumption of joint normality.

Keywords: st0402, treatoprobit, switchoprobit, treatoprobitsim, switchoprobitsim, ordinal outcomes, endogenous binary treatment, treatment effects

1 Introduction

Ordered outcomes appear frequently in social science; consumer preferences, injury severity, political attitudes, and self-assessed health are among the many outcomes that appear as ordinal in survey and vital statistics data.[1] The specification of these outcomes in a treatment-effects framework relies crucially on the handling of unobservables. If one assumes that they are unimportant to the selection process, a model such as that executed by the command `teffects` could be appropriate. Alternatively, if one believes that unobservable confounders are important to the selection and outcome processes, it is important to include them in the model specification. Conventionally, and most simply, these errors are often assumed to follow a bivariate normal (BIVN) distribution—such would be an application of `gllamm` or `ssm` (Miranda and Rabe-Hesketh 2006). However, when the assumption of joint normality is violated, estimates can be inconsistent.

To address this problem, Aakvik, Heckman, and Vytlacil (2005) proposed using a latent-factor (LF) structure to model unobserved heterogeneity in treatment and outcome processes. In theory, LFs represent unobserved traits that determine both treatment participation and outcome. Thus they are a potential source of confounding in econometric settings. In practice, they can be incorporated into econometric models by numerical integration—as in Aakvik, Heckman, and Vytlacil (2005)—or by simulation. As shown below, in the latter case, distributions of LFs can be approximated by random draws from (virtually any) continuous probability distribution and then entered into the model much like observed covariates. Because this approach allows the researcher to

1. Greene and Hensher (2010) provide an exhaustive treatment of ordered models and their applications.

relax the assumption of bivariate normality, it has been especially useful in cases where outcomes are marginally nonnormal (Deb and Trivedi 2006a,b). The two commands I introduce here allow the same flexibility in situations in which treatment and outcome are marginally normal.

2 Model

In this section, I outline the two types of ordered-outcomes models fit by the commands `treatoprobitsim` and `switchoprobitsim`. The first is a two-equation treatment-effects model for an ordered outcome. The second is a switching regression model for an ordered outcome, where the outcomes for treated and untreated persons are handled separately.

For both models, we represent the treatment in the following way:

$$T_i = \begin{cases} 1 & \text{if } T_i^* = Z_i\gamma + \upsilon_i > 0 \\ 0 & \text{if } T_i^* = Z_i\gamma + \upsilon_i \leq 0 \end{cases}$$

The treatment-effects model assumes that there is one regime for the outcome. In this model, the outcome is

$$Y_i = \begin{cases} 1 & \text{if } -\infty < X_i\beta + \varepsilon_i \leq \mu_1 \\ 2 & \text{if } \mu_1 < X_i\beta + \varepsilon_i \leq \mu_2 \\ & \dots \\ J-1 & \text{if } \mu_{J-1} < X_i\beta + \varepsilon_i \leq \mu_J \\ J & \text{if } \mu_J < X_i\beta + \varepsilon_i \leq \infty \end{cases}$$

for $j = 1, \dots, J$ possible outcomes and where the index $Y_i^* = X_i\beta + \varepsilon_i$. In the endogenous switching model, the outcome is

$$Y_{0i} = \begin{cases} 1 & \text{if } -\infty < X_{0i}\beta_0 + \varepsilon_{0i} \leq \mu_{01} \\ 2 & \text{if } \mu_{01} < X_{0i}\beta_0 + \varepsilon_{0i} \leq \mu_{02} \\ & \dots \\ J-1 & \text{if } \mu_{0J-1} < X_{0i}\beta_0 + \varepsilon_{0i} \leq \mu_{0J} \\ J & \text{if } \mu_{0J} < X_{0i}\beta_0 + \varepsilon_{0i} \leq \infty \end{cases}$$

for the untreated group and

$$Y_{1i} = \begin{cases} 1 & \text{if } -\infty < X_{1i}\beta_1 + \varepsilon_{1i} \leq \mu_{11} \\ 2 & \text{if } \mu_{11} < X_{1i}\beta_1 + \varepsilon_{1i} \leq \mu_{12} \\ & \dots \\ J-1 & \text{if } \mu_{1J-1} < X_{1i}\beta_1 + \varepsilon_{1i} \leq \mu_{1J} \\ J & \text{if } \mu_{1J} < X_{1i}\beta_1 + \varepsilon_{1i} \leq \infty \end{cases}$$

for the treated group for $j = 1, \dots, J$ ordered outcomes. In the endogenous switching model, the latent outcome indices are $Y_{0i}^* = X_{0i}\beta_0 + \varepsilon_{0i}$ and $Y_{1i}^* = X_{1i}\beta_1 + \varepsilon_{1i}$.

2.1 Latent-factor approach

One approach to fitting such a model via maximum likelihood is to assume that υ and ε are distributed as BIVN. As mentioned, working under this assumption when it is not true can yield inconsistent estimates of model parameters. An alternative option is to reformulate the model such that

$$\begin{aligned} \upsilon_i &= \lambda_T \eta_i + \zeta_i \\ \varepsilon_i &= \lambda_Y \eta_i + \iota_i \end{aligned}$$

where we assume that the marginal distributions of ζ and ι are normal but that η need not be. In a maximum likelihood setting, one could integrate out the LFs. So, for the treatment-effects specification, the likelihood function would be

$$\begin{aligned} L_i = \prod_{i=1}^{N} \int_{-\infty}^{\infty} & \Phi\{\tau \times (Z_i\gamma + \lambda_T \eta_i)\} \\ & \times \sum_{k=1}^{K} \{I \times (Y_i = k)\}\{\Phi(\mu_k - X_i\beta + \lambda_Y \eta_i) - \Phi(\mu_{k-1} - X_i\beta + \lambda_Y \eta_i)\} d\eta \end{aligned}$$

where Φ is the standard normal distribution, I is an indicator function, $\tau = 2 \times T_i - 1$, $\mu_0 = -\infty$, $\mu_K = \infty$, and $K = J + 1$. While this can be accomplished by numerical integration, our approach is instead to simulate the distribution of η by taking random draws from its chosen distribution. In this case, the likelihood function for the treatment-effects model is

$$\begin{aligned} L_i = \frac{1}{S} \prod_{i=1}^{N} \sum_{s=1}^{S} & \Phi\{\tau \times (Z_i\gamma + \lambda_T \eta_i)\} \\ & \times \sum_{k=1}^{K} \{I \times (Y = k)\}\{\Phi(\mu_k - X_i\beta + \lambda_Y \eta_i) - \Phi(\mu_{k-1} - X_i\beta + \lambda_Y \eta_i)\} \end{aligned}$$

