No. 11-06
2011

AN ORDERED CATEGORICAL RESPONSE MODEL WITH ENDOGENOUS SWITCHING: SIMULATION EXERCISES AND AN APPLICATION TO STATE HEALTH.

Johnatan Cardona Jimenez
Andrés Ramírez Hassan

Documentos de trabajo
Economía y Finanzas
Centro de Investigaciones Económicas y Financieras (CIEF)

UNIVERSIDAD EAFIT
Abierta al mundo

Foto: Juan Carlos Ramírez

manuscript No.
(will be inserted by the editor)

An ordered categorical response model with endogenous switching: Simulation Exercises and an Application to State Health

Andrés Ramírez Hassan · Johnatan Cardona Jimenez

Received: date / Accepted: date

Abstract We develop and estimate a full information maximum likelihood ordered categorical response model with endogenous switching. Simulation exercises show that estimated parameters are unbias and consistent. And application to Internet access indicates that there are selection and moral hazard.

Keywords Endogenous Switching · Categorical Response · Evolution Strategy

1 Introduction

In this work we developed a specification, estimation and simulation framework for an ordered categorical response model with endogenous switching. Specifically, we analyze the effect of endogenous binary choice indicator (treatment) on an ordinal dependent variable (outcome). Specification is based on latent variable structure and Maximum Likelihood is used to estimate the model. An evolutive algorithm is used to maximize the likelihood function (see Storn and Price (1997)), and some simulation exercises are done in order to analyze small sample propertier of the estimators.

The approach in this paper is based on connecting treatment and outcome via unobservabed heterogeneity where we assume a multivariate Normal distribution for the stochastic heterogeneity associated to each variable. Advantage of the multivariate normal specification is that it imposes no bounds on the correlations and many might find the setup more familiar (Deb and P.Trivedi, 2006). However, there is one mayor disadvantages associated to specify a multivariate Normal distribution. The error in the outcome equation is a multiple of the error in the treatment equation, plus some noise that is independent of treatment decision. The Bayes theorem is used to build the joint distribution of endogenous treatment and outcome using an initial specification of the marginal distribution for treatment and the conditional distribution for outcome.

Storn and Price (1997) performed an stochastic optimization algorithm which was called Differential Evolution (DE). This algorithm is robust, easy to use, and lends itself very well to parallel computation. Mayer et al. (2005) apply a DE in the optimization of a beef model and they find that DE performs better than Genial (a real-value genetic algorithm). They showed that the DE parameters work effectively but can be improved in the sense that smaller populations can be considerably more efficient. Yang et al. (2007) and others propose two new efficient DE variants, named DECCI and DECC, for high-dimensional optimization. Experimental results show that the algorithms have superior performance on a set of widely used benchmark functions. Babu and Angira (2006) perform an improvmen to original DE that attempts to speed up convergence rate. Ardia et al. (2011) present an introduction to the R package DEoptim which implements the DE algorithm and give a view its utility for financial applications by solving a non-convex portfolio optimization problem.

Munkin and Trivedi (2008) build a Bayesian ordered Probit model with endegenous selection which is used to analyze the effects of different types of medical insurance plans on the level of hospital utilization. Deb and

The author would like to thank Universidad EAFIT, especially the Direction of Research for the funding.

Andrés Ramírez Hassan
Universidad EAFIT, Carrera 49 7 Sur 50. Bloque 26 oficina 204
Tel.: +00 57 4 29619549
E-mail: aramir21@eafit.edu.co

Johnatan Cardona Jimenez
second address

P.Trivedi (2006) use latent factor structure for a class of nonlinear, non-normal microeconometric model of treatment and outcome with selection. Their specification permits both negative and positive correlations, although it places bounds on the correlation. Prieger (2002) uses the copula approach to estimate a selectivity model to health care usage. However, the copula also places restrictions on the pattern of allowable correlations. Melo et al. (2002) use simultaneous equations methods, including discrete factor estimation, to test the effect of medicare HMOs on utilization when strong controls for selection bias are imposed. Cardon and Hendel (2001) use moment conditions to built...

The rest of the paper is organized as follows. Section (2) describes the model. Section (3) presents the simulations exercises. Section (4) shows an application on Internet Access and Section (5) concludes.

2 The model

We present an extension of the Roy model (Roy, 1951) or standard switching regression model with endogenous switching (Maddala, 1983)[1] where the observed outcomes are ordered categorical responses.

We define the ordered response endogenous switching model as follows. A latent variable y_1^* determines whether the outcome observed is y_2 or y_3. Specifically, we observe whether y_1^* is positive or negative,

$$y_1 = \begin{cases} 1, & y_1^* > 0 \\ 0, & y_1^* \leqslant 0 \end{cases} \tag{1}$$

And observe exactly one of y_2 or y_3 according to

$$y = \begin{cases} y_2, & y_1^* > 0 \\ y_3, & y_1^* \leqslant 0 \end{cases} \tag{2}$$

where,
$y_2 = j \Leftrightarrow \alpha_{2,j-1} < y_2^* \leqslant \alpha_{2,j}$, $j = 1,2,\ldots,m$, $\alpha_{2,0} = -\infty$ and $\alpha_{2,m} = \infty$.
$y_3 = k \Leftrightarrow \alpha_{3,k-1} < y_3^* \leqslant \alpha_{3,k}$, $k = 1,2,\ldots,m$, $\alpha_{3,0} = -\infty$ and $\alpha_{3,m} = \infty$.
y_2^* and y_3^* are latent variables.

