JENA ECONOMIC RESEARCH PAPERS

2010 – 062

The Impact of Truancy on Educational Attainment: 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

The Impact of Truancy on Educational Attainment: A Bivariate Ordered Probit Estimator with Mixed Effects

Franz Buscha
Centre for Employment Research
University for Westminster, London, UK
buschaf@wmin.ac.uk

Anna Conte
Strategic Interaction Group
Max-Planck-Institut für Ökonomik, Jena, Germany
and
Department of Economics and Quantitative Methods
University for Westminster, London, UK
aconte@econ.mpg.de; A.Conte@westminster.ac.uk

Abstract: This paper investigates the relationship between educational attainment at age 16 and truancy. Using data from the Youth Cohort Study of England and Wales, we estimate the causal impact that truancy has on GCSE examination outcomes. Problematic is that both truancy and attainment are measured as ordered responses requiring a bivariate ordered probit model to account for the potential endogeneity of truancy. Furthermore, we extent the 'naïve' bivariate ordered probit estimator to include mixed effects which allows us to estimate the distribution of the truancy effect on educational attainment. This estimator offers a more flexible parametric setting to recover the causal effect of truancy on education and results suggest that the impact of truancy on education is indeed more complex than implied by the naïve estimator.

JEL: I20; C35; C51
Keywords: educational attainment, truancy, bivariate ordered probit, mixed effects

The authors gratefully acknowledge valuable advice and comments from Paul Latreille (Swansea), Franco Peracchi (Tor Vergata, Rome), seminar participants at the International Workshop in Applied Education Economics 2010 (Catanzaro) and seminar participants at the Work, Pensions and Employment Group annual conference 2010 (Bristol).

1. Introduction

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, 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; Farrington, 1996). 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 reports attesting a record rate of school sessions being missed (1.03% - DCSF, 2009). Moreover, the long term trend does not look favourable with truancy rates now higher in 2009 than they were in 1997. This is even though the UK government has spent over £885 million on anti-truancy policies during the period 1997/98 to 2003/04, and set targets to cut truancy by a third during this time period. 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.[1]

Romer's (1993) study was one of the first to argue that truancy is a significant phenomenon and that evidence suggested a strong association between poor exam performance and high levels of truancy. Since then, a number of other studies have examined the relationship between truancy and educational attainment and have come to similar conclusions; namely, that truancy is a strong indicator of poor educational outcomes (Durden and Ellis, 1995; Marburger, 2001, Kirby and McElroy, 2003). The majority of this evidence relates truancy to educational outcomes at a university level and relatively few studies have examined the impact of truancy on educational performance at age 16 when education is still compulsory. Indeed, the compulsion by law to attend school until the age of 16 creates an interesting scenario where truancy can be seen as a rational phenomenon for youths whose discount rates are high and have no other way of escaping additional years of schooling.[2] Within this context, and given that age 16 signifies the start of the 'school-to-work' transition, determining exactly to what extent truancy causally affects compulsory educational attainment becomes an important question.

For the UK, Bosworth (1994) explores the determinants and effects of truancy for pupils in their final year of compulsory schooling. Using data from the Youth Cohort Study (YCS 3), his findings suggest that boys are slightly more likely to truant than girls and that individuals from single parent families are more likely to truant. Individuals who reported high levels of truancy are less likely to obtain good educational scores at age 16. However, even individuals who report fairly low levels of truancy significantly tend to obtain lower grades; these findings suggest that it is not so much the hours of truancy that impacts, rather the signal that truanting represents. Payne (2001), also using data from the Youth Cohort Studies (YCS 9 and 10), finds that pupils who stayed-on and went into post-compulsory education (further education, ages 16-

---

[1] Whilst this paper focuses on the UK impact of school truancy, a likewise problem exists for the US. A good overview of the US literature is provided by the National Centre for School Engagement (Heilbrunn, 2007) whose report highlights the strong detrimental effect that truancy can have and how research surrounding truancy is still in its infancy.

[2] See Harmon, *et al.* 2003 for a discussion of optimal schooling choices.

18) are much more likely to drop-out before they reach the end (at age 18), if they truanted during their final year of compulsory schooling (age 16).

However, because truancy is an individual choice, it is likely to be driven by unobserved factors such as motivation and satisfaction which clearly also impact on educational outcomes. In this sense, truancy could be considered an endogenous variable and consequently requires to be treated as such in any educational regression. Furthermore, one could argue that education determines truancy: those not doing well at school become disillusioned and want to leave the compulsory education system as soon as possible because they believe they can gain a higher pay-off in the labour market by working early. However, because pupils are bound by law to remain in school until the age of 16, they can only accomplish this by truanting. Seen in such a light, the causal nature of truancy on education is reversed and any causal impact of truancy on education is nil.