where S is the number of simulation draws, and the λs are loading factors that describe the dependence between the unobservables in treatment and outcome processes. For the endogenous switching model, the likelihood function is

$$\begin{aligned} L_i = \frac{1}{S} \prod_{i=1}^{N} \sum_{s=1}^{S} \sum_{\ell=0}^{\ell=1} & \{I \times (T_i = \ell)\} \times \Phi\{\tau \times (Z_i\gamma + \lambda_{\ell T} \eta_i)\} \times \sum_{\ell=0}^{\ell=1} \{I \times (T_i = \ell)\} \\ & \times \sum_{k=1}^{K} \{I \times (Y_i = k)\}\{\Phi(\mu_{\ell k} - X_{\ell i}\beta_\ell + \lambda_{\ell Y} \eta_{\ell i}) - \Phi(\mu_{\ell k-1} - X_{\ell i}\beta_\ell + \lambda_{\ell Y} \eta_{\ell i})\} \end{aligned}$$

where $\ell \in (0, 1)$.

In implementing these estimators, we use Halton-based sequences to draw from the distributions of the LFs. The implementation of Halton sequences in Mata is described

in Drukker and Gates (2006). Deb and Trivedi (2006a) and Train (2009) suggest that there are three advantages of using Halton sequences: 1) they cover the domain of the distribution more evenly than traditional pseudorandom generators; 2) they reduce the variance of the simulated likelihood function caused by the negative correlation of draws across observations; and 3) they substantially reduce required computational time.

2.2 Marginal effects

The estimated treatment effects are of particular interest in this context, specifically the average treatment effect (ATE) and the average treatment effect for the treated (ATT). The former is defined as the effect of treatment on a person selected at random from the given population relative to the effect on that person had he or she not received the treatment. Let δ be the coefficient on the endogenous treatment dummy variable in this model. For the treatment-effects specification,

$$\begin{aligned}\text{ATE}_j^T = \frac{1}{N}\frac{1}{S}\sum_{i=1}^{N}\sum_{s=1}^{S}&[\Phi\{\mu_k - (X_i\beta + \delta + \lambda\eta_{is})\} - \Phi\{\mu_{k-1} - (X_i\beta + \delta + \lambda\eta_{is})\}]\\ &- [\Phi\{\mu_k - (X_i\beta + \lambda\eta_{is})\} - \Phi\{\mu_{k-1} - (X_i\beta + \lambda\eta_{is})\}]\end{aligned}$$

and for the switching regression,

$$\begin{aligned}\text{ATE}_k^S = \frac{1}{N}\frac{1}{S}\sum_{i=1}^{N}\sum_{s=1}^{S}&[\Phi\{\mu_{1k} - (X_{1i}\beta_1 + \lambda_1\eta_{is})\} - \Phi\{\mu_{1k-1} - (X_{1i}\beta_1 + \lambda_1\eta_{is})\}]\\ &- [\Phi\{\mu_{0k} - (X_{0i}\beta_0 + \lambda_0\eta_{is})\} - \Phi\{\mu_{0k-1} - (X_{0i}\beta_0 + \lambda_0\eta_{is})\}]\end{aligned}$$

Here $k = 1, \ldots, K$, $K = J + 1$, and J is the number of choices. $\mu_0 = -\infty$ and $\mu_K = \infty$, and Φ is the standard normal cumulative distribution. $T = \ell$ for $\ell \in (0, 1)$ signifies that the treatment indicator has been set to 0 or 1. The T and S superscripts refer to the treatment effects and switching models, respectively.

The ATT estimates the difference in outcomes for a person who adopted the treatment; that is, it tells us, conditional on the treatment, the difference between the treated and untreated state for a given person. It is

$$\begin{aligned}\text{ATT}_j^T = \frac{1}{N}\frac{1}{S}\sum_{i=1}^{N}\frac{1}{E\{\Phi(Z_i\gamma)\}}\Bigg(&\sum_{s=1}^{S}\Phi(Z_i\gamma + \eta_{is})\\ &\times [\Phi\{\mu_j - (X_i\beta + \delta + \lambda\eta_{is})\} - \Phi\{\mu_{j-1} - (X_i\beta + \delta + \lambda\eta_{is})\}\\ &- \Phi\{\mu_j - (X_i\beta + \lambda\eta_{is})\} + \Phi\{\mu_{j-1} - (X_i\beta + \lambda\eta_{is})\}]\Bigg)\end{aligned}$$

for the treatment-effects model. For the switching model, it is

$$\begin{aligned}\text{ATT}_j^S = \frac{1}{N}\frac{1}{S}\sum_{i=1}^{N}\frac{1}{E\{\Phi(Z_i\gamma)\}}\Bigg(&\sum_{s=1}^{S}\sum_{\ell=0}^{\ell=1}\{I\times(T_i=\ell)\}\,\Phi(Z_i\gamma+\eta_{is})\\ &\times[\Phi\{\mu_{1j}-(X_{1i}\beta_1+\lambda_1\eta_{is})\}-\Phi\{\mu_{1,j-1}-(X_{1i}\beta_1+\lambda_1\eta_{is})\}\\ &-\Phi\{\mu_{0j}-(X_{0i}\beta_0+\lambda_0\eta_{is})\}+\Phi\{\mu_{0,j-1}-(X_{0i}\beta_0+\lambda_0\eta_{is})\}]\Bigg)\end{aligned}$$

See Aakvik, Heckman, and Vytlacil (2005) for more detail. Here we normalize the λ_T to one, as is customary in the literature.

3 Maximum simulated likelihood estimation

The commands `treatoprobitsim` and `switchoprobitsim` fit the treatment and switching regression models, respectively. The syntax for both commands is

`treatoprobitsim` | `switchoprobitsim` *depvar* [*indepvars*] [*if*] [*in*] [*weight*] ,
`treatment(`*depvar*$_t$ `=` [*varlist*$_t$]`)` `simulationdraws(`#`)` [`facdensity(`*string*`)`
`facscale(`*real*`)` `facskew(`*real*`)` `startpoint(`*integer*`)` `facmean(`*real*`)`
`vce(`*string*`)` `sesimulations(`*integer*`)` `mixpi(`*integer*`)`]

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

3.1 Estimation options

`treatment(`*depvar*$_t$ `=` [*varlist*$_t$]`)` specifies the participation index (coded as zero or one). `treatment()` is required.

`simulationdraws(`#`)` specifies the number of draws from the distribution of the LF. `simulationdraws()` is required.

