JENA ECONOMIC RESEARCH PAPERS

2009 – 103

A Bivariate Ordered Probit Estimator with Mixed Effects

by

Franz Buscha
Anna Conte

www.jenecon.de

ISSN 1864-7057

The JENA ECONOMIC RESEARCH PAPERS is a joint publication of the Friedrich Schiller University and the Max Planck Institute of Economics, Jena, Germany. For editorial correspondence please contact markus.pasche@uni-jena.de.

Impressum:

Friedrich Schiller University Jena
Carl-Zeiss-Str. 3
D-07743 Jena
www.uni-jena.de

Max Planck Institute of Economics
Kahlaische Str. 10
D-07745 Jena
www.econ.mpg.de

A Bivariate Ordered Probit Estimator with Mixed Effects

Franz Buscha
Centre of Employment Research
University of Westminster
buschaf@wmin.ac.uk

Anna Conte
Strategic Interaction Group
Max-Planck-Institut für Ökonomik, Jena
aconte@econ.mpg.de

Abstract: In this paper, we discuss the derivation and application of a bivariate ordered probit model with mixed effects. Our approach allows one to estimate the distribution of the effect (gamma) of an endogenous ordered variable on an ordered explanatory variable. By allowing gamma to vary over the population, our estimator offers a more flexible parametric setting to recover the causal effect of an endogenous variable in an ordered choice setting. We use Monte Carlo simulations to examine the performance of the maximum likelihood estimator of our system and apply this to a relevant example from the UK education literature.[1]

JEL: C35; C51; I20
Keywords: bivariate ordered probit, maximum likelihood, mixed effects, truancy

Acknowledgements: We would like to thank Paul Latreille for his comments.

[1] Supplemental materials containing STATA code to implement our estimator are available online

1. Introduction

When a suspected endogenous explanatory variable is encountered in the applied economics literature, instrumental variable methods are often applied to estimate a causal and consistent effect. As such, instrumental variable estimation has long been a mainstay of the econometric literature and is arguably one of the most commonly used empirical methodologies. However, when both the dependent variable and the suspected endogenous variable take the form of categorical data, standard IV techniques (such as two-stage least squares) often break down and more complicated analytical techniques are required.[2]

When both the dependent variable and the endogenous variable take a binary form, a bivariate probit model can be used (Greene, 2008: 827).[3] When both the dependent variable and the endogenous variable take the form of ordered categorical data, then a bivariate ordered probit model can be applied (Greene and Hensher, 2009: 223).[4] Finally, when the dependent variable is ordered with more than two choices and the endogenous variable is binary, then a semi-ordered bivariate probit model is needed to correctly estimate the system (Greene and Hensher, 2009: 225).[5]

In this paper, we aim to add to this literature by presenting an ordered probit estimator with mixed effects. To this extent, Section 2 outlines some of the recent literature, whilst section 3 derives the estimator. Section 4 presents simulation results whilst section 5 applies our estimator to a relevant example from the UK education literature where binary or categorical answers are a frequent occurrence.

2. Literature

A variety of papers have been written which make use of a bivariate ordered probit estimator. A convenient overview of the literature is provided by Greene and Hensher (2009: 226) who highlight approximately 25 different papers making use of this methodology in a variety of circumstances from 1991 to 2007.

Likely, the first application of the bivariate ordered probit model goes back to Calhoun (1989, 1991, 1994) who provides a technical description of the estimator in addition to a computer programme for practical implementation of the estimator in Fortran (Calhoun, 1998). This was followed by two applied examples examining the relationship between desired family size and the number of children born (Calhoun, 1991; 1994). More recent examples of published papers making use of a bivariate ordered probit estimator include Dawson and Dobson (2009), who examine the role that a variety of factors play on the home vs. away team discipline scores, and Kawakatsu and Largey (2009) who propose an EM algorithm for bivariate ordered probit models with endogenous regressors instead of direct numerical

---

[2] For a more nuanced argument see Angrist's (2001) discussion of limited dependent variable models with dummy endogenous regressors.

[3] These can be extended to the multivariate case when more than one binary endogenour regressor is on the 'right-hand-side' of the equation. See Cappellari and Jenkins (2003, 2006). Implemented as a Stata routine `mvprobit` by Cappellari and Jenkins (2003).

[4] Implemented as a Stata routine `bioprobit` by Sajaia (2008)

[5] It should be noted that the semi-ordered bivariate probit estimator is a special case of the bivariate ordered probit estimator and does not require special modifications to the likelihood function. However, for expositional purposes we will highlight examples in our paper using both types of estimators.

optimization methods. Papers which make use of a semi-ordered bivariate probit approach include studies by Weiss (1993), Armstrong and McVicar (2000), McVicar and McKee (2002) and Ramanna (2008). However, not all applications of a bivariate ordered probit model are the result of a suspected endogenous regressor in explaining an ordered response. Sometimes the seemingly unrelated specification is used to improve estimation efficiency.

Until recently, the practical hurdle in implementing such a bivariate ordered probit estimator has been quite steep. Perhaps partially because the bivariate ordered probit model has found little exposition in econometric textbooks and maybe because the computational hurdle of programming one's own likelihood is relatively complex. Thankfully, both LIMDEP and Stata now support the estimation of such bivariate ordered choice models. For Stata a good description of the bivariate ordered probit estimator and its practical implementation is provided by Sajaia (2008) who describes the bivariate ordered probit estimator and introduces a routine which enables user's a relatively easy practical implementation of ordered-ordered and semi-ordered models. The programme can be used both as a SURE estimator and as a recursive systems estimator with one endogenous variable.

Within this context we want to contribute to the literature by using methodology emerging in the behavioural economics literature that aims at estimating the distribution over the population of the relevant parameters of the model under investigation instead of just reporting a point estimate of these parameters. See for example, Botti *et al.* (2008) and Conte *et al.* (2009). With this in mind, we work out a modified version of a bivariate ordered probit (semi-ordered bivariate probit) that allows the effect of an endogenous ordered (bivariate) variable to be heterogeneous across the population. In other words, we assume that, *ceteris paribus*, the effect of the endogenous variable can differ individual by individual and that such differences are captured by an appropriate choice of the distribution function for this effect.

3. Estimator

Assume that two latent variables y_{1i}^* and y_{2i}^* are determined by the following system of equations:

$$\begin{cases} y_{1i}^* = x_{1i}'\beta_1 + \varepsilon_{1i} \\ y_{2i}^* = \gamma_i y_{1i}^* + x_{2i}'\beta_2 + \varepsilon_{2i} \end{cases}, \text{ with } \gamma_i \sim N(\mu_\gamma, \sigma_\gamma^2) \text{ and } \begin{pmatrix} \varepsilon_{1i} \\ \varepsilon_{2i} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right). \quad (1)$$

Here, x_{1i} and x_{2i} are vectors of observables, β_1 and β_2 are a vector of parameters, γ_i is a scalar representing the effect that y_{1i}^* has on y_{2i}^* for individual *i*, and ε_{1i} and ε_{2i} are two error terms, assumed to be jointly normal with correlation coefficient ρ and uncorrelated with everything else in the model; in particular, $E(x_{1i}\varepsilon_{1i}) = 0$ and $E(x_{2i}\varepsilon_{2i}) = 0$.[6]

The basic idea underlying our model is that when trying to explain people's choices at stake are both observed factors, represented by x_{1i} and x_{2i} (for example demographic variables) and unobserved

[6] The derivation of the reduced form of the system in eq. (1) follows in the Appendix.

factors (for example motivation) embedded in γ_i and in the joint distribution of ε_{1i} and ε_{2i}. In this model, we assume that there is heterogeneity between individuals in the way the latent variable y_{1i}^* influences y_{2i}^*. To capture such an effect we adopt a continuous mixture approach by estimating the distribution of this effect over the population. In other words, we assume that each individual draws their own γ_i from a distribution and what we do here is to estimate the parameters of the underlying distribution of γ_i.