Given (1) and (2), for $y_1 = 1$ we observe $y = y_2 = j$, with probability equal to the probability that $y_1 = 1$ times the conditional probability of $y_2 = j$ given that $y_1 = 1$. Thus for $y_2 = j$ the density is $f(y_2 = j|y_1 = 1)f(y_1 = 1)$, $j = 1,2,\ldots,m$. For $y_1 = 0$ we observe $y = y_3 = k$, with probability equal to the probability that $y_1 = 0$ times the conditional probability of $y_3 = k$ given that $y_1 = 0$, then the density is $f(y_k = k|y_1 = 0)f(y_1 = 0)$,$k = 1,2,\ldots,m$.

The joint density for one observation can be written as

$$f(y_1,y) = \begin{cases} f(y_2 = j|y_1 = 1)f(y_1 = 1), & y_2 = j, y_1 = 1, j = 1,2,\ldots,m \\ f(y_3 = k|y_1 = 0)f(y_1 = 0), & y_3 = k, y_1 = 0, k = 1,2,\ldots,m \end{cases}$$

or

$$f(y_1,y) = \left(\prod_{j=1}^{m} f(y_2 = j|y_1 = 1)^{y_j} f(y_1 = 1)\right)^{y_1} \left(\prod_{k=1}^{m} f(y_3 = k|y_1 = 0)^{y_k} f(y_1 = 0)\right)^{1-y_1} \tag{3}$$

where

$$y_j = \begin{cases} 1, & y_2 = j \\ 0, & y_2 \neq j \end{cases} \tag{4}$$

and

$$y_k = \begin{cases} 1, & y_3 = k \\ 0, & y_3 \neq k \end{cases} \tag{5}$$

Assumption 1 *We assume that the latents variables are linear in regressors with additive erros, i.e.* $y_1^* = x_1^T\beta_1 + \varepsilon_1$, $y_2^* = x_2^T\beta_2 + \varepsilon_2$ *and* $y_3^* = x_3^T\beta_3 + \varepsilon_3$.

[1] This model is also known as Tobit type 5 model (Amemiya, 1985).

Assumption 2 *The correlated errors are joint normal*

$$\begin{bmatrix}\varepsilon_1\\ \varepsilon_2\\ \varepsilon_3\end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix}0\\0\\0\end{bmatrix}, \begin{bmatrix}1 & \sigma_{12} & \sigma_{13}\\ \sigma_{21} & \sigma_2^2 & \sigma_{23}\\ \sigma_{31} & \sigma_{32} & \sigma_3^2\end{bmatrix}\right) \tag{6}$$

For the ordered response model with endogenous switching we use the following propositions.

Proposition 1 *Given (6) then the errors have the following form: $\varepsilon_2 = \sigma_{12}\varepsilon_1 + \xi_2$, where $\xi_2 \sim \mathcal{N}(0, \sigma_2^2 - \sigma_{12}^2)$, and the random variable ξ_2 is independient of ε_1. And $\varepsilon_3 = \sigma_{13}\varepsilon_1 + \xi_3$, where $\xi_3 \sim \mathcal{N}(0, \sigma_3^2 - \sigma_{13}^2)$, and the random variable ξ_3 is independient of ε_1.*

See (Heckman, 1979) and (Cameron and Trivedi, 2005).

Proposition 2 *Suppose $z \sim \mathcal{N}(0,1)$ and c a constant cutoff. Then the left-truncated moments of z are*

- $E(z|z>c) = \dfrac{\phi(c)}{1-\Phi(c)}$
- $E(z^2|z>c) = 1 + \dfrac{c\phi(c)}{1-\Phi(c)}$
- $Var(z|z>c) = 1 + \dfrac{c\phi(c)}{1-\Phi(c)} - \left(\dfrac{\phi(c)}{1-\Phi(c)}\right)^2$

See Appendix (6.1).

Theorem 3 *Given assumptions (1) and (2) then $f(y_1 = 1) = \Phi(x_1^T\beta_1)$ and $f(y_1 = 0) = 1 - \Phi(x_1^T\beta_1)$.*

See Appendix (6.2).

