United States Department of Agriculture

Estimating Treatment Effects for Ordered Outcomes Using Maximum Simulated Likelihood

Christian A. Gregory

Economic Research Service, USDA

Stata Users Conference, July 30-31, Columbus OH

The views expressed are those of the author and should not be attributed to ERS or USDA.

Background and Motivation

- ordered outcomes ubiquitous in social sciences
- used in many circumstances with latent variables
  - health status
  - injury severity
  - political preferences
  - disability status
  - grades
  - food security status
- Greene and Hensher (2010) provide a comprehensive overview

Background and Motivation

- How to handle ordered outcomes in context of bivariate treatment?
- Depends upon beliefs about unobservables:
  - unobservables in participation and outcome uncorrelated, use `teffects`
  - correlated unobservables: use `glamm` or `ssm` (Miranda and Rabe-Hesketh, 2006)
- Concerns
  - joint normality violated – estimates biased and inconsistent
  - quadrature routine in `gllamm` and `ssm` can be slow to converge
- Bayesian methods: Munkin and Trivedi (2008), Deb et al. (2006), Li and Tobias (2008), Li and Tobias (2014)

Background and Motivation

- A strategy: specify unobservables as latent factor (Aakvik et al., 2005).
- Advantages
  - can be specified as entering into treatment/outcome linearly
  - latent factor can follow any continuous distribution
  - current application: use halton-sequence Monte Carlo draws to improve in speed
- This method has been advantageous when outcomes are known not to follow normal distribution (Deb and Trivedi, 2006)
- We use it here to offer same flexibility for situation in which treatment and outcome belived to be marginally normal.

Where We Are Going

- Four estimators
- Model
- Latent Factor Approach
- Syntax
- Monte Carlo Results
- Examples
- Helpful Hints

Four estimators

Error Structure	Outcome Regime	
	Single	Treated/Untreated
Bivariate Normal	`treatoprobit`	`switchoprobit`
Latent Factor	`treatoprobitsim`	`switchoprobitsim`

Model

- For both models, we represent the treatment in the following away.

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

- Treatment effects model assumes a single regime for outcome:

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

Model

- Endogenous switching, separate regimes for treated and untreated:

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

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

- for $j = 1 \dots J$ possible outcomes and where the index $Y^*_{i,+} = X_{i,+}\beta + \varepsilon_{i,+}$

Latent Factor Approach

- Conventionally, assume that υ and $\varepsilon \sim \Phi_2(0,1)$
- We reformulate the model such that

$$\begin{aligned} \upsilon_i &= \lambda_T \eta_i + \zeta_i \\ \varepsilon_i &= \lambda_Y \eta_i + \iota_i, \end{aligned} \tag{1}$$

for treatment effects model, or

$$\upsilon_i = \lambda_T \eta_i + \zeta_i$$

$$\varepsilon_{i0} = \lambda_{Y0} \eta_{i0} + \iota_{i0} \tag{2}$$

$$\varepsilon_{i1} = \lambda_{Y1} \eta_{i1} + \iota_{i1} \tag{3}$$

for the switching model, where we assume that the marginal distributions of ζ and ι are normal, but that η need not be.

Latent Factor Approach

- Use Monte Carlo draws from chosen distribution of η. Likelihood function (treatment effect estimator) then is:

$$L = \frac{1}{S} \prod_{i=1}^{N} \sum_{s=1}^{S} \Phi(\tau * (Z_i \gamma + \lambda_T \eta_i)) \times$$
$$\sum_{k=1}^{K} (I * (Y = k)) \{ \Phi(\mu_k - X_i \beta + \lambda_Y \eta_i) - \Phi(\mu_{k-1} - X_i \beta + \lambda_Y \eta_i) \}, \quad (4)$$

- $\tau = 2 * T_i - 1$
- S is the number of simulation draws
- λs are loading factors–describe dependence between treatment and outcome.

Latent Factor Approach

- For switching estimator, likelihood is:

$$L = \frac{1}{S}\prod_{i=1}^{N}\sum_{s=1}^{S}\sum_{\ell=0}^{\ell=1}(I*(T_i=\ell))\times\Phi(\tau*(Z_i\gamma+\lambda_{\ell T}\eta_i))*\sum_{\ell=0}^{1}(I*(T_i=\ell))\times$$
$$\sum_{k=1}^{K}(I*(Y_i=k))\{\Phi(\mu_{\ell k}-X_{\ell i}\beta_\ell+\lambda_{\ell Y}\eta_{\ell i})-\Phi(\mu_{\ell k-1}-X_{\ell i}\beta_\ell+\lambda_{\ell Y}\eta_{\ell i})\}, \tag{5}$$

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

Marginal Effects: ATE

- Let δ be coefficient on treatment indicator. Then the average treatment effect (ATE) for the treatment effect model is

$$ATE_j^T = \frac{1}{N}\frac{1}{S}\sum_{i=1}^{N}\sum_{s=1}^{S}\{\Phi(\mu_k-(X_i\beta+\delta+\lambda\eta_{is}))-\Phi(\mu_{k-1}-(X_i\beta+\delta+\lambda\eta_{is}))\} - \{\Phi(\mu_k-(X_i\beta+\lambda\eta_{is}))-\Phi(\mu_{k-1}-(X_i\beta+\lambda\eta_{is}))\} \quad (6)$$