However, we do not observe the realisation of the two latent variables y_{1i}^* and y_{2i}^*. What we observe, instead, are the two categorical variables y_{1i} and y_{2i}. These are respectively linked to y_{1i}^* and y_{2i}^* by the following observational rules:

$$y_{1i} = \begin{cases} 1 & if\ y_{1i}^* \leq j_{11} \\ \vdots & \vdots \\ l & if\ j_{1l-1} < y_{1i}^* \leq j_{1l} \\ \vdots & \vdots \\ L & if\ j_{1L-1} < y_{1i}^* \end{cases} \qquad y_{2i} = \begin{cases} 1 & if\ y_{2i}^* \leq j_{21} \\ \vdots & \vdots \\ m & if\ j_{2m-1} < y_{2i}^* \leq j_{2m}, \\ \vdots & \vdots \\ M & if\ j_{2M-1} < y_{2i}^* \end{cases} \tag{2}$$

where the $j_{..}$ are cut-points to be estimated along with the other parameters of the model.[7]

The probability of observing $y_{1i} = l$ and $y_{2i} = m$ for individual *i* is:

$$\begin{aligned} Pr(y_{1i} = l, y_{2i} = m) \\ &= \Phi(j_{1l} - x'_{1i}\beta_1, (j_{2m} - \gamma_i x'_{1i}\beta_1 - x'_{2i}\beta_2)\lambda_i, \tilde{\rho}_i) \\ &- \Phi(j_{1l-1} - x'_{1i}\beta_1, (j_{2m} - \gamma_i x'_{1i}\beta_1 - x'_{2i}\beta_2)\lambda_i, \tilde{\rho}_i) \\ &- \Phi(j_{1l} - x'_{1i}\beta_1, (j_{2m-1} - \gamma_i x'_{1i}\beta_1 - x'_{2i}\beta_2)\lambda_i, \tilde{\rho}_i) \\ &+ \Phi(j_{1l-1} - x'_{1i}\beta_1, (j_{2m-1} - \gamma_i x'_{1i}\beta_1 - x'_{2i}\beta_2)\lambda_i, \tilde{\rho}_i) \end{aligned} \tag{3}$$

where $\Phi(.,.,.)$ is the bivariate standard normal cumulative distribution function and λ_i and $\tilde{\rho}_i$ are respectively defined as follows: $\lambda_i = 1/\sqrt{\gamma_i^2 + 2\gamma_i\rho + 1}$ and $\tilde{\rho}_i = \lambda_i(\gamma_i + \rho)$.

The log-likelihood contribution of individual *i* is:

$$L_i = \int_{-\infty}^{\infty} \prod_{l=1}^{L} \prod_{m=1}^{M} Pr(y_{1i} = l, y_{2i} = m)^{I(y_{1i}=l, y_{2i}=m)} f(\gamma_i; \mu_\gamma, \sigma_\gamma)\, d\gamma_i\ , \tag{4}$$

[7] The cut-points meet the following conditions: $j_{10} < \cdots < j_{1l} < \cdots < j_{1L}$, with $j_{10} = -\infty$ and $j_{1L} = \infty$; $j_{20} < \cdots < j_{2m} < \cdots < j_{2M}$, with $j_{20} = -\infty$ and $j_{2M} = \infty$.

where $f(\gamma_i; \mu_\gamma, \sigma_\gamma)$ is the normal density function for the random variable γ_i, and $I(y_{1i} = l, y_{2i} = m)$ is an indicator function that equals one when $y_{1i} = l$ and $y_{2i} = m$.

The sample log-likelihood function $lnL = \sum_{i=1}^{N} lnL_i$ is maximised using 20-point Gauss-Hermite quadrature. The program is written in STATA version 11.0, and is available from the authors on request.

4. Monte Carlo simulations

To examine the small sample properties of our estimator we implement Monte Carlo simulations. We generate x_{1i} and z_{1i} as independent standard normal random variables and ε_{1i} and ε_{2i} as standard normal random variables with correlation ρ. Moreover, we simulate semi-ordered and ordered-ordered conditions of the bivariate ordered probit model with mixed effects. The latent variables y_{1i}^* and y_{2i}^* for the semi-ordered model are generated by the following process:

$$\begin{aligned} y_{1i}^* &= -0.5 + 1x_{1i} + 1z_{1i} + \varepsilon_{1i} \\ y_{2i}^* &= \gamma_i y_{1i}^* - 2.5x_{1i} + \varepsilon_{2i} \end{aligned} \tag{5}$$

Where $\gamma_i \sim N(\mu_\gamma = 0.5, \sigma_\gamma = 0.5)$ and subject to the observational rule:

$$y_{1i} = \begin{cases} 0 = y_{1i}^* \leq 0 \\ 1 = y_{1i}^* > 0 \end{cases} \qquad y_{2i} = \begin{cases} 1 = y_{2i}^* \leq -2 \\ 2 = -1 < y_{2i}^* \leq -1 \\ 3 = -1 < y_{2i}^* \leq 0 \\ 4 = 0 < y_{2i}^* \leq 1 \\ 5 = 1 < y_{2i}^* \end{cases} \tag{6}$$

Data for the ordered-ordered model is subject to the following data generating process:

$$\begin{aligned} y_{1i}^* &= 1x_{1i} + 1z_{1i} + \varepsilon_{1i} \\ y_{2i}^* &= \gamma_i y_{1i}^* - 2.5x_{1i} + \varepsilon_{2i} \end{aligned} \tag{7}$$

Where $\gamma_i \sim N(\mu_\gamma = 0.5, \sigma_\gamma = 0.5)$ and the observational rule is:

$$y_{1i} = \begin{cases} 1 = y_{1i}^* \leq -2 \\ 2 = -2 < y_{1i}^* \leq 0.5 \\ 3 = 0.5 < y_{1i}^* \end{cases} \qquad y_{2i} = \begin{cases} 1 = y_{2i}^* \leq -3 \\ 2 = -3 < y_{2i}^* \leq -1 \\ 3 = -1 < y_{2i}^* \leq 1 \\ 4 = 1 < y_{2i}^* \end{cases} \tag{8}$$

We run 1000 replications for values of $\rho = \{-0.9, -0.5, 0, 0.5, 0.9\}$ for observations, $Obs = \{200, 500, 1000, 5000\}$. We report the sample mean estimates of μ_γ, σ_γ and ρ in addition to reporting the root mean square error (RMSE). Our simulation results indicate that our mixed estimator performs well in recovering the true parameters, even in small samples. The additional categories in an ordered-ordered model appear to reduce the RMSE marginally.