`facdensity(`*string*`)` specifies the density of the LF. *string* may be `normal`, `uniform`, `logit`, `chi2`, `lognormal`, `gamma`, or `mixture`. The default is `facdensity(normal)`. `mixture` produces η as a two-factor mixture of normals. The mixing proportion for an $N(0,1)$ is specified by `mixpi()` as an integer between 0 and 100. For this option, `facmean()` and `facscale()` specify the mean and scale, respectively, of a component to be mixed with this $N(0,1)$. `facdensity(mixture)` is available only for `switchoprobitsim`.

`facscale(`*real*`)` specifies the scale of the LF distribution. The default is `facscale(1)`.

`facskew(`*real*`)` specifies the skewness of the LF distribution for use with the option `facdensity(chi2)`. The default is `facskew(2)`.

`startpoint(`*integer*`)` specifies the starting point for the Halton-sequence draws that are used to simulate the LF distribution. The default is `startpoint(5)`.

`facmean(`*real*`)` is particularly useful with the `gamma` distribution option; because all the LFs are normalized to mean zero, this parameter essentially controls the skewness of the gamma distribution used.

`vce(`*string*`)` specifies how to estimate the variance–covariance matrix corresponding to the parameter estimates. `cluster` *clustvar* specifies the cluster standard errors using *clustvar*. `robust` computes the robust variance–covariance matrix.

`sesimulations(`*integer*`)` specifies the number of draws of the parameter vector used in computing standard errors of ATEs and ATT. The default is `sesimulations(100)`.[2]

`mixpi(`*integer*`)` specifies the mixing proportion for a two-component mixture of normals. The default is `mixpi(50)`.

3.2 Postestimation

Syntax

`predict` [*type*] *newvar* [*if*] [*in*] [`, p11 p1#` `p0#` `te#` `tt#` `sete#` `sett#` `ptr xbout xbout0 xbout1 lf`]

Options

`p11` calculates the joint probability of participation in treatment and outcome 1; the default.

`p1#` calculates the joint probability of participation in treatment and outcome #.

`p0#` calculates the joint probability of nonparticipation in treatment and outcome #.

`te#` calculates the treatment effect on outcome #.

`tt#` calculates the treatment effect on the treated for outcome #.

`sete#` calculates the standard error of the treatment effect on outcome #.

`sett#` calculates the standard error of the treatment effect on the treated for outcome #.

`ptr` calculates the probability of treatment.

2. We take `sesimulations()` draws from the distribution of the parameter vector and calculate the ATT for each draw. The standard deviation of those marginal effects is reported as the standard error of the treatment on the treated. The results of this procedure are qualitatively similar to those using the delta method, but they are more consistently within the expected bounds of zero and one.

`xbout` (for use with `treatoprobitsim` postestimation) calculates the linear predictions for the outcome variable.

`xbout0` (for use with `switchoprobitsim` postestimation) calculates the linear predictions for the outcome variable for the untreated group.

`xbout1` (for use with `switchoprobitsim` postestimation) calculates the linear predictions for the outcome variable for the treated group.

`lf` calculates the likelihood contribution for each observation.

4 Monte Carlo simulations

In this section, I present results from Monte Carlo experiments in which I vary the distribution of the unobservables in the data-generating process (DGP) and compare estimates from `treatoprobitsim` and `switchoprobitsim` with those of `treatoprobit` and `switchoprobit`. The latter models impose bivariate normality on the error structure, whereas the former allow the researcher latitude in choosing the mixing distribution. I performed these experiments using 1,000, 2,000, and 5,000 observations with different DGPs, which correspond to different distributions for the LF (normal, logit, chi-squared, gamma, lognormal, and a nonparametric two-component finite mixture of normals). For each DGP and sample-size combination, I performed 200 experiments and used 100 draws from the distribution of the unobserved component.

4.1 Treatment-effects model

Table 1 provides parameter estimates from Monte Carlo experiments using the treatment-effects estimator described above. The leftmost column shows the true values of the parameters, while each column to the right shows the parameters produced by the estimator using different assumptions for the DGP—normal (Φ), logit, gamma with a mean of 4 [$\Gamma(4)$], chi-squared (χ^2), lognormal [$\ln(\Phi)$], and a finite mixture of normals. The finite mixture of normals is $0.75 \times N(0,1) + 0.25 \times N(4,25)$. I model the chi-squared distribution using a gamma distribution of the same mean, and I use the lognormal distribution to fit the model using a finite mixture. The top panel shows results for 1,000 observations, the middle panel for 2,000, and the bottom panel for 5,000.

The estimator does quite well for both the normal and the logit distributions. Looking at the table with these two DGPs, we can see that the estimator's accuracy increases as N increases. We can also see that for the skewed distributions, the estimator does fairly well at estimating the parameters in the outcome equation, including (most notably) the relative magnitude and size of the cutpoints (μ). This will be important when calculating the ATEs (below). The estimates of the selection equation parameters, although somewhat biased, are still plausible even though the coefficient of x_1 in the lognormal model remains stubbornly large.