- For the switching regression, it is

$$ATE_k^S = \frac{1}{N}\frac{1}{S}\sum_{i=1}^{N}\sum_{s=1}^{S}\{\Phi(\mu_{1k}-(X_{1i}\beta_1+\lambda_1\eta_{is}))-\Phi(\mu_{1k-1}-(X_{1i}\beta_1+\lambda_1\eta_{is}))\} - \{\Phi(\mu_{0k}-(X_{0i}\beta+\lambda_0\eta_{is}))-\Phi(\mu_{0k-1}-(X_{0i}\beta_0+\lambda_0\eta_{is}))\} \quad (7)$$

Marginal Effects: ATT

- Let δ be coefficient on treatment indicator. Then the average treatment effect on the treated (ATT) for the treatment effect model is

$$
\begin{aligned}
ATT_j^T = \frac{1}{N}\frac{1}{S}\sum_{i=1}^{N}\frac{1}{E(\Phi(Z_i\gamma))}\Big[\sum_{s=1}^{S}\Phi(Z_i\gamma+\eta_{is})\times \\
\{\Phi(\mu_j-(X_i\beta+\delta+\lambda\eta_{is}))-\Phi(\mu_{j-1}-(X_i\beta+\delta+\lambda\eta_{is})) \\
-\Phi(\mu_j-(X_i\beta+\lambda\eta_{is}))+\Phi(\mu_{j-1}-(X_i\beta+\lambda\eta_{is}))\}\Big] \quad (8)
\end{aligned}
$$

Marginal Effects: ATT

- For the switching regression, it is

$$ATT_j^S = \frac{1}{N}\frac{1}{S}\sum_{i=1}^{N}\frac{1}{E(\Phi(Z_i\gamma))}\Big[\sum_{s=1}^{S}\sum_{\ell=0}^{\ell=1}(I*(T_i=\ell))\Phi(Z_i\gamma+\eta_{is})\times$$
$$\{\Phi(\mu_{1j}-(X_{1i}\beta_1+\lambda_1\eta_{is}))-\Phi(\mu_{1,j-1}-(X_{1i}\beta_1+\lambda_1\eta_{is}))$$
$$-\Phi(\mu_{0j}-(X_{0i}\beta_0+\lambda_0\eta_{is}))+\Phi(\mu_{0,j-1}-(X_{0i}\beta_0+\lambda_0\eta_{is}))\}.\Big] \quad (9)$$

- As is conventional for these models, we normalize λ_T to unity.

Syntax and Options

- Command syntax
  - `treat/switchoprobitsim` *depvar* [*indvars*] [*if*] [*in*] [*weight*] , `treat(`*depvar*$_T$= *varlist*`)` `simulationdraws(`*integer*`)` [`facdensity(`*string*`)` `facskew(`*real*`)` `facscale(`*real*`)` `startpoint(`*integer*`)` `vce(`*string*`)` `sesim(`*integer*`)` *maximize options*]
- Options
  - `treatment(`*depvar*$_T$= *varlist*`)` specifies treatment index as 0 or 1.
  - `sim(`*integer*`)` specifies the number of simulation draws from the distribution of η.
  - `facdensity(`*string*`)` specifies the density of the latent factor: default is standard normal; other options are `uniform`, `logit`, `gamma`, `chi2`, `lognormal` and `mixture` are also premitted.

Options

- `facskew`(*real*) is for use with the `chi2` option; default is 2.
- `facmean`(*real*) is particularly useful with `gamma` distribution option, essentially controls skewness of gamma distribution used; also, with `mixture` option, specifies the mean of Φ to be mixed with $\Phi(0,1)$
- `facscale`(*real*) specifies scale of distribution; default is 1. Also, specifies scale of mixing distribution with `mixture` option.
- `mixpi`*(integer (0-100))* specifies the weight on the $\Phi(0,1)$ in mixing specification.
- `startpoint`(*integer*) specifies the starting point for Halton sequence draws; default is 1.
- `sesim`(*integer*) number of simulations used to calculate standard error of ATT; default is 100.
- `vce`(*string*) specifies robust or cluster for variance estimation.

Postestimation

- `predict` predicts `p11` the probability of the first outcome for the treated group; this is the default.
- `predict` *varname*, `p0`*i* predicts the probability of outcome *i* for the untreated group.
- `predict` *varname*, `p1`*i* predicts the probability of outcome *i* for the treated group.

Postestimation

- `predict` *varname*, `tt`*i* predicts the average treatment effect on the treated for outcome *i*.
- `predict` *varname*, `te`*i* predicts the average treatment effect for outcome *i*.
- `predict` *varname*, `sett`*i* predicts the standard error of the average treatment effect on the treated for outcome *i*.
- `predict` *varname*, `sete`*i* predicts the standard error of average treatment effect for outcome *i*.