Semi-ordered bivariate probit

Obs = 200

Rho	μ_y	RMSE	σ_y	RMSE	ρ	RMSE
-0.9	0.4921	0.0741	0.4802	0.1130	-0.8991	0.0488
-0.5	0.5044	0.1094	0.5184	0.1838	-0.5011	0.1365
0	0.5228	0.1412	0.5515	0.2223	0.0095	0.1676
0.5	0.5393	0.1660	0.5524	0.2178	0.5209	0.1436
0.9	0.5214	0.1592	0.4900	0.1691	0.9030	0.0482

Obs = 500

Rho	μ_y	RMSE	σ_y	RMSE	ρ	RMSE
-0.9	0.4974	0.0486	0.4982	0.0693	-0.9004	0.0306
-0.5	0.5019	0.0656	0.5061	0.1149	-0.5007	0.0821
0	0.5063	0.0822	0.5080	0.1257	0.0034	0.1016
0.5	0.5080	0.0941	0.5070	0.1247	0.5064	0.0852
0.9	0.5084	0.0971	0.5025	0.1078	0.9073	0.0343

Obs = 1000

Rho	μ_y	RMSE	σ_y	RMSE	ρ	RMSE
-0.9	0.4985	0.0315	0.4937	0.0488	-0.8997	0.0211
-0.5	0.5005	0.0444	0.4949	0.0726	-0.5010	0.0563
0	0.5039	0.0550	0.4972	0.0842	-0.0007	0.0735
0.5	0.5050	0.0606	0.4976	0.0888	0.4993	0.0616
0.9	0.5067	0.0645	0.5006	0.0763	0.9054	0.0239

Obs = 5000

Rho	μ_y	RMSE	σ_y	RMSE	ρ	RMSE
-0.9	0.4995	0.0150	0.4965	0.0205	-0.8993	0.0094
-0.5	0.4994	0.0205	0.4976	0.0320	-0.4993	0.0265
0	0.5003	0.0256	0.4978	0.0378	-0.0002	0.0324
0.5	0.5007	0.0284	0.4997	0.0366	0.5010	0.0254
0.9	0.4999	0.0284	0.4971	0.0321	0.9007	0.0099

Ordered-ordered bivariate probit

Obs = 200

Rho	μ_y	RMSE	σ_y	RMSE	ρ	RMSE
-0.9	0.4971	0.0729	0.4723	0.1082	-0.8943	0.0502
-0.5	0.5092	0.1000	0.5121	0.1599	-0.5015	0.1166
0	0.5346	0.1455	0.5647	0.2184	0.0124	0.1518
0.5	0.5256	0.1489	0.5254	0.1781	0.5259	0.1387
0.9	0.5006	0.1504	0.4589	0.1387	0.8823	0.0591

Obs = 500

Rho	μ_y	RMSE	σ_y	RMSE	ρ	RMSE
-0.9	0.4997	0.0436	0.4994	0.0647	-0.9021	0.0324
-0.5	0.5043	0.0587	0.5068	0.0915	-0.5029	0.0721
0	0.5067	0.0757	0.5084	0.1091	0.0015	0.0964
0.5	0.5116	0.0924	0.5104	0.1066	0.5081	0.0785
0.9	0.4965	0.0928	0.4858	0.0845	0.9008	0.0404

Obs = 1000

Rho	μ_y	RMSE	σ_y	RMSE	ρ	RMSE
-0.9	0.5007	0.0324	0.5018	0.0468	-0.9033	0.0234
-0.5	0.5022	0.0434	0.4997	0.0642	-0.5027	0.0517
0	0.5043	0.0528	0.4998	0.0760	-0.0030	0.0656
0.5	0.5047	0.0590	0.4994	0.0750	0.5001	0.0575
0.9	0.5066	0.0666	0.5004	0.0610	0.9029	0.0321

Obs = 5000

Rho	μ_y	RMSE	σ_y	RMSE	ρ	RMSE
-0.9	0.4996	0.0140	0.4979	0.0198	-0.8997	0.0098
-0.5	0.4995	0.0186	0.4976	0.0281	-0.4996	0.0230
0	0.4998	0.0243	0.4980	0.0341	-0.0005	0.0288
0.5	0.5008	0.0278	0.4989	0.0327	0.4991	0.0248
0.9	0.5009	0.0292	0.5000	0.0276	0.9016	0.0138

Monte Carlo Simulations: 1000 Replications

5. Empirical example

Choices made in the early years of an individual's life have long-lasting effects into adulthood and one of the most important decisions undertaken by young people in the UK is at the age of 16. This is when compulsory education ends and students sit General Certificate of Secondary Education examinations (GCSE). Those who receive low educational attainment scores are more likely to experience lower wages, higher unemployment and, in general, are likely to experience a lower quality of life (Bradley and Nguyen, 2004). One major factor in determining educational outcomes is truancy which has been identified as a strong predictor of low educational attainment (Bosworth, 1994) and of 'poor life outcomes' (Hibbert and Fogelman, 1990). In the UK unauthorised school absence before the age of sixteen is illegal; however, official truancy statistics generally make for grim reading with the most recent truancy statistics reporting a record rate of school sessions being missed (1.03% - DCSF, 2009). Moreover, the long term trend does not look favourable (rising truancy rate) even though the UK government has spent over £885 million on anti-truancy policies during the period 1997/98 to 2003/04. A report by the Select Committee on Public Accounts (2006) estimated that the cost of absent pupils was £1.6 billion in missed education in the school year 2003/04. Clearly, studies which examine the determinants of truancy and the impact that truancy may have on educational outcomes can be considered important given such a social and political context.

One of the main data sources for the UK education literature which examines the schooling experiences of 16 to 18 year olds (and contains detailed records on truancy and educational attainment) is the Youth Cohort Study of England and Wales (YCS). The YCS is specifically designed to monitor the behaviour and decisions made by a representative sample of the UK school population as they transit from compulsory education (age 16) to further education and higher education, or to the labour market. The YCS data is a longitudinal dataset designed to follow individuals over the course of 3 years (3 sweeps starting at age 16) and now has eleven cohorts with nearly thirty sweeps. However, in this study we make use of only the first sweep as truancy information is not available in later sweeps. Moreover, we make use of the restricted version of the dataset which enables us to map local economic conditions to individuals (we have this information for YCS 11). Although the YCS data uses a multi-stage stratified random sampling procedure, differences in selection and response rates may still be an issue. We therefore make use of included weights which correct for differential selection probabilities, correct the ethnicity boost, and take into account non-response bias.

The YCS records truancy and educational attainment as ordered variables. Moreover, the measuring of ordered educational outcomes is relatively common in the UK as the UK does not use grade point averages like the United States, but instead relies on categorical targets – for example in 2008 the DCSF set a national target of 60% of pupils achieving 5 or more A*-C grades at GCSE. This is reflected in the data sources whereby the YCS survey 11 (pupils eligible to leave school in 2000-01) returns the following descriptive statistics:

Table 2: Truancy and Educational Attainment in the Youth Cohort Study 11, Sweep 1.