Theorem 4 *Given assumptiosn (1) and (2) then*

$$f(y_2 = j|y_1 = 1) = \Phi\left(\frac{\alpha_{2,j} - x_2^T\beta_2 - \mu_{\varepsilon_2|\varepsilon_1 > -x_1^T\beta_1}}{\sigma_{\varepsilon_2|\varepsilon_1 > -x_1^T\beta_1}}\right) - \Phi\left(\frac{\alpha_{2,j-1} - x_2^T\beta_2 - \mu_{\varepsilon_2|\varepsilon_1 > -x_1^T\beta_1}}{\sigma_{\varepsilon_2|\varepsilon_1 > -x_1^T\beta_1}}\right) \tag{7}$$

and

$$f(y_3 = k|y_1 = 0) = \Phi\left(\frac{\alpha_{3,k} - x_3^T\beta_3 - \mu_{\varepsilon_3|\varepsilon_1 \leqslant -x_1^T\beta_1}}{\sigma_{\varepsilon_3|\varepsilon_1 \leqslant -x_1^T\beta_1}}\right) - \Phi\left(\frac{\alpha_{3,k-1} - x_3^T\beta_3 - \mu_{\varepsilon_3|\varepsilon_1 \leqslant -x_1^T\beta_1}}{\sigma_{\varepsilon_3|\varepsilon_1 \leqslant -x_1^T\beta_1}}\right) \tag{8}$$

where $\mu_{\varepsilon_2|\varepsilon_1 > -x_1^T\beta_1} = \sigma_{12}\lambda(x_1^T\beta_1)$, $\mu_{\varepsilon_3|\varepsilon_1 \leqslant -x_1^T\beta_1} = -\sigma_{13}\lambda(-x_1^T\beta_1)$, $\sigma^2_{\varepsilon_2|\varepsilon_1 > -x_1^T\beta_1} = \sigma_2^2 - \sigma_{12}^2\lambda(x_1^T\beta_1)\left[x_1^T\beta_1 + \lambda(x_1^T\beta_1)\right]$, $\sigma^2_{\varepsilon_3|\varepsilon_1 \leqslant -x_1^T\beta_1} = \sigma_3^2 + \sigma_{13}^2\lambda(-x_1^T\beta_1)\left[x_1^T\beta_1 - \lambda(-x_1^T\beta_1)\right]$ and $\lambda(x_1^T\beta_1) = \dfrac{\phi(x_1^T\beta_1)}{\Phi(x_1^T\beta_1)}$ is the inverse Mills ratio.

Observe that whether $\sigma_{12} = \sigma_{13} = 0$, there is not selection effect based on unobservable variables, and we obtain standart ordered Probit model.
See Appendix (6.3).

Assumption 5 *We consider estimation given a sample $(y_{1i}, y_{2i}, y_{3i}, x_{1i}, x_{2i}, x_{3i}), i = 1, 2, \dots, N$, where we assume independence over i.*

Given the density (3) and the assumption (5) then the log-likelihood function is

$$\begin{aligned}\mathcal{L}(\beta,\Sigma) = \sum_{i=1}^{N} y_{1i}\left(\sum_{j=1}^{m} y_{ji}[logf(y_{2i} = j|y_{1i} = 1)] + logf(y_{1i} = 1)\right)\\ + (1-y_{1i})\left(\sum_{k=1}^{m} y_{ki}[logf(y_{3i} = k|y_{1i} = 0)] + logf(y_{1i} = 0)\right)\end{aligned}$$

where $\beta = [\beta_1, \beta_2, \beta_3]^T$ and $\Sigma = \begin{bmatrix}1 & \sigma_{12} & \sigma_{13}\\ \sigma_{21} & \sigma_2^2 & .\\ \sigma_{31} & . & \sigma_3^2\end{bmatrix}$, because $\sigma_{23} = \sigma_{32}$ can not be identified.
Differentiating with respect to the parameters of the model we can get the Full Information Maximum Likelihood Estimators (FIMLE).

Theorem 6 *Given assumptions (1) and (2) then the Average Treatmen Effect is*

$$ATE = E(y_{2i}|y_{1i}=1) - E(y_{2i}|y_{1i}=0) = \sum_{j=1}^{m} f(y_{2i}=j|y_{1i}=1)j - \sum_{j=1}^{m} f(y_{2i}=j|y_{1i}=0)j \tag{9}$$

where $f(y_{2i}=j|y_{1i}=0) = \Phi\left(\frac{\alpha_{2,j} - x_{2i}^T\beta_2 - \mu_{\varepsilon_2|\varepsilon_1 \leqslant -x_{1i}^T\beta_1}}{\sigma_{\varepsilon_2|\varepsilon_1 \leqslant -x_{1i}^T\beta_1}}\right) - \Phi\left(\frac{\alpha_{2,j-1} - x_2^T\beta_{2i} - \mu_{\varepsilon_2|\varepsilon_1 \leqslant -x_{1i}^T\beta_1}}{\sigma_{\varepsilon_2|\varepsilon_1 \leqslant -x_{1i}^T\beta_1}}\right)$, $\mu_{\varepsilon_2|\varepsilon_1 \leqslant -x_{1i}^T\beta_1} = -\sigma_{12}\lambda(-x_{1i}^T\beta_1)$ *and* $\sigma^2_{\varepsilon_2|\varepsilon_1 \leqslant -x_{1i}^T\beta_1} = \sigma_2^2 + \sigma_{12}^2\lambda(-x_{1i}^T\beta_1)\left[x_{1i}^T\beta_1 - \lambda(-x_{1i}^T\beta_1)\right]$.