Therefore, in order to estimate the causal impact that truancy has on education we must take the potential endogeneity of truancy into account. However, many of the aforementioned studies do little to control for the possible endogeneity of truancy in their attainment regressions. Only in recent years have studies by Stanca (2006), Chen and Lin (2008) and Dobkin *et al.* (2009) attempted to use more innovative methods (such as random design, panel data and regression discontinuity) to recover a *causal* effect on education by controlling for unobservable characteristics which could influence both truancy and attainment. However, none of these studies examine the impact of truancy on attainment before individuals are eligible to leave school. On the contrary, we argue that it is important to identify the causal nature of truancy on education before self-selection into further and higher education takes place.

In this paper, we try to address the issue of endogeneity of truancy, but we go even further. It is our belief that, even after having dealt with truancy as an endogenous variable, the relationship linking truancy behaviour to educational attainment is more complex than how it can be approximated by conditional mean estimators. In other words, we think that even in a *ceteris paribus* condition there might still be substantial individual differences in the intensity of truancy effect on education. For this reason we estimate the causal impact of truancy on the educational outcomes of 16 year olds in the UK by using a modified bivariate ordered probit model which allows us to estimate the distribution of the effect of truancy on attainment over the population. By appropriately modelling the parameter of interest according to a suitable distribution, we aim to take the heterogeneity of the truancy effect into account in any estimates of educational attainment. Results support our intuition and suggest that the impact of truancy on educational attainment is more complex than assumed by conditional mean estimators and that the effect of truancy on educational attainment is more heterogeneous for boys than for girls.

The next section describes the data used in this study. Section 3 outlines our methodology, derives the bivariate ordered probit estimator with mixed effects and discusses the identification of the model. Section 4 presents the results, whilst section 5 concludes.

2. Data

One of the main data sources for the UK education literature which examines the schooling experiences of students in their final year of compulsory schooling and beyond, and contains records on truancy and educational outcomes at age 15/16, 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 to further education and higher education, or to the labour market. The YCS data is a longitudinal dataset designed to follow individuals over 3 years (3 consecutive sweeps starting at age 15/16) and now has 13 cohorts with over 35 sweeps. In this study we make use of only the first sweep as truancy information is not available in later interviews. 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 10 which corresponds to student eligible to leave school at the end of the 1998/99 school year). 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 the included sample 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. 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 Department for Children Schools and Families (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 information for YCS 10 returns the following descriptive statistics:

Table 1: Truancy and Educational Attainment at Age 16 in the Youth Cohort Study, Sweep 1.

Truancy	*Frequency*	*Percent*	*Weighted Percent*	*Highest Educational Attainment at 16*	*Frequency*	*Percent*	*Weighted Percent*
Never	8,647	68.51%	65.00%	None	545	4.22%	4.22%
For the odd day or lesson	3,014	23.88%	24.98%	1-4 GCSE D-G grade	200	1.55%	3.34%
Particular days or lessons	607	4.81%	6.23%	5+ GCSE D-G grade	1,063	8.24%	18.03%
For several days at a time	188	1.49%	1.97%	1-4 GCSE A*-C grade	2,788	21.60%	25.49%
For weeks at a time	166	1.32%	1.82%	5+ GCSE A*-C grade	8,310	64.39%	48.92%
Total	12,622	100.00%	100.00%	Total	12,906	100.00%	100.00%

Source: YCS 10 (surveys those eligible to leave school in 1998-1999)

The data at hand – that can be interpreted as a representative sample of the UK school population in their final year of compulsory schooling – suggests that the majority of youths did not truant. However, a substantial portion (31%) engaged in minor forms of truancy whilst a small proportion (4%) engaged in more serious forms of truancy. Approximately half of all pupils achieved the highest grade category of 5 or more A*-C grade GCSE's, whilst a further 25% achieved the second highest category of 1 to 4 A*-C grades. A small proportion (7.5%) achieved none or poor grades.

3. Methodology

3.1 A Bivariate Ordered Probit model with Mixed Effects

Generally, when a suspected endogenous explanatory variable is encountered in the applied micro-economics literature, instrumental variable methods are often applied to estimate a causal and consistent effect.[3] As such, instrumental variable estimation technique 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.[4]

[3] We will not discuss the LATE implications of the instrumental variable estimator here.