Truancy	*Frequency*	*Percent*	*Educational Attainment at 16*	*Frequency*	*Percent*
Never	10,922	65.93	None	615	3.68
For the odd day or lesson	4,226	25.51	1-4 GCSE D-G grade	215	1.29
Particular days or lessons	867	5.23	5+ GCSE D-G grade	1,329	7.95
For several days at a time	289	1.74	1-4 GCSE A*-C grade	3,722	22.28
For weeks at a time	262	1.58	5+ GCSE A*-C grade	10,826	64.80
Total	16,566	100	Total	16,707	100

However, given that truancy is an individual choice, likely to be motivated by a series of unobserved characteristics, it is likely that truancy could be considered to be an endogenous variable in an educational attainment regression. Estimation techniques which do not account for the potential endogeneity of truancy may thus provide consistent estimates. In this example we treat truancy as an endogenous variable and use the previously outlined bivariate ordered probit estimator with mixed effects in both a semi-ordered and ordered-ordered fashion to estimate the causal impact of truancy on educational attainment. The economic model we estimate argues that our model consists of two unobserved latent factors which are determined by a series of exogenous characteristics (x_{1i} and x_{2i}). Furthermore, one of the latent factors is determined by the other latent factor:

$$\begin{aligned} T_i^* &= x_{1i}'\beta_1 + z_2'\beta_2 + \varepsilon_1 \\ E_i^* &= x_{2i}'\beta_2 + \gamma_i T_i^* + \varepsilon_2 \end{aligned} \tag{9}$$

The unobserved latent variables T_i^* (truancy) and E_i^* (educational outcomes) are related to their respective observed outcomes as follows:

$$Truancy_i = \begin{cases} 1 = \text{Never} & \text{if} \quad T_i < c_{11} \\ 2 = \text{Odd day} & \text{if} \quad c_{11} < T_i \le c_{12} \\ 3 = \text{Particular day} & \text{if} \quad c_{12} < T_i \le c_{13} \\ 4 = \text{Several days} & \text{if} \quad c_{13} < T_i \le c_{14} \\ 5 = \text{Weeks at a time} & \text{if} \quad c_{14} < T_i \end{cases} \qquad Education_i = \begin{cases} 1 = \text{None} & \text{if} \quad E_i < c_{21} \\ 2 = \text{1-4 GCSE D-G} & \text{if} \quad c_{21} < E_i \le c_{22} \\ 3 = \text{5+ GCSE D-G} & \text{if} \quad c_{22} < E_i \le c_{23} \\ 4 = \text{1-4 GCSE A*-C} & \text{if} \quad c_{23} < E_i \le c_{24} \\ 5 = \text{5+ GCSE A*-C} & \text{if} \quad c_{24} < E_i \end{cases} \tag{10}$$

If ε_1 and ε_2 are distributed as bivariate standard normal with correlation ρ, and γ_i is a mixed effect which is distributed standard normal with mean μ_γ and standard deviation σ_γ over the sample of individuals *i*, then the above system can be estimated using a bivariate ordered probit model with mixed effects. The assumption that $\gamma_i \sim N(\mu_\gamma, \sigma_\gamma)$ is made *a priori* as we assume that the impact that truancy has on an individual's attainment may be positively or negatively distributed around an unknown mean. Moreover, there is little reason to assume that this distribution is asymmetric around the mean. Finally we assume that there is 'clumping' near the mean and few individuals experience extreme positive or negative effects. The standard normal distribution suits these assumptions well. If the true effect of truancy is similar for every individual *i* then σ_γ will be estimated as 0 and the estimator collapses into a 'standard' bivariate ordered probit estimator. The distributional assumption of γ_i must be made on a case by case basis as other relationships may require different distributions. For example, one could imagine that the impact health-checkups on length of survival will always have positive effects which may be distributed with an exponential decay and thereby warrant a power function.

In addition to using a vector of standard exogenous controls such as school type, ethnicity, parental education, gender, housing and disability status (Bosworth, 1994; Bradley and Taylor, 2004) we also make use of several instruments to help identify the recursive system of equations. In particular we make use of the fact that we have local authority district codes available in our data (restricted version) and use this to match in a variety of local labour market characteristics from NOMIS.[8] Specifically we make use of the median local part-time pay as there is a there is a literature which argues that truancy, rather than being an irrational phenomenon with little causal explanation, has strong roots in rational choice by individuals who maximise their expected pay-offs as they change their behaviour in response to appropriate economic incentives Burgess *et al.* (2002). Therefore, there is likely to be a correlation between truancy and local labour market conditions. However, at the same time, economic incentives such as local unemployment or local wage rates are also likely to affect educational decisions. This is because high local wages or low local unemployment change the opportunity cost of an additional year of schooling (and hence the educational outcomes).

Nonetheless, we argue that truancy is likely to be related to local *part-time* wages since by law under-16s must attend school until they are 16. Full-time employment is thus not a viable option for truanting under-16 year olds whilst part-time work and part-time pay are viable options and should hence influence the truanting decision. Dustman *et al.* (1997), for example, show that part-time working is a strong predictor of truancy in the UK. Conversely, post-school aspirations, and therefore educational outcomes, are likely to be related to local *full-time* wages and local unemployment, rather than local part-time wages. We argue that young people who are looking ahead and wish to determine their educational investment are likely to be influenced by 'stronger' economic indicators that local part-time pay. This line of reasoning allows us to use local part-time pay as an exogenous instrument which is uncorrelated with educational attainment. Using our instrument in independent (naïve) ordered probit regressions shows that the local median part-time pay has a positive and statistically significant effect on truancy (higher local part-time wages results in higher truancy) whilst it has a statistically insignificant effect on

[8] A service provided by the Office for National Statistic which gives free access to detailed UK labour market statistics. See www.nomisweb.co.uk

educational outcomes.[9] We thus argue that the right conditions for our instrument (correlated with truancy, uncorrelated with education) are fulfilled.

For expositional purposes we split the sample into two groups – those from a high parental socio-economic background (higher professionals, lower professionals, higher technical and intermediate occupations) and those from a low parental socio-economic background (lower supervisory, semi-routine, routine and other occupations). Moreover we estimate the semi-ordered case and the ordered-ordered case to highlight the difference between truancy and no truancy vs. different intensities of truancy, as measured in Table 1. Finally, for comparison purposes we provide estimates of the same model specification using the more restrictive assumptions of the 'standard' bivariate ordered probit, which constrains γ_i to a singular value. Table 3 present the estimates from the 'standard' model whilst Table 4 presents the estimates from the model with mixed effects.

[9] These are available upon request.

Table 3: Semi-Ordered and Ordered-Ordered Bivariate Probit Estimation ***without*** *Mixed Effects*