Table 1. Monte Carlo parameter estimates: Treatment-effects model

	Model DGP						
	True	Φ	Logit	$\Gamma(4)$	χ^2	$\ln(\Phi)$	Mixture
				$N = 1000$			
treat: x_1	−2.25	−2.27	−2.24	−2.85	−2.37	−3.53	−2.22
treat: x_2	0.24	0.25	0.25	0.31	0.23	0.35	0.27
treat: _cons	0.80	0.80	0.79	0.83	0.54	0.71	0.62
out: x_1	−2.56	−2.85	−2.69	−2.40	−2.59	−2.30	−2.40
out: x_{treat}	−0.75	−0.84	−0.79	−0.90	−0.94	−0.79	−1.01
out: λ	−0.55	−0.60	−0.58	−0.40	−0.68	−0.30	−0.58
out: μ_1	−2.32	−2.59	−2.43	−2.33	−2.62	−2.35	−2.62
out: μ_2	−1.37	−1.53	−1.43	−1.42	−1.62	−1.45	−1.64
out: μ_3	0.60	0.65	0.63	0.48	0.53	0.44	0.44
out: μ_4	1.23	1.36	1.30	1.08	1.22	1.05	1.09
				$N = 2000$			
treat: x_1	−2.25	−2.26	−2.25	−2.83	−2.38	−3.48	−2.81
treat: x_2	0.24	0.23	0.25	0.28	0.23	0.35	0.27
treat: _cons	0.80	0.81	0.80	0.84	0.54	0.70	0.61
out: x_1	−2.56	−2.67	−2.55	−2.35	−2.55	−2.29	−2.35
out: x_{treat}	−0.75	−0.78	−0.79	−0.87	−0.90	−0.75	−0.96
out: λ	−0.55	−0.58	−0.53	−0.39	−0.67	−0.32	−0.58
out: μ_1	−2.32	−2.42	−2.33	−2.28	−2.58	−2.33	−2.56
out: μ_2	−1.37	−1.43	−1.39	−1.39	−1.59	−1.44	−1.59
out: μ_3	0.60	0.63	0.57	0.48	0.53	0.47	0.47
out: μ_4	1.23	1.28	1.20	1.07	1.21	1.08	1.10
				$N = 5000$			
treat: x_1	−2.25	−2.25	−2.18	−2.82	−2.37	−3.49	−2.80
treat: x_2	0.24	0.24	0.21	0.29	0.23	0.35	0.27
treat: _cons	0.80	0.80	0.79	0.84	0.53	0.69	0.61
out: x_1	−2.56	−2.57	−2.64	−2.36	−2.55	−2.28	−2.35
out: x_{treat}	−0.75	−0.77	−0.73	−0.85	−0.89	−0.72	−0.96
out: λ	−0.55	−0.54	−0.60	−0.41	−0.68	−0.33	−0.58
out: μ_1	−2.32	−2.34	−2.39	−2.28	−2.58	−2.33	−2.57
out: μ_2	−1.37	−1.39	−1.39	−1.38	−1.58	−1.43	−1.59
out: μ_3	0.60	0.59	0.67	0.50	0.55	0.50	0.46
out: μ_4	1.23	1.22	1.30	1.10	1.23	1.12	1.10

Marginal effects

Table 2 shows the marginal effects for the models whose parameter estimates are shown in table 1. The top, middle, and bottom panels show the estimates for 1,000, 2,000, and 5,000 observations, respectively. From left to right, each panel of three columns shows the results comparing the true marginal effect with a BIVN and the LF models. We show the true model fits in each panel because the ATEs are functions of the data, and they change slightly from simulated dataset to simulated dataset. In considering the ATEs, we look at both the point estimates for each outcome and the total transition probabilities (TTPs) predicted by the model. Because the marginal effects for all the outcomes must equal zero, we can look at the total (or absolute value) of the transitions from or to different outcomes.

Regardless of the sample size, the estimates for the BIVN and normal LF model are essentially identical, which is to be expected. A comparison of the LF logit and BIVN models also shows them to be very similar, although the LF model appears to deal with kurtosis slightly better than the BIVN model. The advantages of the LF models are most evident when the distribution of unobservables is skewed; no matter the sample size, for the chi-squared, gamma, and lognormal distributions, the TTPs fit by the BIVN model are off by 100% or more, while the estimates from the LF model are quite close to the true effects. This is particularly evident in the table for the chi-squared, gamma, and lognormal models when $N = 5000$: the marginal effects predicted by the BIVN model are off by more than 300% in these cases, whereas the results from the LF model show little bias.

Table 2. Monte Carlo results: ATEs, treatment-effects model

$N = 1000$

	Normal			Logit			Gamma		
	True	BIVN	LF	True	BIVN	LF	True	BIVN	LF
Outcome 1	0.085	0.087	0.085	0.084	0.099	0.088	0.085	0.224	0.106
Outcome 2	0.017	0.016	0.016	0.018	0.015	0.016	0.017	0.024	0.018
Outcome 3	0.000	0.001	0.001	0.000	0.003	0.001	0.000	0.011	0.001
Outcome 4	−0.010	−0.009	−0.009	−0.011	−0.008	−0.009	−0.010	−0.009	−0.011
Outcome 5	−0.091	−0.095	−0.094	−0.092	−0.109	−0.096	−0.092	−0.250	−0.115
	Chi-squared			Lognormal			Mixture		
	True	BIVN	LF	True	BIVN	LF	True	BIVN	LF
Outcome 1	0.084	0.291	0.099	0.085	0.178	0.093	0.085	0.073	0.108
Outcome 2	0.017	0.029	0.018	0.017	0.018	0.018	0.017	0.006	0.020
Outcome 3	0.001	0.015	0.002	0.000	0.010	0.004	−0.000	0.006	0.005
Outcome 4	−0.010	−0.011	−0.010	−0.010	−0.007	−0.010	−0.010	−0.001	−0.011
Outcome 5	−0.092	−0.324	−0.109	−0.091	−0.200	−0.105	−0.091	−0.083	−0.123

Table 2. Monte Carlo results: ATEs, treatment-effects model

$N = 2000$									
	Normal			Logit			Gamma		
	True	BIVN	LF	True	BIVN	LF	True	BIVN	LF
Outcome 1	0.085	0.086	0.086	0.085	0.074	0.088	0.084	0.259	0.103
Outcome 2	0.017	0.016	0.017	0.017	0.013	0.016	0.017	0.028	0.018
Outcome 3	0.000	0.001	0.001	0.000	0.000	0.001	0.001	0.014	0.001
Outcome 4	−0.010	−0.009	−0.010	−0.010	−0.008	−0.009	−0.010	−0.010	−0.011
Outcome 5	−0.092	−0.094	−0.094	−0.092	−0.080	−0.095	−0.092	−0.291	−0.112
	Chi-squared			Lognormal			Mixture		
	True	BIVN	LF	True	BIVN	LF	True	BIVN	LF
Outcome 1	0.084	0.320	0.095	0.085	0.203	0.087	0.085	0.033	0.103
Outcome 2	0.017	0.032	0.017	0.017	0.022	0.018	0.017	0.002	0.020
Outcome 3	0.001	0.017	0.002	0.000	0.012	0.004	−0.000	0.004	0.005
Outcome 4	−0.010	−0.012	−0.010	−0.010	−0.009	−0.010	−0.010	0.001	−0.010
Outcome 5	−0.092	−0.356	−0.105	−0.092	−0.228	−0.099	−0.091	−0.039	−0.118
$N = 5000$									
	Normal			Logit			Gamma		
	True	BIVN	LF	True	BIVN	LF	True	BIVN	LF
Outcome 1	0.085	0.084	0.086	0.085	0.052	0.077	0.085	0.277	0.100
Outcome 2	0.017	0.016	0.017	0.016	0.009	0.013	0.017	0.030	0.018
Outcome 3	0.000	0.000	0.000	−0.001	−0.001	−0.000	0.000	0.015	0.001
Outcome 4	−0.010	−0.010	−0.010	−0.010	−0.005	−0.008	−0.010	−0.011	−0.011
Outcome 5	−0.092	−0.091	−0.093	−0.091	−0.055	−0.082	−0.092	−0.310	−0.108
	Chi-squared			Lognormal			Mixture		
	True	BIVN	LF	True	BIVN	LF	True	BIVN	LF
Outcome 1	0.085	0.344	0.093	0.085	0.234	0.083	0.085	0.048	0.103
Outcome 2	0.017	0.034	0.017	0.017	0.026	0.017	0.017	0.003	0.020
Outcome 3	0.000	0.017	0.002	0.000	0.014	0.004	0.001	0.005	0.006
Outcome 4	−0.010	−0.013	−0.009	−0.010	−0.011	−0.010	−0.010	0.000	−0.010
Outcome 5	−0.092	−0.382	−0.102	−0.092	−0.263	−0.095	−0.092	−0.056	−0.118