Monte Carlo Results

Table: Monte Carlo Results: ATE's, Treatment Effects Model, $N = 5,000$

	DGP					
	Normal			Logit		
	True	BiVN	LF	True	BiVN	LF
Outcome 1	0.085	0.084	0.086	0.085	0.052	0.077
Outcome 2	0.017	0.016	0.017	0.016	0.009	0.013
Outcome 3	0.000	0.000	0.000	-0.001	-0.001	-0.000
Outcome 4	-0.010	-0.010	-0.010	-0.010	-0.005	-0.008
Outcome 5	-0.092	-0.091	-0.093	-0.091	-0.055	-0.082

Monte Carlo Results

Table: Monte Carlo Results: ATE's, Treatment Effects Model, $N = 5,000$

	DGP					
	Gamma			Chi Squared		
	True	BiVN	LF	True	BiVN	LF
Outcome 1	0.085	0.277	0.100	0.085	0.344	0.093
Outcome 2	0.017	0.030	0.018	0.017	0.034	0.017
Outcome 3	0.000	0.015	0.001	0.000	0.017	0.002
Outcome 4	-0.010	-0.011	-0.011	-0.010	-0.013	-0.009
Outcome 5	-0.092	-0.310	-0.108	-0.092	-0.382	-0.102

Monte Carlo Results

Table: Monte Carlo Results: ATE's, Treatment Effects Model, $N = 5,000$

	DGP					
	Log Normal			Mixture		
	True	BiVN	LF	True	BiVN	LF
Outcome 1	0.085	0.234	0.083	0.085	0.048	0.103
Outcome 2	0.017	0.026	0.017	0.017	0.003	0.020
Outcome 3	0.000	0.014	0.004	0.001	0.005	0.006
Outcome 4	-0.010	-0.011	-0.010	-0.010	0.000	-0.010
Outcome 5	-0.092	-0.263	-0.095	-0.092	-0.056	-0.118

Monte Carlo Results

Table: Monte Carlo Results: ATE's, Switching Model, $N = 5,000$

	DGP					
	Normal			Logit		
	True	BiVN	LF	True	BiVN	LF
Outcome 1	-0.156	-0.153	-0.156	-0.156	-0.140	-0.155
Outcome 2	-0.117	-0.105	-0.101	-0.117	-0.096	-0.074
Outcome 3	0.089	0.071	0.075	0.089	0.048	0.057
Outcome 4	0.184	0.186	0.182	0.184	0.188	0.173

Monte Carlo Results

Table: Monte Carlo Results: ATE's, Switching Model, $N = 5,000$

	DGP					
	Gamma			Chi Squared		
	True	BiVN	LF	True	BiVN	LF
Outcome 1	-0.156	-0.145	-0.192	-0.157	-0.155	-0.171
Outcome 2	-0.117	-0.162	-0.081	-0.117	-0.150	-0.114
Outcome 3	0.089	0.046	0.071	0.089	0.083	0.085
Outcome 4	0.184	0.261	0.201	0.184	0.222	0.200

Monte Carlo Results

Table: Monte Carlo Results: ATE's, Switching Model, $N = 5,000$

	DGP					
	Log Normal			Mixture		
	True	BiVN	LatentF	True	BiVN	LatentF
Outcome 1	-0.156	-0.158	-0.167	-0.157	-0.158	-0.170
Outcome 2	-0.117	-0.178	-0.135	-0.117	-0.171	-0.121
Outcome 3	0.089	0.098	0.094	0.089	0.091	0.088
Outcome 4	0.184	0.238	0.208	0.184	0.238	0.203

Example

Table: Example: Food Security and SNAP

	ATE: Treatment Effects Model		
	BiVN	LF Logit	LF Gamma
High Food Security	.23	.21	.12
Marginal Food Security	-.04	-.04	-.03
Low Food Security	-.06	-.06	-.04
Very Low Food Security	-.13	-.11	-.05
	N=28,831		

Data: National Health Interview Survey, 2011-2013, Low Income Sample

Example

Table: Example: Food Security and SNAP

	ATE: Switching Model		
	BiVN	LF Logit	LF Mixture
High Food Security	.01	-.01	-.27
Marginal Food Security	.09	-.09	.06
Low Food Security	-.05	-.03	.10
Very Low Food Security	-.14	.13	.11
	N=28,831		

Data: National Health Interview Survey, 2011-2013, Low Income Sample

Comments and Hints

- `-sim` routines report a likelihood ratio test of independent (`treat`) and single (`switch`) regimes.
- Using the `mixture` option makes tests of regime differences difficult. Good robustness check if you don't care about nuisance parameters.
- $\sim$ 100 simulation draws is nearly optimal in terms of accuracy in most applications; $\leq$ 80 is not recommended
- Models using different distributions are, in general, not nested. Model selection is crucial. Test proposed by Vuong (1989) can be useful / easy to calculate.

Going Further

- copula based modeling of dependence structures
- benefits of modeling with and without counterfactuals

Thank You!

Thank You!