And

$$ATE = E(y_{3i}|y_{1i}=0) - E(y_{3i}|y_{1i}=1) = \sum_{k=1}^{m} f(y_{3i}=k|y_{1i}=0)k - \sum_{k=1}^{m} f(y_{3i}=k|y_{1i}=1)k \tag{10}$$

where $f(y_{3i}=k|y_{1i}=1) = \Phi\left(\frac{\alpha_{3,k} - x_{3i}^T\beta_3 - \mu_{\varepsilon_3|\varepsilon_1 > -x_{1i}^T\beta_1}}{\sigma_{\varepsilon_3|\varepsilon_1 > -x_{1i}^T\beta_1}}\right) - \Phi\left(\frac{\alpha_{3,k-1} - x_{3i}^T\beta_3 - \mu_{\varepsilon_3|\varepsilon_1 > -x_{1i}^T\beta_1}}{\sigma_{\varepsilon_3|\varepsilon_1 > -x_{1i}^T\beta_1}}\right)$, $\mu_{\varepsilon_3|\varepsilon_1 > -x_{1i}^T\beta_1} = \sigma_{13}\lambda(x_{1i}^T\beta_1)$ *and* $\sigma^2_{\varepsilon_3|\varepsilon_1 > -x_{1i}^T\beta_1} = \sigma_3^2 - \sigma_{13}^2\lambda(x_{1i}^T\beta_1)\left[x_{1i}^T\beta_1 + \lambda(x_{1i}^T\beta_1)\right]$.

See Appendix (6.3).

3 Simulation exercises

We will now present a simulation exercise to show some properties of unbiasedness and consistency asociated to model estimation ((11) and (12)) by maximizing the likelihood. **Differencial Evolution (DE)** is used to provide maximum likelihood estimation. We have fixed parameter values β_1, β_{11}, β_2, β_{22}, β_3, β_{33}, σ_2^2,σ_3^2, σ_{12}, σ_{13} and we simulate errors ε_1, ε_2 and ε_3 for obtain responses y_1^*, y_2^* and y_3^*. The support of the covariates was fixed: $x_1 \in [0,40]$, $x_{11} \in [0,80]$, $x_2 \in [0,100]$, $x_{22} \in [80,150]$, $x_3 \in [0,100]$, $x_{33} \in [80,150]$. And different sample size was evaluated: $n = 50, 200, 400, 1000$. From figure 1 we show the algorithm employed for parameter estimation.

Model Specification

$$y_1 = \begin{cases} 1, & y_1^* > 0 \\ 0, & y_1^* \leqslant 0 \end{cases} \tag{11}$$

And observe exactly one of y_2 or y_3 according to

$$y = \begin{cases} y_2, & y_1^* > 0 \\ y_3, & y_1^* \leqslant 0 \end{cases} \tag{12}$$

where,
$y_2 = j \Leftrightarrow \alpha_{2,j-1} < y_2^* \leqslant \alpha_{2,j}$, $j = 1,2,\ldots,m$, $\alpha_{2,0} = -\infty$ and $\alpha_{2,m} = \infty$.
$y_3 = k \Leftrightarrow \alpha_{3,k-1} < y_3^* \leqslant \alpha_{3,k}$, $k = 1,2,\ldots,m$, $\alpha_{3,0} = -\infty$ and $\alpha_{3,m} = \infty$.
$y_1^* = x_1\beta_1 + x_{11}\beta_{11} + \varepsilon_1$; $y_2^* = x_2\beta_2 + x_{22}\beta_{22} + \varepsilon_2$; $y_3^* = x_3\beta_3 + x_{33}\beta_{33} + \varepsilon_3$.

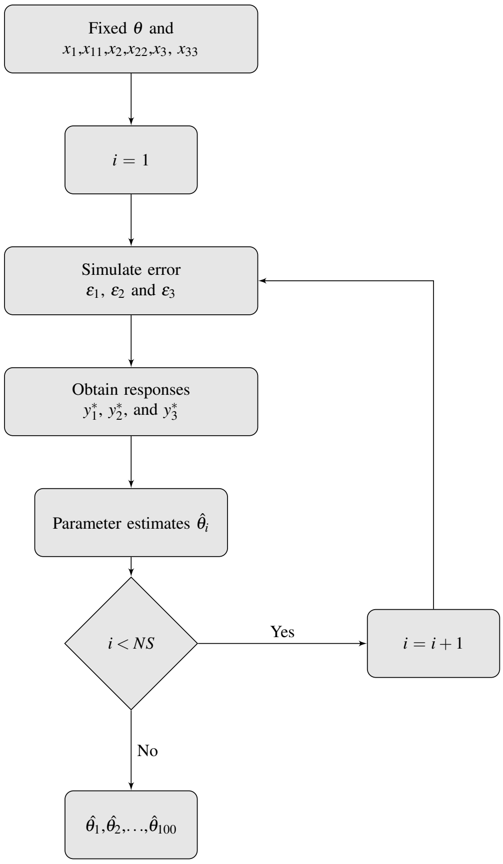

Fig. 1 Algorithm to estimate $\theta = (\beta_1, \beta_{11}, \beta_2, \beta_{22}, \beta_3, \beta_{33}, \sigma_2^2, \sigma_3^2, \sigma_{12}, \sigma_{13}, \alpha_{2,j-1}, \alpha_{2,j}, \alpha_{3,k-1}, \alpha_{3,k})$. $NS = 100$ is the number of simulations.