4.2 Switching model

Table 3 shows the parameter estimates from simulations using the endogenous switching model. The top, middle, and bottom panels again show parameters for samples of 1,000, 2,000, and 5,000, respectively.[3] The most obvious characteristic of the experiments is that for skewed distributions of the LF, the parameter estimates for the selection-equation variables, λ_0, and λ_1, deteriorate notably. This does not seem to be helped by increasing the number of observations. The parameters for the outcome equations in

3. The scale of the factor density was set to 1 for all of these simulations.

the skewed density models fare somewhat better. Parameters in the normal and logistic models are consistent with expectations.

Table 3. Monte Carlo parameter estimates: Endogenous switching model

	Model DGP						
	True	Φ	Logit	$\Gamma(4)$	χ^2	$\ln(\Phi)$	Mixture
				$N = 1000$			
treat: x_1	−3.25	−3.27	−3.27	−3.87	−5.85	−6.24	−5.95
treat: x_2	−0.54	−0.55	−0.52	−0.66	−0.95	−1.04	−0.99
treat: _cons	0.80	0.81	0.80	0.85	1.37	1.38	1.34
out0: x_1	1.00	1.06	1.03	0.98	1.02	1.04	1.04
out1: x_1	−1.00	−1.05	−1.02	−0.97	−0.93	−0.92	−0.93
out0: λ_0	0.54	0.54	0.53	0.28	0.09	0.03	0.04
out1: λ_1	−0.63	−0.64	−0.63	−0.58	−0.23	−0.18	−0.20
out0: μ_{01}	−1.04	−1.07	−1.02	−0.75	−0.91	−0.92	−0.86
out0: μ_{02}	0.97	1.03	1.01	1.21	1.07	1.11	1.11
out0: μ_{03}	1.80	1.90	1.86	1.94	1.85	1.91	1.90
out1: μ_{11}	−2.04	−2.10	−2.08	−2.00	−1.87	−1.85	−1.79
out1: μ_{12}	−0.07	−0.06	−0.06	−0.05	−0.09	−0.10	−0.05
out1: μ_{13}	0.80	0.85	0.82	0.81	0.73	0.72	0.74
				$N = 2000$			
treat: x_1	−3.25	−3.26	−3.25	−3.82	−5.82	−6.16	−5.91
treat: x_2	−0.54	−0.54	−0.54	−0.65	−0.95	−1.02	−0.98
treat: _cons	0.80	0.80	0.80	0.86	1.40	1.39	1.34
out0: x_1	1.00	1.03	1.03	0.97	1.01	1.02	1.02
out1: x_1	−1.00	−1.02	−1.03	−0.93	−0.92	−0.92	−0.93
out0: λ_0	0.54	0.55	0.57	0.25	0.08	0.02	0.04
out1: λ_1	−0.63	−0.65	−0.64	−0.53	−0.23	−0.17	−0.18
out0: μ_{01}	−1.04	−1.06	−1.09	−0.71	−0.88	−0.89	−0.85
out0: μ_{02}	0.97	1.00	0.99	1.18	1.07	1.08	1.10
out0: μ_{03}	1.80	1.85	1.85	1.89	1.84	1.88	1.88
out1: μ_{11}	−2.04	−2.10	−2.07	−1.86	−1.85	−1.79	−1.73
out1: μ_{12}	−0.07	−0.08	−0.05	−0.04	−0.10	−0.11	−0.05
out1: μ_{13}	0.80	0.82	0.83	0.77	0.71	0.71	0.74
				$N = 5000$			
treat: x_1	−3.25	−3.25	−3.23	−3.81	−5.79	−6.19	−5.90
treat: x_2	−0.54	−0.54	−0.53	−0.63	−0.97	−1.03	−0.99
treat: _cons	0.80	0.80	0.79	0.83	1.39	1.40	1.35
out0: x_1	1.00	1.01	1.01	0.96	1.00	1.03	1.03
out1: x_1	−1.00	−1.01	−1.01	−0.92	−0.91	−0.91	−0.92
out0: λ_0	0.54	0.55	0.55	0.26	0.08	0.02	0.03
out1: λ_1	−0.63	−0.64	−0.64	−0.54	−0.21	−0.16	−0.17
out0: μ_{01}	−1.04	−1.06	−1.05	−0.71	−0.88	−0.88	−0.83
out0: μ_{02}	0.97	0.98	0.98	1.17	1.05	1.09	1.10
out0: μ_{03}	1.80	1.82	1.82	1.89	1.82	1.88	1.88
out1: μ_{11}	−2.04	−2.06	−2.07	−1.85	−1.79	−1.75	−1.69
out1: μ_{12}	−0.07	−0.07	−0.08	−0.04	−0.09	−0.10	−0.06
out1: μ_{13}	0.80	0.81	0.80	0.76	0.71	0.70	0.73

Marginal effects

Table 4 shows the marginal effects for the endogenous switching models. For the models using symmetric, unimodal distributions (normal and logit), the LF and the BIVN perform comparably well. Interestingly, although the parameter estimates for the LF models using skewed distributions of η are inconsistent, the marginal effects still show an advantage to using the LF structure. In particular, the TTP for the LF models is generally closer to the true model, and the distributions appear closer to the true models. For example, when $N = 5000$, the TTP is 0.273; when the simulated DGP is a gamma mixing distribution, the LF model estimate of the TTP is identical, while the BIVN estimate is 0.307. While there is still bias in the LF estimates of the marginal effects for each category, the estimates are generally better than with the BIVNs. For example, even in the gamma model for $N = 5000$, the distribution of marginal effects is more accurate for the LF model, although the point estimate for outcome 1 is more accurate for the BIVN.