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

When both the dependent variable and the endogenous variable take a binary form, a bivariate probit model can be used (Greene, 2008: 827).[5] 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).[6] 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).[7] In recent years a variety of papers have been written which make use of bivariate ordered probit and semi-ordered probit estimators and a convenient overview of this 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.

Our variables of interest, as shown in Table 1, suggest that a bivariate 'ordered-ordered' methodology would be an appropriate solution. However, we modify the existing bivariate ordered probit estimator (hereafter called the naïve bivariate ordered probit estimator) to include a mixed effect. This methodology, emerging in the behavioural economics literature,[8] 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. In other words, we assume that, *ceteris paribus*, the effect of the endogenous variable on the dependent variable can differ individual by individual and that such differences are captured by an appropriate choice of the distribution function for this effect. The reason for such a hypothesis is that truancy and education are complex phenomena that go beyond observable explanatory factors. Our approach aims to capture nuances in the truancy effect on educational attainment; nuances which cannot be captured by simply framing a particular individual in a specific cohort of people identified by a certain combination of observables.[9]

Assume that the two latent variables attitude to truancy and educational outcome, respectively T_i^* and E_i^*, are determined by the following system of equations:

$$\begin{cases} T_i^* = x'_{Ti}\beta_T + \varepsilon_{Ti} \\ E_i^* = x'_{Ei}\beta_E + \gamma_i T_i^* + \varepsilon_{Ei} \end{cases}, \quad \text{with } \gamma_i \sim N(\mu_\gamma, \sigma_\gamma^2) \text{ and } \begin{pmatrix} \varepsilon_{Ti} \\ \varepsilon_{Ei} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right). \quad (1)$$

Here, x_{Ti} and x_{Ei} are vectors of observables, β_T and β_E are a vector of parameters, γ_i represents the effect that T_i^* has on E_i^* for individual *i*, and ε_{Ti} and ε_{Ei} are two error terms, assumed to be jointly normal with correlation coefficient ρ and uncorrelated with everything else in the model; in particular, $E(x_{Ti}\varepsilon_{Ti}) = 0$ and $E(x_{Ei}\varepsilon_{Ei}) = 0$.[10]

---

[5] These can be extended to the multivariate case when more than one binary endogenous 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).

[6] Implemented as a Stata routine bioprobit by Sajaia (2008).

[7] 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.

[8] See, for example, Botti *et al.* (2008) and Conte *et al.* (2011).

[9] The mixed effect approach has further advantages; even if a large vector of explanatory variables were available, a 'kitchen sink' approach can sometimes cause more harm than good in terms of collinear variables or introducing further endogenous regressors (which in case of simultaneous regressors can bias all other parameter estimates). Furthermore, whilst one can use interactions terms within the regression specification to model more complex truancy effects, the interpretation of these is notoriously difficult in a non-linear setting.

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

The basic idea underlying our model is that, when trying to explain people's choices, at stake are both observed factors, represented by x_{Ti} and x_{Ei} (for example demographic variables) and unobserved factors (for example motivation) embedded in γ_i and in the joint distribution of ε_{Ti} and ε_{Ei}. In this model, we assume that there is heterogeneity between individuals in the way the latent variable T_i^* influences E_i^*. 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 normal distribution with mean μ_γ and standard deviation σ_γ, 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 T_i^* and E_i^*. What we observe, instead, are the two categorical variables T_i and E_i. The unobserved latent variables T_i^* and E_i^* are related to their respective observed outcomes, according to the classification in Table 1, by the following observational rules:

$$T_i = \begin{cases} 1 = Never & if\ T_i^* \leq j_{11} \\ 2 = Odd\ day & if\ j_{11} < T_i^* \leq j_{12} \\ 3 = Particular\ day & if\ j_{12} < T_i^* \leq j_{13} \\ 4 = several\ days & if\ j_{13} < T_i^* \leq j_{14} \\ 5 = Weeks\ at\ a\ time & if\ j_{14} \leq T_i^* \end{cases} \qquad E_i = \begin{cases} 1 = None & if\ E_i^* \leq j_{21} \\ 2 = 1-4\ GCSE\ D-G & if\ j_{21} < E_i^* \leq j_{22} \\ 3 = 5+GCSE\ D-G & if\ j_{22} < E_i^* \leq j_{23}, \\ 4 = 1-4\ GCSE\ A^*-C & if\ j_{23} < E_i^* \leq j_{24} \\ 5 = 5+GCSE\ A^*-C & if\ j_{24} \leq E_i^* \end{cases} \qquad (2)$$

where the $j_{..}$ are cut-points to be estimated along with the other parameters of the model.[11]