The simulation exercises was performed in the R software through DEoptim package. For attempt to speed up convergence rate, we have vectorized and parallelized all code. The machine characteristic from the simulation exercises was run are as follows: DEll PowerdEdge 2950, 8 processors Intel EM64T, 4 GB of RAM and Linux/GNU Centos 5.4 operating system.

Table 1 Results of simulation study for ordered categorical response model with endogenous switching. Fixed parameters values of the model (11) and (12) and their estimation under different sample sizes.

		$\hat{\theta}$ under different sample size ($ee(\hat{\theta})$)			
		Sample size			
Parameters	Real value	50	200	400	1000
β_1	0.075	0.048(0.042)	0.055(0.014)	0.079(0.026)	0.075(0.010)
β_{11}	0.05	0.036(0.035)	0.039(0.025)	0.065(0.036)	0.040(0.012)
β_2	0.13	0.092(0.014)	0.090(0.018)	0.101(0.016)	0.121(0.012)
β_{22}	0.11	0.13(0.037)	0.119(0.030)	0.121(0.026)	0.105(0.019)
β_3	0.14	0.171(0.090)	0.159(0.080)	0.152(0.032)	0.141(0.018)
β_{33}	0.12	0.18(0.023)	0.131(0.078)	0.14(0.041)	0.124(0.030)
σ_2^2	2.83	3.891(0.627)	2.994(0.461)	2.982(0.023)	2.891(0.018)
σ_3^2	3.10	4.104(0.439)	3.992(0.131)	3.892(0.016)	3.502(0.013)
σ_{12}	60	55.20(2.081)	58.10(1.139)	58.561(1.511)	58.901(1.021)
σ_{13}	50	45.01(2.013)	45.00(1.634)	48.000(1.091)	48.000(1.039)
$\alpha_{2,j-1}$	12.65	8.001(1.852)	9.901(1.658)	10.001(1.204)	11.001(1.018)
$\alpha_{2,j}$	24.93	26.99(1.011)	26.19(0.767)	25.99(0.823)	24.99(0.612)
$\alpha_{3,k-1}$	13.07	10.46(1.068)	12.01(0.587)	12.06(0.987)	12.84(0.423)
$\alpha_{3,k}$	26.67	28.29(1.688)	26.92(1.348)	26.89(0.818)	26.34(0.751)

From the table 1 we can see the real or fixed values of parameters model (11) and (12) and their estimation under different simulation scenaries. We can observe that when the sample size grows, the parameters estimation is better in the sense that parameters estimated seems to be near of the real or fixed values.

Assessment of the estimation quality

For evaluation of the estimation quality for the model parameters ((11) and (12)), we have measured the relative error which is defined in the expression (13):

$$\delta(\hat{\theta}) = \left|\frac{\theta - \hat{\theta}}{\theta}\right| \tag{13}$$

In the simulation process we have obtained several values of the parameters estimated under different simulated samples (see diagram 1), then we have defined a average relative error:

$$\bar{\delta}(\hat{\theta}) = \sum_{i=1}^{NS} \delta(\hat{\theta}_i)/NS. \tag{14}$$

From the table 2 we show the average relative error for the parameters estimated. We can see that the relative error decreases when the sample size grows, which makes sense with the properties of unbiased and consistency.

Table 2 Results of simulation study for ordered categorical response model with endogenous switching. Average relative error for assessment of the parameters estimation under different sample sizes.

	Average relative error $\bar{\delta}(\hat{\theta})$ Sample size			
Parameters	50	200	400	1000
β_1	39.42%	27%	5.922%	0.09%
β_{11}	28.99%	22%	28.23%	22%
β_2	28.99%	30%	24.96%	8.53%
β_{22}	19.03%	9.818%	10.66%	10.10%
β_3	20.03%	13%	9.925%	8.76%
β_{33}	23.53%	19.16%	16.65%	17.33%
σ_2^2	38.78%	6.65%	5.85%	2.20%
σ_3^2	34.52%	28.70%	26.71%	12.99%
σ_{12}	9.10%	3.96%	4.25%	1.86%
σ_{13}	10.23%	9.56%	5.45%	3.94%
$\alpha_{2,j-1}$	35.73%	21.94%	20.92%	12.77%
$\alpha_{2,j}$	9.31%	6.24%	4.28%	1.68%
$\alpha_{3,k-1}$	22.76%	8.93%	7.91%	2.60%
$\alpha_{3,k}$	7.10%	1.23%	0.84%	2.19%

4 Application: State Health

We have performed an application of the model 1 and 2 with data from Medellin Living Standars Survey (ECV/2007). We take the State Health as the response variable which has three levels (1: bad health, 2: good health, 3: excellent health). As covariates we have taken the gender (Male:1, Female:0), the age, the socioeconomic stratum which has six levels (Stratum1: Under-Lower stratum, Stratum2:Lower stratum, Stratum3:Medium-Low strata, Stratum4: Average strata, Stratum5: Medium-High strata, Stratum6: High), health care spending, other household expenses, and whether perform or not physical exercise (1:yes, 0:no). As endogenous switching we take whether an individual is covered or uncovered by Subsidized Regimen (1:covered, 0:uncovered).