Table 4. Monte Carlo results: ATEs, endogenous switching model

$N = 1000$

DGP	Normal			Logit			Gamma		
	True	BIVN	LatentF	True	BIVN	LatentF	True	BIVN	LatentF
Outcome 1	−0.156	−0.154	−0.154	−0.156	−0.150	−0.161	−0.157	−0.148	−0.188
Outcome 2	−0.116	−0.100	−0.104	−0.116	−0.087	−0.070	−0.117	−0.157	−0.087
Outcome 3	0.088	0.072	0.078	0.089	0.049	0.059	0.089	0.047	0.076
Outcome 4	0.184	0.182	0.179	0.184	0.188	0.172	0.185	0.258	0.199
	Chi-squared			Lognormal			Mixture		
	True	BIVN	LatentF	True	BIVN	LatentF	True	BIVN	LatentF
Outcome 1	−0.156	−0.157	−0.170	−0.156	−0.157	−0.166	−0.156	−0.153	−0.166
Outcome 2	−0.117	−0.148	−0.116	−0.117	−0.179	−0.138	−0.116	−0.171	−0.121
Outcome 3	0.089	0.082	0.086	0.089	0.096	0.094	0.089	0.085	0.085
Outcome 4	0.184	0.222	0.200	0.184	0.240	0.209	0.183	0.239	0.202

$N = 2000$

	Normal			Logit			Gamma		
	True	BIVN	LatentF	True	BIVN	LatentF	True	BIVN	LatentF
Outcome 1	−0.156	−0.155	−0.157	−0.156	−0.135	−0.144	−0.157	−0.146	−0.191
Outcome 2	−0.117	−0.103	−0.102	−0.117	−0.093	−0.082	−0.117	−0.162	−0.083
Outcome 3	0.089	0.071	0.076	0.089	0.046	0.055	0.089	0.047	0.072
Outcome 4	0.184	0.187	0.182	0.183	0.182	0.171	0.185	0.261	0.201
	Chi-squared			Lognormal			Mixture		
	True	BIVN	LatentF	True	BIVN	LatentF	True	BIVN	LatentF
Outcome 1	−0.157	−0.162	−0.177	−0.157	−0.159	−0.167	−0.157	−0.152	−0.164
Outcome 2	−0.116	−0.149	−0.112	−0.117	−0.178	−0.134	−0.117	−0.172	−0.123
Outcome 3	0.089	0.085	0.088	0.089	0.097	0.093	0.089	0.087	0.085
Outcome 4	0.185	0.225	0.202	0.184	0.241	0.208	0.184	0.238	0.202

$N = 5000$

	Normal			Logit			Gamma		
	True	BIVN	LatentF	True	BIVN	LatentF	True	BIVN	LatentF
Outcome 1	−0.156	−0.153	−0.156	−0.156	−0.140	−0.155	−0.156	−0.145	−0.192
Outcome 2	−0.117	−0.105	−0.101	−0.117	−0.096	−0.074	−0.117	−0.162	−0.081
Outcome 3	0.089	0.071	0.075	0.089	0.048	0.057	0.089	0.046	0.071
Outcome 4	0.184	0.186	0.182	0.184	0.188	0.173	0.184	0.261	0.201
	Chi-squared			Lognormal			Mixture		
	True	BIVN	LatentF	True	BIVN	LatentF	True	BIVN	LatentF
Outcome 1	−0.157	−0.155	−0.171	−0.156	−0.158	−0.167	−0.157	−0.158	−0.170
Outcome 2	−0.117	−0.150	−0.114	−0.117	−0.178	−0.135	−0.117	−0.171	−0.121
Outcome 3	0.089	0.083	0.085	0.089	0.098	0.094	0.089	0.091	0.088
Outcome 4	0.184	0.222	0.200	0.184	0.238	0.208	0.184	0.238	0.203

5 Examples

In this section, we use `treatoprobit` and `switchoprobit` to compare output for models that assume BIVN errors with those that use an LF structure with different mixing distributions. For our example, we use data from the National Health Interview Survey, which began fielding the 10-item food security module to assess 30-day household food security status in 2011. Of particular interest to researchers is the estimate of the effect of participation in the Supplemental Nutrition Assistance Program (SNAP) (formerly food stamps) on food security. One customary way to delineate food security status is to treat it as an ordered variable: 1 = high food security (0 affirmative responses to questions indicating food-insecure conditions); 2 = marginal food security (1–2 affirmative responses); 3 = low food security (3–5 affirmative responses); and 4 = very low food security ($\geq$ 6 affirmative responses).[4] We fit the model using BIVN errors (using `treatoprobit`) and then using an LF structure with normal, extreme value (logit), and gamma distributions. We show selected parameters from only the first and last of these models, and we provide marginal effects from all four.

```
. use nhisdataex
. local race hispanic black aian asian other_r
. local employ employed lookingfw retired wkdisabled
. local famstruc singadult multadult singpar
. local rhs age_p married hhsize female `famstruc´ `employ´ `race´
```

4. See Bickel et al. (2000) for more on the construction of the food security measure.

```
. treatoprobit fsstatd `rhs´ [pweight=normwgt], treat(snap `rhs´) vce(robust)
  (output omitted)
Treatment Effects Ordered Probit Regression     Number of obs     =     28,799
                                                Wald chi2(16)     =   3255.30
Log pseudolikelihood = -43914.348               Prob > chi2       =    0.0000
```