Variables		Semi-ordered bivariate probit						Ordered-ordered bivariate probit					
		Low SEC			High SEC			Low SEC			High SEC		
Dependent	Independent	Coeff	S.E	P-val	Coeff	S.E	P-val	Coeff	S.E	P-val	Coeff	S.E	P-val
Truancy	Female	0.116	0.042	0.005	0.051	0.029	0.078	0.144	0.038	0.000	0.037	0.028	0.186
	Selective school	-0.317	0.166	0.056	-0.265	0.059	0.000	-0.381	0.140	0.006	-0.299	0.055	0.000
	Independent school	-0.528	0.184	0.004	-0.434	0.055	0.000	-0.591	0.160	0.000	-0.463	0.051	0.000
	Ethnicity white	0.079	0.056	0.160	-0.049	0.048	0.305	0.089	0.052	0.086	-0.036	0.045	0.421
	Parents have one A-level	0.054	0.063	0.392	-0.086	0.037	0.019	0.060	0.057	0.294	-0.098	0.036	0.006
	Parents have one degree	0.227	0.080	0.004	-0.077	0.034	0.023	0.211	0.068	0.002	-0.076	0.033	0.022
	Council house	0.280	0.048	0.000	0.461	0.065	0.000	0.335	0.044	0.000	0.507	0.060	0.000
	Rented house	0.272	0.071	0.000	0.378	0.068	0.000	0.303	0.063	0.000	0.386	0.065	0.000
	Other house	0.615	0.179	0.001	0.179	0.129	0.164	0.828	0.181	0.000	0.287	0.124	0.021
	Disability	0.240	0.102	0.019	0.076	0.074	0.303	0.342	0.096	0.000	0.126	0.074	0.087
	Median local pay	0.278	0.196	0.157	0.484	0.151	0.001	0.249	0.178	0.161	0.451	0.154	0.003
	Had part-time job	0.263	0.045	0.000	0.332	0.032	0.000	0.210	0.041	0.000	0.325	0.031	0.000
	constant/cut11	1.917	0.931	0.040	2.865	0.714	0.000	1.792	0.842	0.033	2.703	0.728	0.000
	cut12							2.720	0.842	0.001	3.694	0.729	0.000
	cut13							3.178	0.841	0.000	4.205	0.733	0.000
	cut14							3.531	0.842	0.000	4.553	0.735	0.000
Education	Female	0.225	0.046	0.000	0.319	0.033	0.000	0.205	0.055	0.000	0.320	0.034	0.000
	Selective school	1.510	0.352	0.000	1.330	0.193	0.000	1.478	0.341	0.000	1.320	0.193	0.000
	Independent school	1.096	0.336	0.001	0.502	0.108	0.000	1.095	0.326	0.001	0.492	0.109	0.000
	Ethnicity white	-0.173	0.045	0.000	-0.060	0.056	0.288	-0.173	0.045	0.000	-0.056	0.057	0.326
	Parents have one A-level	0.187	0.059	0.002	0.278	0.040	0.000	0.177	0.062	0.004	0.278	0.040	0.000
	Parents have one degree	0.022	0.077	0.780	0.439	0.040	0.000	0.013	0.077	0.871	0.444	0.040	0.000
	Council house	-0.509	0.038	0.000	-0.643	0.067	0.000	-0.520	0.038	0.000	-0.639	0.069	0.000
	Rented house	-0.480	0.058	0.000	-0.359	0.078	0.000	-0.485	0.058	0.000	-0.359	0.077	0.000
	Other house	-0.734	0.154	0.000	-0.414	0.110	0.000	-0.796	0.168	0.000	-0.390	0.112	0.000
	Disability	-0.414	0.076	0.000	-0.454	0.075	0.000	-0.436	0.077	0.000	-0.448	0.076	0.000
	cut21	-1.333	0.312	0.000	-2.388	0.318	0.000	-1.261	0.354	0.000	-2.398	0.328	0.000
	cut22	-1.056	0.300	0.000	-2.166	0.315	0.000	-0.989	0.336	0.003	-2.170	0.324	0.000
	cut23	-0.144	0.262	0.582	-1.345	0.307	0.000	-0.105	0.281	0.708	-1.334	0.313	0.000
	cut24	0.665	0.232	0.004	-0.581	0.301	0.054	0.672	0.242	0.005	-0.562	0.304	0.065
	γ	0.209	0.122	0.086	-0.229	0.102	0.024	0.251	0.144	0.081	-0.238	0.106	0.025
	ρ	-0.457	0.108	0.000	-0.072	0.101	0.474	-0.534	0.119	0.000	-0.117	0.104	0.263
	Log pseudolikelihood		-10428.41			-15081.626			-12481.075			-17503.258	
	N		4684			10356			4684			10356	

Source: Youth Cohort Study of England and Wales 11 (2001/2002), weighted analysis

Table 4: Semi-Ordered and Ordered-Ordered Bivariate Probit Estimation ***with*** *Mixed Effects*

Variables		Semi-ordered bivariate probit						Ordered-ordered bivariate probit					
		Low SEC			High SEC			Low SEC			High SEC		
Dependent	Independent	Coeff	S.E	P-val	Coeff	S.E	P-val	Coeff	S.E	P-val	Coeff	S.E	P-val
Truancy	Female	0.112	0.039	0.004	0.055	0.029	0.060	0.113	0.039	0.003	0.038	0.029	0.188
	Selective school	-0.373	0.208	0.072	-0.288	0.065	0.000	-0.426	0.182	0.019	-0.292	0.060	0.000
	Independent school	-0.351	0.241	0.145	-0.491	0.065	0.000	-0.519	0.170	0.002	-0.500	0.060	0.000
	Ethnicity white	0.068	0.053	0.199	-0.023	0.050	0.651	0.070	0.050	0.167	-0.061	0.046	0.184
	Parents have one A-level	0.075	0.060	0.212	-0.078	0.037	0.036	0.067	0.058	0.245	-0.102	0.036	0.005
	Parents have one degree	0.260	0.077	0.001	-0.055	0.035	0.116	0.218	0.068	0.001	-0.069	0.034	0.046
	Council house	0.305	0.045	0.000	0.473	0.062	0.000	0.340	0.044	0.000	0.514	0.060	0.000
	Rented house	0.286	0.068	0.000	0.354	0.071	0.000	0.312	0.065	0.000	0.392	0.069	0.000
	Other house	0.636	0.162	0.000	0.213	0.122	0.080	0.870	0.181	0.000	0.329	0.123	0.007
	Disability	0.189	0.093	0.042	0.042	0.074	0.570	0.319	0.098	0.001	0.105	0.076	0.170
	Log median local pay	0.084	0.174	0.629	0.554	0.154	0.000	0.139	0.020	0.000	0.149	0.025	0.000
	Had part-job	0.256	0.042	0.000	0.325	0.033	0.000	0.205	0.041	0.000	0.314	0.031	0.000
	constant/cut11	-0.996	0.825	0.227	-3.218	0.728	0.000	-0.051	0.087	0.556	-0.130	0.112	0.245
	cut12							0.891	0.090	0.000	0.871	0.115	0.000
	cut13							1.367	0.091	0.000	1.394	0.113	0.000
	cut14							1.758	0.093	0.000	1.792	0.117	0.000
Education	Female	0.353	0.095	0.000	0.508	0.052	0.000	0.260	0.067	0.000	0.410	0.044	0.000
	Selective school	2.528	0.621	0.000	2.579	0.367	0.000	1.832	0.369	0.000	1.883	0.240	0.000
	Independent school	1.886	0.497	0.000	1.089	0.183	0.000	1.234	0.414	0.003	0.859	0.154	0.000
	Ethnicity white	-0.297	0.071	0.000	-0.113	0.079	0.153	-0.210	0.051	0.000	-0.066	0.068	0.329
	Parents have one A-level	0.269	0.102	0.008	0.418	0.064	0.000	0.199	0.072	0.006	0.354	0.052	0.000
	Parents have one degree	0.003	0.117	0.978	0.703	0.069	0.000	0.006	0.089	0.946	0.568	0.059	0.000
	Council house	-0.712	0.075	0.000	-0.888	0.125	0.000	-0.576	0.046	0.000	-0.770	0.105	0.000
	Rented house	-0.656	0.092	0.000	-0.440	0.119	0.000	-0.531	0.064	0.000	-0.391	0.097	0.000
	Other house	-1.167	0.257	0.000	-0.625	0.173	0.000	-0.890	0.183	0.000	-0.483	0.147	0.001
	Disability	-0.607	0.145	0.000	-0.697	0.124	0.000	-0.493	0.091	0.000	-0.570	0.103	0.000
	cut21	-2.643	0.327	0.000	-2.955	0.228	0.000	-1.983	0.148	0.000	-2.293	0.199	0.000
	cut22	-2.104	0.236	0.000	-2.420	0.191	0.000	-1.655	0.111	0.000	-1.940	0.171	0.000
	cut23	-0.776	0.077	0.000	-0.922	0.119	0.000	-0.651	0.074	0.000	-0.787	0.124	0.000
	cut24	0.285	0.162	0.078	0.232	0.105	0.027	0.215	0.150	0.151	0.194	0.116	0.094
	μ_v	0.276	0.172	0.110	-0.726	0.157	0.000	0.260	0.168	0.123	-0.480	0.149	0.001
	σ_v	1.284	0.240	0.000	1.369	0.144	0.000	0.551	0.115	0.000	0.867	0.134	0.000
	ρ	-0.704	0.120	0.000	0.101	0.153	0.510	-0.577	0.132	0.000	-0.038	0.144	0.793
	Log pseudolikelihood		-10395.465			-14990.441			-8709.142			-10662.568	
	N		4684			10356			4684			10356	