Table 3 Results of the application for health state.

Parameter	(y_1^*)	(y_2^*)	(y_3^*)
HC spending	-28378.74	-13018.10	-3971.54
household expenses	37772.97	72466.84	84838.67
Stratum2		1983.78	-82068.95
Stratum3		47027.93	90097.30
Stratum4		43562.83	-28881.18
Stratum5		18004.75	64250.01
Stratum6		-93242.42	-30468.89
Gender		-52201.02	-882.36
phy. exer.		41076.65	8127.56

Table 4 Results of the application for health state.

Parameter	Estimation
σ_2^2	42050.31
σ_{12}	92493.36
σ_3^2	6412.72
σ_{13}	10381.27
α_{21}	-62107.83
α_{22}	-6854.83
α_{31}	31241.53
α_{32}	67028.99

5 Conclusion

We have developed an ordered categorical response model with endogenous switching, and we have shown through simulation exercises that its estimators are consistent, because when the sample size grows the estimated parameters tend to be near the real or fixed values. We have employed Evolution Strategy (ES) for maximizing the likelihood, and it has shown that is a good method because it does not need initial points for the parameters, but it needs a lot of computer time to perform its searches.
From the application for State Health we can see that there are selection and moral hazard.

References

Amemiya, T. (1985). *Advanced Econometrics*. Cambridge University Press, United States.

Ardia, D., Boudt, K., Carl, P., Mullen, K., and Peterson, B. (2011). Differential evolution with deoptim. *The R Journal*, 3(1):27–34.

Babu, B. and Angira, R. (2006). Modified differential evolution (mde) for optimization of non-linear chemical processes. *Computers and Chemical Engineering*, 30:989–1002.

Cameron, C. and Trivedi, P. (2005). *Microeconometrics: Methods and Applications*. Cambridge University Press, United States.

Cardon, J. and Hendel, I. (2001). Asymmetric information in health insurance: Evidence from the national medical expenditure survey. *The RAND Journal of Economics*, 32:408–427.

Deb, P. and P.Trivedi (2006). Specification and simulated likelihood estimation of a non-normal treatment-outcome model with selection: Application to health care utilization. *Econometrics Journal*, 9:307–331.

Heckman, J. (1979). Sample selection as a specification error. *Econometrica*, 47(3):152–161.

Maddala, G. (1983). *Limited-Dependent and Qualitative Variables in Economics*. Cambridge University Press, United Kingdom.

Mayer, D., Kinghorn, B., and Archer, A. (2005). Differential evolution : an easy and efficient evolutionary algorithm for model optimization. *Agricultural Systems*, 83:315–328.

Melo, M., Stearns, S., and Norton, E. (2002). Do medicare hmos still reduce health services use after controlling for slection bias? *Health Economics*, 11:323–340.

Munkin, M. and Trivedi, P. (2008). Bayesian analysis of the ordered probit model with endegenous selection. *Journal of Econometrics*, 143:334–348.

Prieger, J. (2002). A flexible parametric selection model for non-normal data with application to health care usage. *Journal of applied econometrics*, 17:367–392.

Roy, A. (1951). Some thoughts on the distribution of earnings. *Oxford Economic Papers*, 3:135–146.

Storn, R. and Price, K. (1997). Differential evolution: A simple and efficient heuristic for global optimization over continuous spaces. *Journal of Global Optimization*, 11:341–359.

Yang, Z., Tang, K., and Yao, X. (2007). Differential evolution for high-dimensional function optimization. *IEEE Congress on Evolutionary Computation*.

6 Appendix

6.1 Appendix 1

Proof of proposition (2).

Proof Given a random variable $z \sim \mathcal{N}(0,1)$ then

$$\begin{aligned}
E(z|z>c) &= \left(\frac{1}{1-\Phi(c)}\right)\int_c^\infty z\left(\frac{1}{\sqrt{2\pi}}\exp(-z^2/2)\right)dz \\
&= \left(\frac{1}{1-\Phi(c)}\right)\int_c^\infty \frac{d}{dz}\left(\frac{-1}{\sqrt{2\pi}}\exp(-z^2/2)\right)dz \\
&= \left(\frac{1}{1-\Phi(c)}\right)\frac{-1}{\sqrt{2\pi}}\exp(-z^2/2)|_c^\infty \\
&= \frac{\phi(c)}{1-\Phi(c)}
\end{aligned}$$