	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
snap						
age_p	-.000046	.0009737	-0.05	0.962	-.0019544	.0018624
married	-.4736267	.0295239	-16.04	0.000	-.5314924	-.415761
hhsize	.0939572	.0102386	9.18	0.000	.0738899	.1140245
female	.0804729	.0229467	3.51	0.000	.0354981	.1254477
singadult	-.5883686	.0439667	-13.38	0.000	-.6745419	-.5021954
multadult	-.517028	.0361596	-14.30	0.000	-.5878995	-.4461565
singpar	.3316473	.0386366	8.58	0.000	.2559209	.4073737
employed	-.347665	.0301505	-11.53	0.000	-.4067589	-.2885712
lookingfw	.2599816	.0396727	6.55	0.000	.1822245	.3377388
retired	-.2743294	.0823616	-3.33	0.001	-.435755	-.1129037
wkdisabled	.7274979	.0436212	16.68	0.000	.6420018	.8129939
hispanic	.1500298	.0271558	5.52	0.000	.0968053	.2032542
black	.4412894	.0288837	15.28	0.000	.3846784	.4979004
aian	.4165094	.1146362	3.63	0.000	.1918266	.6411921
asian	-.3969417	.0551626	-7.20	0.000	-.5050584	-.288825
other_r	.1596228	.0731043	2.18	0.029	.0163411	.3029045
_cons	-.6224873	.0694857	-8.96	0.000	-.7586769	-.4862977
fsstatd						
age_p	-.0002116	.0007819	-0.27	0.787	-.0017442	.001321
married	-.3152767	.0250112	-12.61	0.000	-.3642977	-.2662557
hhsize	.0436623	.0093384	4.68	0.000	.0253594	.0619652
female	.0776843	.0192436	4.04	0.000	.0399675	.115401
singadult	-.2385724	.0375653	-6.35	0.000	-.3121991	-.1649456
multadult	-.2463461	.0303752	-8.11	0.000	-.3058804	-.1868119
singpar	.1341663	.0348165	3.85	0.000	.0659272	.2024054
employed	-.1436358	.026113	-5.50	0.000	-.1948163	-.0924553
lookingfw	.4455762	.0342109	13.02	0.000	.3785241	.5126283
retired	-.3846497	.0715479	-5.38	0.000	-.5248809	-.2444184
wkdisabled	.8175104	.0377016	21.68	0.000	.7436166	.8914043
hispanic	.2741635	.0225514	12.16	0.000	.2299636	.3183634
black	.3660647	.0250125	14.64	0.000	.317041	.4150884
aian	.377977	.0868844	4.35	0.000	.2076866	.5482674
asian	-.2989624	.0439825	-6.80	0.000	-.3851666	-.2127583
other_r	.331947	.0617394	5.38	0.000	.21094	.452954
snap	-.7378056	.0452764	-16.30	0.000	-.8265457	-.6490655
/cut1	.2497769	.0594021	4.20	0.000	.1333509	.366203
/cut2	.5886479	.0598069	9.84	0.000	.4714286	.7058672
/cut3	1.028602	.0615911	16.70	0.000	.9078858	1.149318
/atanh_rho	.9786962	.0550681	17.77	0.000	.8707646	1.086628
rho	.7525009	.0238854			.7017624	.795644

```
Test of independent equations = 315.86. Probability of independent equations 0.
. estimates store treatbinorm
```

```
. forvalues i = 1/4 {
  2.         predict atebinorm`i´, te`i´
  3.         }
. treatoprobitsim fsstatd `rhs´ [pweight=normwgt], treatment(snap=`rhs´)
> simulationdraws(100) facdensity(normal) vce(robust)
```

(*output omitted*)

Treatment-effects Latent Factor Ordered Probit Regression

				Number of obs	= 28,799
				Wald chi2(16)	= 3195.03
Log pseudolikelihood = -43923.01				Prob > chi2	= 0.0000

	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
snap						
age_p	-4.05e-06	.00139	-0.00	0.998	-.0027283	.0027202
married	-.6679034	.0419763	-15.91	0.000	-.7501754	-.5856313
hhsize	.1332539	.0145025	9.19	0.000	.1048296	.1616782
female	.115858	.0325509	3.56	0.000	.0520595	.1796565
singadult	-.8302665	.0620713	-13.38	0.000	-.951924	-.7086091
multadult	-.7250665	.0513938	-14.11	0.000	-.8257964	-.6243365
singpar	.4769149	.0550523	8.66	0.000	.3690143	.5848155
employed	-.4921286	.0425036	-11.58	0.000	-.575434	-.4088231
lookingfw	.3652094	.0562598	6.49	0.000	.2549421	.4754767
retired	-.3698807	.1191494	-3.10	0.002	-.6034092	-.1363522
wkdisabled	1.024297	.0615694	16.64	0.000	.9036237	1.144971
hispanic	.2022707	.03831	5.28	0.000	.1271844	.277357
black	.6212871	.0410516	15.13	0.000	.5408274	.7017468
aian	.5833615	.162661	3.59	0.000	.2645517	.9021713
asian	-.5707779	.0783332	-7.29	0.000	-.7243082	-.4172476
other_r	.2222753	.1043758	2.13	0.033	.0177025	.4268481
_cons	-.8783926	.0984808	-8.92	0.000	-1.071412	-.6853737
fsstatd						
age_p	-.0010645	.0025596	-0.42	0.678	-.0060812	.0039522
married	-.9289767	.2609557	-3.56	0.000	-1.44044	-.417513
hhsize	.1210846	.0453401	2.67	0.008	.0322197	.2099495
female	.2414904	.0886705	2.72	0.006	.0676994	.4152814
singadult	-.6390487	.2157401	-2.96	0.003	-1.061891	-.2162059
multadult	-.6802417	.1976243	-3.44	0.001	-1.067578	-.2929051
singpar	.3583444	.1521919	2.35	0.019	.0600536	.6566351
employed	-.3889823	.1377068	-2.82	0.005	-.6588827	-.1190819
lookingfw	1.418526	.3698478	3.84	0.000	.6936372	2.143414
retired	-1.139736	.31616	-3.60	0.000	-1.759398	-.5200737
wkdisabled	2.546814	.6647396	3.83	0.000	1.243949	3.84968
hispanic	.8697599	.2196563	3.96	0.000	.4392415	1.300278
black	1.097521	.2880627	3.81	0.000	.5329282	1.662113
aian	1.135213	.4011641	2.83	0.005	.3489461	1.92148
asian	-.8897508	.2743625	-3.24	0.001	-1.427491	-.3520101
other_r	1.065179	.3344399	3.18	0.001	.409689	1.720669
snap	-1.857074	.5715196	-3.25	0.001	-2.977232	-.7369166
/cut1	.9631284	.2758039	3.49	0.000	.4225626	1.503694
/cut2	2.09923	.5091329	4.12	0.000	1.101348	3.097112
/cut3	3.580336	.8500518	4.21	0.000	1.914265	5.246407
/lambda	3.010609	.845674	3.56	0.000	1.353118	4.6681

```
Notes:
1. 100 Halton sequence-based quasirandom draws per observation
2. Latent factor density is normal
3. Standard deviation of factor density is 1
4. Test of independent equations = 12.67. Probability of independent equations 0.
. estimates store treatnormlf
. forvalues i = 1/4 {
  2.         predict atetrnorm`i´, te`i´
  3.         }
. summarize ate*, separator(4)
```