Source: Youth Cohort Study of England and Wales 11 (2001/2002), weighted analysis

Before discussing the results of the bivariate ordered probit with mixed effects, a quick comparison between the 'standard' bivariate ordered probit model and the 'mixed' bivariate ordered probit model suggests that, especially for the ordered-ordered case, the mixed effects model significantly reduces the estimated log-likelihood. Moreover, the standard error of the cut-points has been reduced significantly in the ordered-ordered case. Finally, the estimated values of γ_i in the standard model are comparable to the estimates values of μ_γ in the mixed model.

Results for the semi-ordered and ordered-ordered estimator with mixed effects indicates that the decision truant is more complex than would assumed by a 'standard' bivariate ordered probit model (where one mean effect is estimated for every individual i in the sample). Our results suggest that there is considerably heterogeneity in the impact of truancy on educational outcomes. Both the semi-ordered and ordered-ordered model provide a similar estimate of μ_γ with pupils from disadvantaged socio-economic backgrounds experiencing a mean positive effect of truancy on educational attainment (0.28 and 0.26) whilst pupils from advantaged socio-economic backgrounds experience a mean negative effects (-0.73 or -0.48 depending on the model). The means between both groups are statistically insignificant from each other. Estimates of σ_γ, however, vary substantially by the type of model used (semi-ordered or ordered-ordered). Estimates of the semi-ordered estimator suggest a standard deviation of around 1.3 for both groups whilst estimates from the ordered-ordered model suggest much smaller standard deviations of 0.55 and 0.87 for disadvantaged and advantaged socio-economic groups respectively. The additional information provided in the ordered-ordered model thus substantially reduces the variance of the heterogeneous effect of truancy. Finally, estimates of ρ suggests that truancy is likely to be an endogenous regressor for low socio-economic pupils, whilst not endogenous for pupils from the high socio-economic group. The negative sign on the correlation between truancy and education suggests that there is a tendency over the population of students from low socio-economic background to have lower educational levels the more they truant.[10]

The information provided by our regression estimates can be used to plot the above information in a graphical form, given by Figure 1 and Figure 2. They highlight the differential effect that truancy may have on different groups of individuals with respect to educational outcomes.

[10] We would like to stress the difference between ρ and γ_i – rho measures the correlation between the error terms in the two regressions. The error terms are compromised of any unobservable characteristics which we do not observe and any statistically significant correlation between the two error terms is indicative that that the exogeneity condition of y_1 in regression y_2 cannot be accepted. This is because error term ε_1 now enters the second equation and if it is correlated with error term ε_2, the expected value of both error terms will not be zero, $E(\varepsilon_1, \varepsilon_2) \neq (0,0)$. This violates the zero-conditional mean assumption of the error term. Gamma (γ_i) estimates the effect that y_1 has on y_2 and should not be mistaken for endogeneity. Gamma may influence the estimated value of rho by correctly ascribing a more nuanced impact of y_1 on y_2, and hence reducing the amount of 'unobservables' in the error term. This may, in some instances lead to insignificant error correlation and thus 'eliminate' endogeneity. However, such an event is only possible if the only model misspecification is the mixed effect of gamma and all other unobservable effects are accounted for – a generally unlikely scenario.

Figure 1

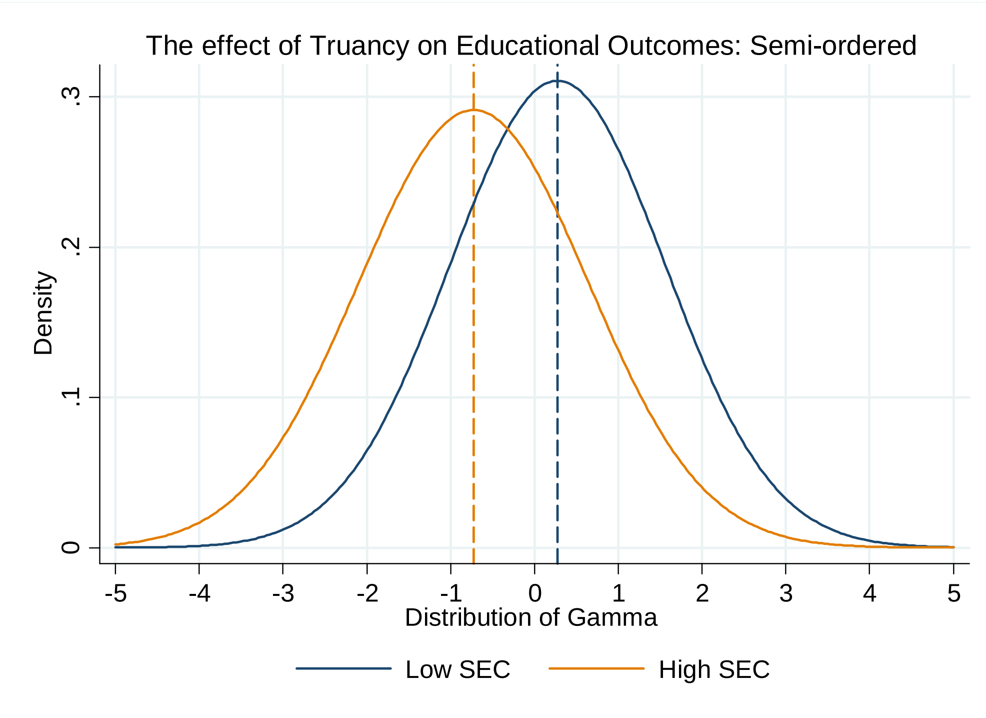
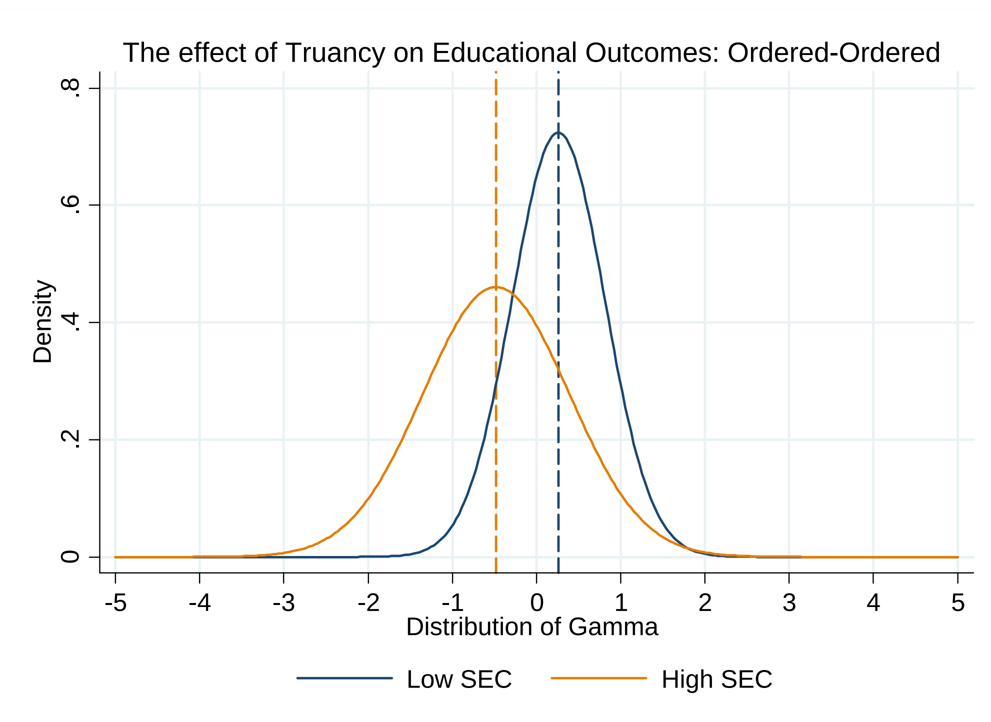