The probability of observing $T_i = l$ and $E_i = m$ for individual *i* is:

$$\begin{aligned} Pr(T_i = l, E_i = m) & \\ &= \Phi(j_{1l} - x'_{Ti}\beta_T, (j_{2m} - \gamma_i x'_{Ti}\beta_T - x'_{Ei}\beta_E)\lambda_i, \tilde{\rho}_i) \\ &- \Phi(j_{1l-1} - x'_{Ti}\beta_T, (j_{2m} - \gamma_i x'_{Ti}\beta_T - x'_{Ei}\beta_E)\lambda_i, \tilde{\rho}_i) \\ &- \Phi(j_{1l} - x'_{Ti}\beta_T, (j_{2m-1} - \gamma_i x'_{Ti}\beta_T - x'_{Ei}\beta_E)\lambda_i, \tilde{\rho}_i) \\ &+ \Phi(j_{1l-1} - x'_{Ti}\beta_T, (j_{2m-1} - \gamma_i x'_{Ti}\beta_T - x'_{Ei}\beta_E)\lambda_i, \tilde{\rho}_i) \end{aligned} \qquad (3)$$

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 likelihood contribution of individual *i* is:

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

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

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 upon request. To examine the small sample properties of our estimator we implement

[11] 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 $j_{2M} = \infty$.

Monte Carlo simulations (Appendix B). We find that our estimator performs well even in small samples (200 observations).

The assumption that $\gamma_i \sim N(\mu_\gamma, \sigma_\gamma^2)$ 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 will collapses into a naïve 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 have on length of survival will always have positive effects which may be distributed with an exponential decay and thereby warrant a power function.[12]

3.2 *Identification*

In order to identify the recursive system of equations (1), we have to impose at least one variable acting as an exclusion restriction on the regressors in x_E. The coefficient on such a variable is then estimated along with the coefficients on several standard exogenous controls, such as socio-economic background, school type, ethnicity, parental education, gender, household composition, housing and disability status (Bosworth, 1994; Bradley and Taylor, 2004).[13] In other words, it is required that at least one of the explanatory variables in the truancy equation does not enter the education equation, i.e. x_E is a subset of x_T.[14] With this in mind, we have to look for at least one variable that in principle contributes to explain T_i^*, but that can be reasonably excluded from the group of variables that are meant to explain educational attainments, E_i^*, in addition to being uncorrelated with any potential unobservable characteristics that drive education (such as ability or motivation).

For example, a cinema in the close neighbourhood of a school might well explain an abnormal truancy level in that school but can certainly not be considered as a direct cause of a bad academic performance. The effect of the cinema on pupils' academic performance can only be explained by means of an indirect chain of events: the presence of the cinema increases absences from school; absences from school determine low educational attainments. In this sense, the presence of a cinema causes students' poor performance but only through the effect it has on truancy.

Generally, the determinants of teenage truancy have not received much academic attention in the UK, although some previous literature by Bosworth (1994), Dustman *et al.* (1997) and Burgess *et al.* (2002) does exists. From these, a variety of factors have been found to influence the truanting decision and a convenient (non-academic) overview of many of such factors is provided by the Illinois State Board of Education Truants' Alternative & Optional Education Program:

---

12 For comparative purposes we also estimate a specification with a log-normally distributed effect: $\ln \gamma_i \sim N(\mu_\gamma, \sigma_\gamma^2)$.

13 A table with descriptive statistics for these controls is provided in Appendix C.

14 For example, suppose that $x_{Ti} = x_{Ei}$; then the system in (1) becomes:

$$\begin{cases} T_i^* = x'_{Ti}\beta_T + \varepsilon_{Ti} \\ E_i^* = x'_{Ti}(\beta_E + \gamma_i\beta_T) + \gamma_i\varepsilon_{Ti} + \varepsilon_{Ei} \end{cases}.$$

Consequently, only β_T is identified, and β_E, γ_i and ρ cannot be estimated consistently.