Similarly,

$$\begin{aligned}
E(z^2|z>c) &= \frac{1}{1-\Phi(c)}\int_c^\infty z^2\phi(z)dz \\
&= \frac{1}{1-\Phi(c)}\int_c^\infty z^2\frac{1}{\sqrt{2\pi}}\exp(-z^2/2)dz \\
&= \frac{1}{1-\Phi(c)}\int_c^\infty z\frac{d}{dz}\left(\frac{-1}{\sqrt{2\pi}}\exp(-z^2/2)\right)dz \\
&= \frac{1}{1-\Phi(c)}\left[z\left(\frac{-1}{\sqrt{2\pi}}\right)\exp(-z^2/2)|_c^\infty - \int_c^\infty\left(-\frac{1}{\sqrt{2\pi}}\exp(-z^2/2)\right)dz\right] \\
&= \frac{1}{1-\Phi(c)}\left[c\phi(c)+(1-\Phi(c))\right] \\
&= 1+\frac{c\phi(c)}{1-\Phi(c)}
\end{aligned}$$

Thus,

$$\begin{aligned}
Var(z|z>c) &= E(z^2|z>c)-[E(z|z>c)]^2 \\
&= 1+\frac{c\phi(c)}{1-\Phi(c)}-\left(\frac{\phi(c)}{(1-\Phi(c))}\right)^2
\end{aligned}$$

□

6.2 Appendix 2

Proof of theorem (3).

Proof

$$\begin{aligned}
f(y_1=1) &= f(y_1^*>0) \\
&= f(x_1^T\beta_1+\varepsilon_1>0) \\
&= f(\varepsilon_1>-x_1^T\beta_1) \\
&= 1-f(\varepsilon_1\leqslant -x_1^T\beta_1) \\
&= 1-\Phi(-x_1^T\beta_1) \\
&= 1-(1-\Phi(x_1^T\beta_1)) \\
&= \Phi(x_1^T\beta_1)
\end{aligned}$$

And

$$\begin{aligned}
f(y_1=0) &= 1-f(y_1=1) \\
&= 1-\Phi(x_1^T\beta_1)
\end{aligned}$$

□

6.3 Appendix 3

Proof of theorems (4) and (6).

Proof

$$
\begin{aligned}
f(y_2=j|y_1=1) &= P(y_2=j|y_1=1)\\
&= P(\alpha_{2,j-1} < y_2^* \leqslant \alpha_{2,j}|y_1^* > 0)\\
&= P(\alpha_{2,j-1} < x_2^T\beta_2+\varepsilon_2 \leqslant \alpha_{2,j}|x_1^T\beta_1+\varepsilon_1 > 0)\\
&= P(\alpha_{2,j-1}-x_2^T\beta_2 < \varepsilon_2 \leqslant \alpha_{2,j}-x_2^T\beta_2|\varepsilon_1 > -x_1^T\beta_1)\\
&= P(\varepsilon_2 \leqslant \alpha_{2,j}-x_2^T\beta_2|\varepsilon_1 > -x_1^T\beta_1) - P(\varepsilon_2 \leqslant \alpha_{2,j-1}-x_2^T\beta_2|\varepsilon_1 > -x_1^T\beta_1)
\end{aligned}
$$

Where the mean and variance of the randon variable ε_2 given $\varepsilon_1 > -x_1^T\beta_1$ will be

$$
\begin{aligned}
E(\varepsilon_2|\varepsilon_1 > -x_1^T\beta_1) &= E(\sigma_{12}\varepsilon_1+\xi_2|\varepsilon_1 > -x_1^T\beta_1)\\
&= \sigma_{12}E(\varepsilon_1|\varepsilon_1 > -x_1^T\beta_1)+E(\xi_2|\varepsilon_1 > -x_1^T\beta_1)\\
&= \sigma_{12}\frac{\phi(x_1^T\beta_1)}{\Phi(x_1^T\beta_1)}
\end{aligned}
$$

And

$$
\begin{aligned}
Var(\varepsilon_2|\varepsilon_1 > -x_1^T\beta_1) &= Var(\sigma_{12}\varepsilon_1+\xi_2|\varepsilon_1 > -x_1^T\beta_1)\\
&= \sigma_{12}^2 Var(\varepsilon_1|\varepsilon_1 > -x_1^T\beta_1)+Var(\xi_2|\varepsilon_1 > -x_1^T\beta_1)+2\sigma_{12}Cov(\varepsilon_1,\xi_2|\varepsilon_1 > -x_1^T\beta_1)\\
&= \sigma_{12}^2\left[1+(-x_1^T\beta_1)\frac{\phi(-x_1^T\beta_1)}{1-\Phi(-x_1^T\beta_1)}-\left[\frac{\phi(-x_1^T\beta_1)}{1-\Phi(-x_1^T\beta_1)}\right]^2\right]+\sigma_2^2-\sigma_{12}^2\\
&= \sigma_2^2-\sigma_{12}^2\left[x_1^T\beta_1\lambda(x_1^T\beta_1)+(\lambda(x_1^T\beta_1))^2\right],
\end{aligned}
$$

where
$\lambda(x_1^T\beta_1) = \dfrac{\phi(x_1^T\beta_1)}{\Phi(x_1^T\beta_1)}$ is the inverse Mills ratio.