Variable	Obs	Mean	Std. Dev.	Min	Max
atebinorm1	28807	.2338827	.0458113	.0622014	.2877998
atebinorm2	28807	-.0395908	.0240583	-.0573134	.0564653
atebinorm3	28807	-.0603047	.0173913	-.073947	.0538011
atebinorm4	28807	-.1339872	.0672631	-.2877998	-.0122491
atetrnorm1	28831	.1906071	.0342642	0	.2317293
atetrnorm2	28831	-.0337568	.0203668	-.0532977	.049203
atetrnorm3	28831	-.0523845	.0141489	-.0686561	.0398286
atetrnorm4	28831	-.1044658	.0519967	-.2316024	0
atetruni1	28831	-.0352651	.0038232	-.0394366	0
atetruni2	28831	.006304	.0037264	-.0068387	.0096515
atetruni3	28831	.0110444	.0020984	-.0004996	.0125319
atetruni4	28831	.0179167	.0074701	0	.0381913
atetrlogit1	28831	.2176944	.0451741	0	.2737896
atetrlogit2	28831	-.0401536	.0245801	-.0596155	.0581049
atetrlogit3	28831	-.0601258	.0173666	-.0780429	.0542948
atetrlogit4	28831	-.1174149	.0623087	-.2737402	0

The measured marginal effect of SNAP on food security status from the model using binomial errors (`atebinorm1`) indicates that program participation increases the probability of high food security—no indications of food insecurity—and decreases the probability of any indications of food-insecure conditions. All the LF models agree on this, but the estimate from the gamma model is roughly half that of the binomial model.

We did the same exercise with the switching regression model, using `switchoprobit` as the reference model that assumes bivariate normality, and then we fit models using BIVN, normal, logit, and gamma mixing distributions. We show only the estimated marginal effects for these models.

```
. summarize atesw*, separator(4)
```

Variable	Obs	Mean	Std. Dev.	Min	Max
ateswbinorm1	28831	.1031208	.1270775	-.1243853	.4059875
ateswbinorm2	28831	.0492451	.0197997	-.0058915	.1450571
ateswbinorm3	28831	.0215504	.0259724	-.0522537	.1672863
ateswbinorm4	28831	-.1739163	.1369277	-.6270664	.0050449
ateswnorm1	28831	-.0239166	.2246735	-.4555572	.5409322
ateswnorm2	28831	-.0561917	.0802961	-.3010055	.1824843
ateswnorm3	28831	-.0537608	.1015212	-.3861693	.2170286
ateswnorm4	28831	.1338691	.1718182	-.7671751	.3114514
ateswgam1	28831	-.08672	.1345656	-.3407514	.3095435
ateswgam2	28831	.0346641	.0241287	-.0262465	.1137453
ateswgam3	28831	.0346484	.0357842	-.0591194	.1254644
ateswgam4	28831	.0174075	.0889321	-.444168	.1432199
ateswlogit1	28831	-.0165626	.206332	-.4375627	.4928646
ateswlogit2	28831	-.0199145	.0502061	-.1501406	.1823081
ateswlogit3	28831	-.0201537	.0654505	-.2108484	.2128532
ateswlogit4	28831	.0566308	.1767924	-.7499791	.2447087
ateswuni1	28831	-.1795781	.1560355	-.4909704	.2872165
ateswuni2	28831	.0322064	.0264156	-.0168294	.1428825
ateswuni3	28831	.0546978	.0447331	-.0416731	.1741264
ateswuni4	28831	.0926739	.0979321	-.3722987	.2334732
ateswchi1	28831	-.1803654	.1104336	-.3971085	.1677944
ateswchi2	28831	.053755	.0255381	0	.1208779
ateswchi3	28831	.0663736	.0336771	-.0164262	.1440052
ateswchi4	28831	.0602368	.0583046	-.225489	.1828505

The switching models highlight further how the choice of mixing distribution can affect the outcome. The BIVN model produces a result consistent with expectations: SNAP reduces the probability of very low food insecurity and increases the probability that recipients report fewer food-insecure conditions. However, the normal LF model suggests just the opposite: SNAP is associated with increased probability of low and very low food security and lower probabilities of high and marginal food security. These two estimates diverging so widely suggests that the distribution of the unobservables is not unimodal or symmetric: the simulation evidence suggests that when the distributions are well behaved in this way, the estimates are similar. The model that uses a logit distribution suggests that SNAP is associated with an increase in very low food security, while the gamma model indicates decreases in both high and very low food security and an increase in the likelihood of other outcomes. The mixture model suggests that SNAP is associated with an increase in marginal, low, and very low food security and a decrease in high food security. Obviously, choosing or averaging the model results would be critical to developing a meaningful understanding of both the econometric results and the underlying selection and outcome processes. Because these models are not generally nested, simple diagnostics will generally not suffice.

6 Conclusion

Until recently, the appeal of using BIVN error structure for treatment-effects models has been partly due to alternatives not being readily available in desktop software. The commands `treatoprobitsim` and `switchoprobitsim` allow the researcher to relax the assumption of bivariate normality of an ordered outcome and a potentially endogenous binary treatment. Although in practice one would want to use instruments to help identify participation, I show that, at least in some contexts, the choice of distribution is decisive for the measurement of treatment effects.

7 Acknowledgment

The views expressed in this article are those of the author and should not be attributed to the Economic Research Service or the USDA.

8 References

Aakvik, A., J. J. Heckman, and E. J. Vytlacil. 2005. Estimating treatment effects for discrete outcomes when responses to treatment vary: An application to Norwegian vocational rehabilitation programs. *Journal of Econometrics* 125: 15–51.

Bickel, G., M. Nord, C. Price, W. Hamilton, and J. Cook. 2000. Guide to Measuring Household Food Security, Revised January 2000. http://www.fns.usda.gov/guide-measuring-household-food-security-revised.

Deb, P., and P. K. Trivedi. 2006a. Maximum simulated likelihood estimation of a negative binomial regression model with multinomial endogenous treatment. *Stata Journal* 6: 246–255.

———. 2006b. Specification and simulated likelihood estimation of a non-normal treatment-outcome model with selection: Application to health care utilization. *Econometrics Journal* 9: 307–331.

Drukker, D. M., and R. Gates. 2006. Generating Halton sequences using Mata. *Stata Journal* 6: 214–228.

Greene, W. H., and D. A. Hensher. 2010. *Modeling Ordered Choices: A Primer*. Cambridge: Cambridge University Press.

Miranda, A., and S. Rabe-Hesketh. 2006. Maximum likelihood estimation of endogenous switching and sample selection models for binary, ordinal, and count variables. *Stata Journal* 6: 285–308.

Train, K. E. 2009. *Discrete Choice Methods with Simulation*. 2nd ed. Cambridge: Cambridge University Press.

About the author

Christian A. Gregory is an agricultural economist at the Economic Research Service, USDA.