Table 2: Possible causes of Teenage Truancy

Child	Family
1. Poor self-concept; low self-esteem	1. Parents lack appreciation for value of education
2. Low academics; particularly behind in reading and math	2. Parents are high-school dropouts
3. Socially isolated; poor inter-personal skills	3. Financial difficulties (lack of adequate nutrition, transportation; inability to pay fees)
4. Feeling of not belonging at school, of being different	4. Ineffective parenting; lack of parental control and supervision
5. Feeling of lack of control over life	5. Familial instability (one-parent household; poor parent-child relationship)
6. Little or no extra-curricular involvement	6. Inappropriate role models
7. Mental and/or emotional instability; childhood depression	7. Child abuse and/or neglect
8. Unidentified learning disabilities	8. Substance abuse; alcoholic parent
9. Vision and/or auditory problems that have gone undiagnosed	9. Parental convenience (older sibling kept home to babysit)
10. Language barriers	
11. Poor health	
12. Negative peer relationships; older, non-school oriented friends	
13. Substance abuse	
14. Fear of school, teachers, and/or administrators	
15. Experienced recent traumatic event (divorce, death of a loved one)	
16. Fear of physical protection going to or at school	
17. Teenage pregnancy and/or parenting	
School	**Community**
1. Uninteresting and irrelevant curriculum	1. Lack of support for school
2. Improper class placement (above or below abilities)	2. Lack of, or unresponsive community service agencies
3. Failure to identify and provide services for problem students	3. Community upheaval and social change
4. Poor pupil-teacher relationships	4. Loss of neighbourhood schools and sense of 'ownership' of schools
5. Insufficient counselling and guidance staff	5. Negative peer influences
6. High student/teacher ratio	6. High incidence of substance abuse, criminal activity
7. Low teacher expectations	7. Gang activity
8. Lack of parent - school communication and involvement	8. Inadequate provision for transportation
9. Too weak, or too rigid administration of policies.	

Source: Illinois State Board of Education's Truants' Alternative and Optional Education Program

Table 2 highlights the conceptual problem we are faced with; the overall majority of factors that influence truancy also influence the education decision. Virtually all child, family and school specific factors are related not only to truancy but also to unobservables which drive the education decision thereby making them invalid choices for a possible exclusion restriction. We therefore decide to go to a community/regional level and make use of the fact that we have local education authority (LEA)[15] codes available in our data (restricted version).[16] We are confident that a regional factor which influences truancy and does not affect education unobservables may exist.

After excluding regions from Wales, because of linkage problems, our remaining sample contains 148 different English LEAs with an average response of 86 students per region. The size and scope of the regions vary with mid-year population estimates for 1999, suggesting an average size of 330,000 inhabitants per LEA (ranging from a minimum of 33,500 for Hartlepool to a maximum of 1.3 million for Kent). Using these regional identifiers, we proceed to impute several additional variables which could potentially serve as exclusion restrictions.

[15] The term LEA has been defunct since the Education and Inspections Act of 2006 and been replaced by Local Authority (LA). However, the term is still used informally and we use the term in that capacity.
[16] Our data also contains school identifiers although these are anonymised and we are unable to link this to other data. The average pupil response at school level is 12 per school.

Table 3: Possible Community causes of Teenage Truancy

Variable	Description	Data Source	Aggregation	Theoretical relation to: Truancy	Theoretical relation to: Education unobservables (ability, motivation & drive)
Local truancy rate	The percentage of half days missed in the 1998/99 school year due to unauthorized absence.	Department for Children, School and Families	148 LEAs	Local truancy rates are likely a determinant of individual truancy – high local truancy increases an individual's propensity to truant	If the transmission from local truancy to individual truancy is through peer effects then these are also likely to influence educational motivation
Local education rate	The proportion of students obtaining 5+ABCs at GCSE in the 1998/99 school year	Department for Children, School and Families	148 LEAs	High performing authorities may have more stringent anti-truancy policies	Local education authorities are likely to directly influence educational motivation through education policies
Local part/full-time pay	Average hourly part/full-time pay in 1999	Office of National Statistics	148 LEAs	Youths in year 11 are generally not allowed to work full-time due to school leaving age legislation (they are required to stay until age 16). However, they are able to work part-time during full-time education. Higher part-time pay may induce higher labour market participation during the school year which will increase truancy.	High labour market returns will alter the expected return to schooling which in turn will influence educational motivation.
Local unemployment rate	Local unemployment rate in 1999	Office of National Statistics	148 LEAs	High unemployment has been correlated with lower truancy (Raffe, 1986)	High unemployment rates likely induce a 'discouraged worker effect' whereby educational motivation increases as unemployment rises as pupils seek to avoid the labour market
Local ASBO rate & Local ASBO breach rate	The number of Anti-Social Behaviour Orders (and breaches) per 1000 of population giving out during 1999 and 2003	Ministry of Justice	39 Criminal Justice System Area's	Various studies have linked crime to truancy (see Prior and Paris, 2005).	If the transmission mechanism is through peer groups then these are also likely to influence educational motivation
Local Urbanicity	The proportion of individuals living in urban/rural environments 2001	Department for Environment, Food and Rural Affairs	148 LEAs	Distance to school may influence the probability of truanting with rural communities likely to have larger distances. Alternatively, a larger share of 'distractions' in urban communities (such as cinemas and parks) may influence truancy there.	Local and rural communities are likely to place difference emphases on education and may thus influence education motivation