This implies that $\varepsilon_2|\varepsilon_1 > -x_1^T\beta_1 \sim \mathcal{N}(\sigma_{12}\lambda(x_1^T\beta_1), \sigma_2^2-\sigma_{12}^2\left[x_1^T\beta_1\lambda(x_1^T\beta_1)+(\lambda(x_1^T\beta_1))^2\right])$.

We define $\mu_{\varepsilon_2|\varepsilon_1>-x_1^T\beta_1} = \sigma_{12}\lambda(x_1^T\beta_1)$ and $\sigma^2_{\varepsilon_2|\varepsilon_1>-x_1^T\beta_1} = \sigma_2^2-\sigma_{12}^2\lambda(x_1^T\beta_1)\left[x_1^T\beta_1+\lambda(x_1^T\beta_1)\right]$.
Thus

$$
\begin{aligned}
f(y_2=j|y_1=1) &= P\left(\frac{\varepsilon_2-\mu_{\varepsilon_2|\varepsilon_1>-x_1^T\beta_1}}{\sigma_{\varepsilon_2|\varepsilon_1>-x_1^T\beta_1}} \leqslant \frac{\alpha_{2,j}-x_2^T\beta_2-\mu_{\varepsilon_2|\varepsilon_1>-x_1^T\beta_1}}{\sigma_{\varepsilon_2|\varepsilon_1>-x_1^T\beta_1}}\right)\\
&- P\left(\frac{\varepsilon_2-\mu_{\varepsilon_2|\varepsilon_1>-x_1^T\beta_1}}{\sigma_{\varepsilon_2|\varepsilon_1>-x_1^T\beta_1}} \leqslant \frac{\alpha_{2,j-1}-x_2^T\beta_2-\mu_{\varepsilon_2|\varepsilon_1>-x_1^T\beta_1}}{\sigma_{\varepsilon_2|\varepsilon_1>-x_1^T\beta_1}}\right)\\
&= \Phi\left(\frac{\alpha_{2,j}-x_2^T\beta_2-\mu_{\varepsilon_2|\varepsilon_1>-x_1^T\beta_1}}{\sigma_{\varepsilon_2|\varepsilon_1>-x_1^T\beta_1}}\right) - \Phi\left(\frac{\alpha_{2,j-1}-x_2^T\beta_2-\mu_{\varepsilon_2|\varepsilon_1>-x_1^T\beta_1}}{\sigma_{\varepsilon_2|\varepsilon_1>-x_1^T\beta_1}}\right)
\end{aligned}
$$

In a similar way can be obtained $f(y_3=k|y_1=0), f(y_3=k|y_1=1)$ and $f(y_2=j|y_1=0)$.

□

References

Amemiya, T. (1985). *Advanced Econometrics*. Cambridge University Press, United States.

Ardia, D., Boudt, K., Carl, P., Mullen, K., and Peterson, B. (2011). Differential evolution with deoptim. *The R Journal*, 3(1):27–34.

Babu, B. and Angira, R. (2006). Modified differential evolution (mde) for optimization of non-linear chemical processes. *Computers and Chemical Engineering*, 30:989–1002.

Cameron, C. and Trivedi, P. (2005). *Microeconometrics: Methods and Applications*. Cambridge University Press, United States.

Cardon, J. and Hendel, I. (2001). Asymmetric information in health insurance: Evidence from the national medical expenditure survey. *The RAND Journal of Economics*, 32:408–427.

Deb, P. and P.Trivedi (2006). Specification and simulated likelihood estimation of a non-normal treatment-outcome model with selection: Application to health care utilization. *Econometrics Journal*, 9:307–331.

Heckman, J. (1979). Sample selection as a specification error. *Econometrica*, 47(3):152–161.

Maddala, G. (1983). *Limited-Dependent and Qualitative Variables in Economics*. Cambridge University Press, United Kingdom.

Mayer, D., Kinghorn, B., and Archer, A. (2005). Differential evolution : an easy and efficient evolutionary algorithm for model optimization. *Agricultural Systems*, 83:315–328.

Melo, M., Stearns, S., and Norton, E. (2002). Do medicare hmos still reduce health services use after controlling for slection bias? *Health Economics*, 11:323–340.

Munkin, M. and Trivedi, P. (2008). Bayesian analysis of the ordered probit model with endegenous selection. *Journal of Econometrics*, 143:334–348.

Prieger, J. (2002). A flexible parametric selection model for non-normal data with application to health care usage. *Journal of applied econometrics*, 17:367–392.

Roy, A. (1951). Some thoughts on the distribution of earnings. *Oxford Economic Papers*, 3:135–146.

Storn, R. and Price, K. (1997). Differential evolution: A simple and efficient heuristic for global optimization over continuous spaces. *Journal of Global Optimization*, 11:341–359.

Yang, Z., Tang, K., and Yao, X. (2007). Differential evolution for high-dimensional function optimization. *IEEE Congress on Evolutionary Computation*.