Figure 2



At first the results of our analysis may seem somewhat counterintuitive as we predict that there is a substantial portion of pupils, in both social classes, who experience positive effects from truancy on educational attainment. Indeed, a mean positive effect for pupils from low socio-economic classification is found.[11] Given that the effect can be considered as causal (if the instrument is 'true') then how can we rationalise these results?

Several factors may explain our findings. Firstly, truancy is unlikely to be a homogenous event. Whilst we are measuring the intensity of truancy (using ordinal scales) there is little information on the *type* of truancy. It is possible that truancy manifests itself in a variety of forms, some of which includes the stereotypical truanting behaviour of 'smoking, drinking and pretty crimes'. However, it also possible truancy is related to more productive means such as absenteeism to go to extra-curricular activities or part-time work, which may carry positive educational externalities. By using a mixed model we are able to estimate the outcomes of such different *types* of truancy. This explains the positive effects of truancy (all gamma's above zero).

The difference in the two distributions can also be explained. At first glance one may expect pupils from higher socio-economic backgrounds to be less affected by truancy than pupils from lower socio-economic backgrounds. However, when truancy is examined from an opportunity cost perspective, the true cost of truanting is likely to be higher for 'rich kids' than for 'poor kids'. The marginal cost of an additional hour of truancy to education is likely to be very high for children from high socio-economic backgrounds (perhaps they miss out on private tuition or an extra hour of high-quality schooling) whilst the marginal cost for children from poor socio-economic backgrounds is lower (they would not miss out on high quality schooling). Hence the higher negative impact that truancy has on pupils from high socio-economic backgrounds. Thus, generally 'rich kids' will do better at school than 'poor kids', but they suffer more when they truant.

Our example highlights the policy aspects of our estimator. If one only estimates the mean of the distribution of γ_i (as done in the 'standard' bivariate ordered probit case) then there is a danger that one underestimates the problem of truancy with respect to educational attainment. In the 'standard' case results suggest an insignificant effect of truancy on educational attainment for those from low socio-economic backgrounds and hence suggestions for policy may be that little should be done for this group. However, as we show, such a policy suggestion may be erroneous if the true effect of truancy on education is more nuanced and a substantial portion of students from low socio-economic backgrounds do experience negative educational effects. In general it appears the effect of truancy on education is more nuanced and warrants further research to identify different types of truancy and motivations behind truancy.

6. Conclusion

In this paper we have outlined a bivariate ordered probit estimator with mixed effects. Given prior distributional assumptions, our estimator allows one to estimate unbiased mixed effects that an ordered

[11] In both estimations (the semi-ordered and ordered-ordered case) the *mean* effect is statistically insignificant from zero. However, a standard normal distribution centred at zero still implies that half of pupils experience positive effects.

endogenous variable has on another ordered outcome. If the instrumental variable conditions are fulfilled these effects can be considered causal. Such an estimator may have many different types of applications and is valid whenever both dependent variable and a suspected endogenous variable take an ordered form (including binary). We highlight the use of our estimator with an example from the education literature to find that the effect of truancy on 16 year olds is a) a distributional effect as opposed to a common mean effect and b) the distributional effects may differ by sub-groups. Future developments of this estimator may include extensions which allow our estimator to be used in a panel context, to capture cluster or group effects in the distribution of γ_i, to allow for *M*-endogenous variable (multivariate ordered probit), or to extent the distributional properties of γ_i beyond a standard normal distribution to allow for a range of distributions to be estimated, such as lognormal, power or triangular distributions. Comparison of such different distributions of γ_i and the appropriate testing procedure to choose a best fit can also be developed.

Supplemental Materials

Whilst the data used in our example are restricted we are happy to provide the computer code used in our simulations. The estimator is implemented as STATA `ml lf` evaluator. It is not difficult to use this code as a template and adapt it for use on 'real' data. The authors will provide the code for the ordered-ordered version (BOP1000.DO) and for the semi-ordered version (SBOB1000.DO) of our estimator upon request. Both files call upon GH20.DO which contains the abscissae and quadrature weights for integration. The authors are happy to provide pointers and more specific help to individual requests.

- BOP1000.DO – Bivariate ordered-ordered probit model performing simulations on 1000 observations with 1000 replications for various values of rho
- SBOP1000.DO – Bivariate semi-ordered probit model performing simulations on 1000 observations with 1000 replications for various values of rho
- GH20.DO - abscissae and quadrature weights for integration

Bibliography

Angrist, J.D. (2001) "Estimation of Limited Dependent Variable Models with Dummy Endogenous Regressors: Simple Strategies for Empirical Practice", *Journal of Business and Economic Statistics*, Vol. 19, No. 1, pp. 2-28

Armstrong, D. and McVicar, J. (2000) "Value Added in Further Education and Vocational Training in Northern Ireland", *Applied Economics*, No. 32, pp. 1727-1736

Bosworth, D.L. (1994) "Truancy and Pupil Performance" *Education Economics*, Vol. 2, No. 3, pp. 243-265

Botti, F., Conte A., Di Cagno D.T. and D'Ippoliti, C. (2008) "Risk attitude in real decision problems" *The B.E. Journal of Economic Analysis & Policy - Advances:* Vol. 8: Iss. 1, Article 6

Bradley, S. and Nguyen, A.N. (2004) "The School-to-work transition" in Johnes, G. and Johnes, J. *International Handbook on the Economics of Education* (pp. 484-521), Edward Elgar, Cheltenham

Bradley, S. and Taylor, J. (2004) "Ethnicity, Educational Attainment and the Transition from School" *The Manchester School*, Vol. 74, No. 3, pp. 317-346

Burgess, S., Gardiner, K., Propper, C. (2002) "The Economic Determinants of Truancy", Centre for Analysis of Social Exclusion, CASE paper 61

Calhoun, C.A. (1989) "BIVOPROB: A Computer Program for Maximum-Likelihood Estimation of Bivariate Ordered-Probit Models for Censored Data", Working Paper, 38, Laxenburg, Austria: International Institute for Applied Systems Analysis

Calhoun, C.A. (1991) "Desired and Excess Fertility in Europe and the United States: Indirect Estimates from World Fertility Survey Data" *European Journal of Population*, Vol. 7, pp. 29-57

Calhoun, C.A. (1994) "The Impact of Children on the Labour Supply of Married Women: Comparative Estimates from European and U.S. Data" *European Journal of Population*, Vol. 10, pp. 293-318.