Table 3 presents various geographical variables we merged into our dataset. Local economic conditions are proxied by full/part-time hourly pay and the unemployment rate. Local crime and delinquency conditions are proxied through Anti-Social Behaviour Order (ASBO) statistics. Unfortunately, these statistics are only collected at a higher regional aggregation. In addition they were only introduced in 1999 and we had to sum ASBO statistics over multiple years to obtain enough differentiation. Local truancy conditions are proxied by information from the National Pupil Absences tables and finally we also introduces a measure of urbanicity which proxies local life-style decisions.

For investigative purposes we have taken a naïve approach where each local variable is independently used as an exclusion restriction without any consideration of theoretical linkages between our relevant variables. We also decide to include two additional individual survey level variables into our regressions acting as proxies for unobserved factors.[17] Results are presented in Appendix D. Estimates show that the mean impact of truancy on education varies considerably with estimated values of μ_γ ranging within the interval [-1, +1]. However, estimates also suggest that few of the variables which act as exclusion restrictions are highly correlated with truancy – the exception being local part-time hourly pay and local truancy rates. Finally, results also show that, although the mixed effects models produce varying values for σ_γ, these values are relatively similar to each other ranging between 0.2 and 0.6. This is important because we will see that, even if we are unable to causally identify the impact of truancy on education, we might be able to identify some measure of heterogeneity around the impact of truancy on attainment.

Having examined results from the naïve identification approach we consider only two variables which could be used as possible exclusion restrictions: local part-time hourly pay and the local truancy rate. Using other local level variables would result in weak identification which in turn may result in high standard errors, imprecise estimates and poor 'traction' in the maximum likelihood estimation process.

Local part-time hourly pay appears to be strongly and significantly correlated with individual propensities to truant. This is likely because individuals in their final year of compulsory schooling are only allowed to work part-time (as opposed to full-time) in the labour market. High local hourly part-time wages are therefore a strong incentive to 'skip school and earn money'. Evidence by Dustmann *et al.* (1997) and Burgess, *et al.* (2002) supports such a view as they find that part-time working and truancy are closely correlated. Assuming that local wages are exogenously determined by macro-economic conditions one could consider the use of local part-time hourly pay as a potential exclusion restriction. However, the usefulness of this variable, or any other local economic indicator, as an exclusion restriction may be marred by an 'encouraged/discouraged worker effect'. If such an effect exists then local economic conditions directly influence education unobservables by changing the expected return to a certain level of education. For example, local unemployment rates at an LEA level have been shown to influence educational decisions (Andrews and Bradley, 1997; Rice 1999) in such a way. High local unemployment leads to a lower probability of obtaining a job after school which in turn reduces the opportunity cost of an additional year of schooling. This directly influences individual motivation towards education and as a result pupils choose to stay-on in school. Local pay conditions likely induce a similar effect and are thus a poor choice for an exclusion restriction as this variable is correlated with education unobservables.

Our choice of a suitable exclusion restriction thus falls on the local truancy rate. Our argument in support of this choice is the following: All pupils who live in a certain area are exposed to peculiar (area-specific) conditions that influence their truanting attitude. Thinking about the cinema example might be particularly useful in order to figure out one of the peculiar conditions we refer to. Explicitly controlling for any possible source of influence on truanting behaviour is impossible, as sample surveys are, in general, not sufficiently detailed for this purpose. An indicator of the local truancy rate captures the effect of all these unobserved external effects on pupils' truanting behaviour and consequently is a natural candidate for an exclusion restriction. Specifically, the UK government has since 1994 published National Pupil

[17] We include "parental help when the individual was younger" and the "survey response type" as unobservable indicators. We argue that the parental help proxy's unobserved parental engagement with the child and survey response type acts an indicator of unobserved motivation (those taking longer to respond are less motivated).

Absence tables every year which measure the level of authorised and unauthorised pupil absence in each LEA. And it is the so-measured local truancy rate that we intend to use as an exclusion restriction.

The effect of the local truancy rate on individual truancy needs not to be confused with the phenomenon known in the literature as peer-group effect or neighbourhood effect. The influence of local truancy rates on individual truancy, as measured at an LEA level, cannot be regarded as a peer-group effect because the regional aggregation of this variable is too large.[18] At such an aggregated level, individuals cannot influence the average local truancy rate and there are unlikely to be spurious peer-group effects due to group-level unobservables. However, the aggregate level of truancy could have an effect on a pupil's propensity to truant because it is meant to capture a tendency to truant triggered off by external conditions common to all the pupils in the area.

Obviously, one might argue that some of these unobserved external conditions/factors might influence individual educational performance as well. In order to counterbalance the effect of such unobservables on educational performance – that might link the local truancy rate directly to the educational attainment – we introduce an indicator of the local education rate (% of pupils gaining 5+ ABCs at GCSE) that is meant to capture the generalised effect of unobserved factors at a local level on individual educational performance (such as the quality of the local education authority or a self-selection process by more able families into particular regions). In addition we also include all other local level factors (including part-time pay) in our regressions to try and control for as many local level unobservables as possible. For example, local unemployment rates and pay conditions will control for any discouraged worker effect and its subsequent impact on education unobservables.

Having controlled for these various regional processes, we argue that the only link between local truancy rates and individual educational attainment is the indirect relation that both variables have via individual truancy. We thus argue that the local truancy rate is uncorrelated with individual education unobservables and our identification criteria are fulfilled. However, even if these assumptions were violated, we would be able to obtain partial inference through the use of simulation techniques (Nevo and Rosen, 2008).

If, for example, the local truancy rate does impact directly on individual educational attainment by being correlated with individual education unobservables, then we can estimate the direction and magnitude of the bias through simulation methods. Assume we have the following recursive system where we set the true value of μ_γ = -0.8 to suggest an overall negative effect of truancy on educational attainment. We set the value of σ_γ = 0.5 to suggest heterogeneity in the impact of truancy on attainment. The value of ρ = 0.5 to create a 'medium' amount of correlation across the error terms.[19] The strength of the exclusion restriction (z) on truancy (T_i^*) is set to 1. Finally, we introduce a direct effect of the exclusion restriction on education (E_i^*) measured by θ.

$$\begin{cases} T_i^* = 1x + 1z + \varepsilon_{Ti} \\ E_i^* = 1x + \gamma_i T_i^* + \theta z + \varepsilon_{Ei} \end{cases}, \quad \text{with} \quad \gamma_i \sim N(-0.8, 0.5) \quad \text{and} \quad \begin{pmatrix} \varepsilon_{Ti} \\ \varepsilon_{Ei} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix} \right) \qquad (6)$$

Figure 1 presents the results of a simulated dataset with 1000 observations for varying values of θ. As expected, biased estimates result when values of θ are different from zero.