Cappellari, L. and Jenkins, S.P. (2003) "Multivariate probit regression using simulated maximum likelihood", *The Stata Journal*, No. 3, pp. 278-294

Cappellari, L. and Jenkins, S.P. (2006) "Calculation of Multivariate Normal Probabilities by Simulation, with Applications to Maximum Simulated Likelihood Estimation", *The Stata Journal*, No. 6, pp. 156-189

Conte, A., Hey J.D., and Moffatt. P.G. (2009) "Mixture models of choice under risk" *Journal of Econometrics*, doi:10.1016/j.jeconom.2009.10.011 (in press)

Dawson, P. and Dobson, S. (2008) "The influence of social pressure and nationality on individual decisions: Evidence from the behaviour of referees" forthcoming in the *Journal of Economic Psychology*, doi:10.1016/j.joep.2009.06.001

DCSF (2009) "Pupil Absence in Spring 2009 Term", Statistical Release, 26th August 2009, London

Dustman, C., Rajah, N. and Smith, S. (1997). "Teenage Truancy, Part-time Working and Wages", *Journal of Population Economics*, Vol. 10, No. 4, pp. 425-442

Greene, W.H. (2008) *Econometric Analysis 6th Edition,* Prentice Hall, New Jersey, USA

Greene, W.H. and Hensher, D.A. (2009) "Modeling Ordered Choices" Mimeo, Stern School of Business, New York (January 2009)

Hibbert, A., Fogelman, K. and Maoner, O. (1990), "Occupational Outcomes of Truancy", *British Journal of Educational Psychology*, Vol. 60, No. 1, pp. 23-36

McVicar, M and McKee, J. (2002) "Part-time Work During Post-Compulsory Education and Examination Performance: Help or Hindrance", *Scottish Journal of Political Economy*, No. 49, Vol. 4, pp. 393-406

Monfardini, C. and Radice, R. (2008) "Testing Exogeneity in the Bivariate Probit Model: A Monte Carlo Study" *Oxford Bulletin of Economics and Statistics*, Vol. 70, No. 2, pp. 271-282

Kawakatsu, H. and Largey, A.G. (2009) "EM algorithms for ordered probit models with endogenous regressors", *Econometrics Journal*, Vol. 12, pp. 164-186

Rivers, D. And Voung, Q.H. (1988) "Limited Information Estimators and Exogeneity Tests for Simultaneous Probit Models"*, Journal of Econometrics*, No. 39, pp. 347-366

Sajaia, Z. (2008) "Maximum Likelihood Estimation of a Bivariate Ordered Probit Model: Implementation and Monte Carlo Simulations", *The Stata journal,* No. 3, Vol. 2, pp. 311-328

Weiss, A (1993) "A Bivariate Ordered Probit Model with truncation: Helmet Use and Motorcycle Injuries", *Applied Statistics*, No. 42, pp. 487-499

Appendix

Let us consider the following transformation of the recursive system in eq. (1)

$$\begin{cases} y_{1i}^* = x_{1i}\beta_1 + \varepsilon_{1i} \\ y_{2i}^* - \gamma_i y_{1i}^* = x_{2i}\beta_2 + \varepsilon_{2i} \end{cases} \tag{A1}$$

that in matrix form is:

$$\Gamma Y^* = \mathrm{B}X + \mathrm{E}\,, \text{ with } \mathrm{E} \sim N\left(\begin{pmatrix}0\\0\end{pmatrix}, \Sigma\right) \text{ and } \Sigma = \begin{bmatrix}1 & \rho\\ \rho & 1\end{bmatrix}, \tag{A2}$$

where:

$$Y^* = \begin{bmatrix}y_{1i}^*\\ y_{2i}^*\end{bmatrix};\ \Gamma = \begin{bmatrix}1 & 0\\ -\gamma_i & 1\end{bmatrix};\ X = \begin{bmatrix}x_{1i}\\ x_{2i}\end{bmatrix};\ \mathrm{B} = \begin{bmatrix}\beta_1 & 0\\ 0 & \beta_2\end{bmatrix};\ \mathrm{E} = \begin{bmatrix}\varepsilon_{1i}\\ \varepsilon_{2i}\end{bmatrix}. \tag{A3}$$

By pre-multiplying both sides by Γ^{-1}, we get:

$$Y^* = \Lambda X + \Upsilon\,, \tag{A4}$$

where $\Lambda = \Gamma^{-1}\mathrm{B}$, $\Upsilon = \Gamma^{-1}\mathrm{E}$, and $\Upsilon \sim N\left(\begin{pmatrix}0\\0\end{pmatrix}, \Omega\right)$, with $\Omega = (\Gamma^{-1})\Sigma(\Gamma^{-1})' = \begin{bmatrix}1 & \rho+\gamma_i\\ \rho+\gamma_i & 1+2\gamma_i\rho+\gamma_i^2\end{bmatrix}$.

Let us define a matrix Θ having on the principal diagonal the inverse squared root of the terms in the principal diagonal of Ω and zero somewhere else. By pre-multiplying eq. (A4) for Θ, we get:

$$\Pi Y^* = \Pi\Lambda X + \Pi\Upsilon\,. \tag{A5}$$

It is worth noting that the error term of the system transformed as such is now:

$$\Pi\Upsilon \sim N\left(\begin{pmatrix}0\\0\end{pmatrix}, \Pi\Omega\Pi\right), \text{ where } \Pi\Omega\Pi = \begin{bmatrix}1 & \tilde{\rho}_i\\ \tilde{\rho}_i & 1\end{bmatrix} \text{ with } \tilde{\rho}_i = \frac{\rho+\gamma_i}{\sqrt{1+2\gamma_i\rho+\gamma_i^2}}. \tag{A6}$$

Basically, we have transformed the system of recursive equations in eq. (1) in the following way in order to fill all its terms in a bivariate standard normal probability distribution function that is available in all the principal statistical packages. The system of equations after our transformation appears then:

$$\begin{cases} y_{1i}^* = x_{1i}\beta_1 + \varepsilon_{1i} \\ \dfrac{y_{2i}^*}{\sqrt{1+2\gamma_i\rho+\gamma_i^2}} = \dfrac{\gamma_i y_{1i}^* + x_{2i}\beta_2}{\sqrt{1+2\gamma_i\rho+\gamma_i^2}} + \dfrac{\gamma_i\varepsilon_{1i}+\varepsilon_{2i}}{\sqrt{1+2\gamma_i\rho+\gamma_i^2}} \end{cases}. \tag{A7}$$