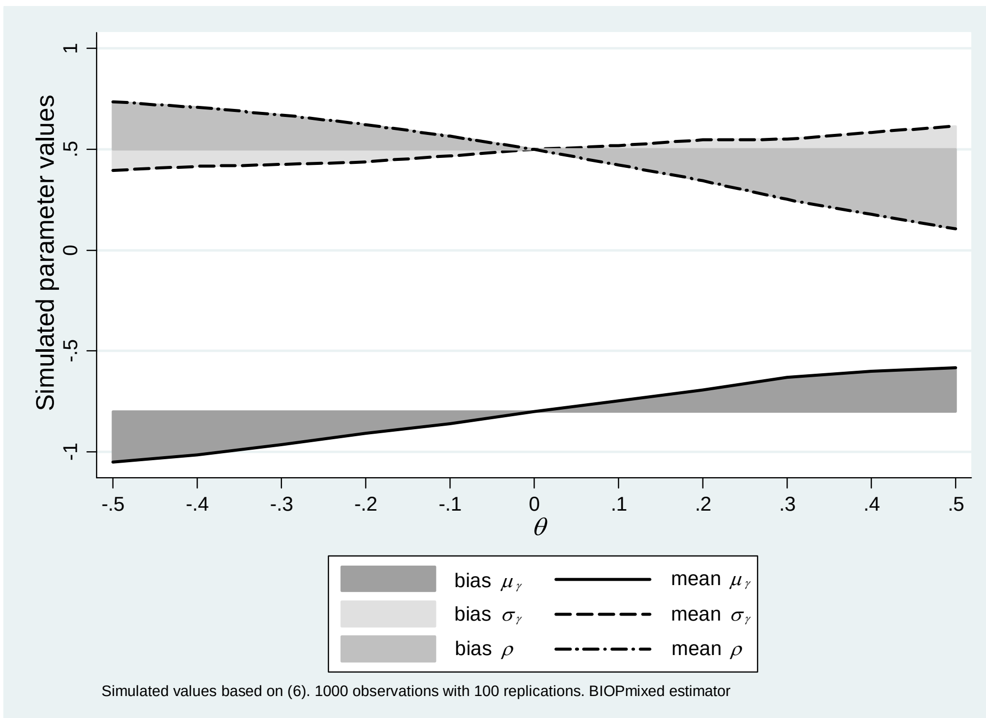
---

18 LEA populations are generally measured in 100,000s. Population estimates for 1999 suggest an average of 19,872 youths aged 15 to 19 per LEA. *Source: Nomis*

19 These values were chosen because they foreshadow our results

However, our simulation also shows that this bias is much less marked in estimates of σ_γ, thus indicating that we still obtain relatively good estimates of the heterogeneity, even when the identification assumptions fail. Finally, using this information we are able to place upper or lower bounds on final estimates. For example, let us assume there is a residual impact of the local truancy rates on individual educational attainment; assuming that the exclusion restriction z impacts negatively on individual educational attainment, then Figure 1 suggest that the estimated values of the impact of individual truancy on attainment would be below the true value. We can thus treat such an estimate as a lower bound and infer that the true impact of truancy on education is likely to be above (less negative than) the estimated value.[20]

Figure 1: The simulated effect of imperfect identification using the BIOPmixed estimator

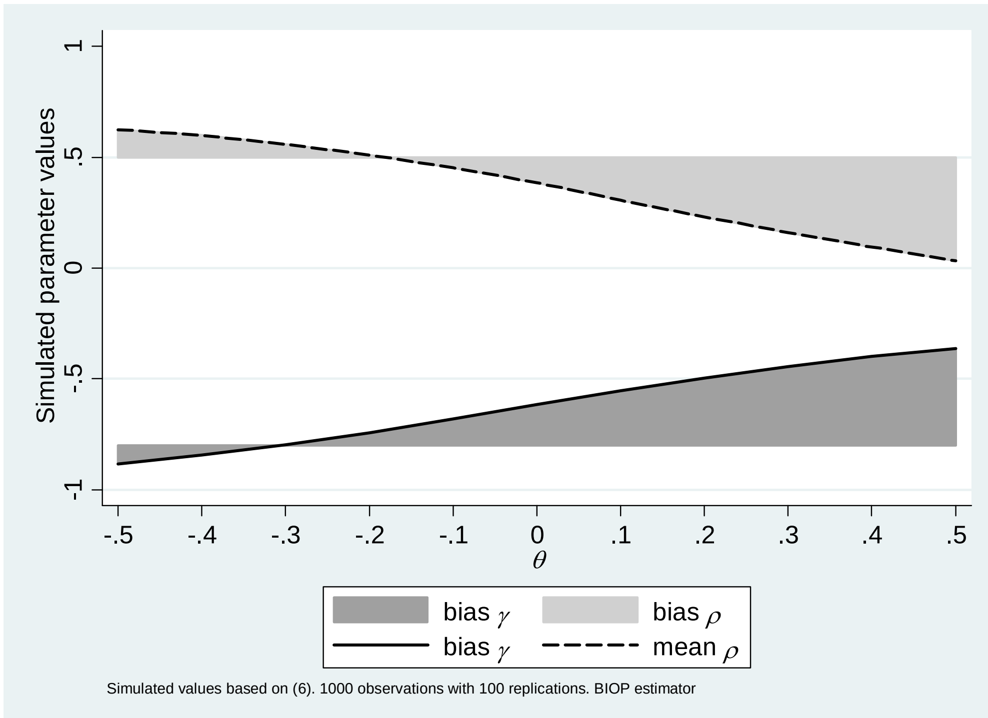


In such a setting, it is interesting to see that, if the effect of truancy on education is heterogeneous but heterogeneity is neglected, then the point estimates of a naïve bivariate ordered probit are biased even when $\theta = 0$. Figure 2 reports the results, based on 100 replications, of a Monte Carlo simulation similar to the one that generated the results in Figure 1. These simulations assume that the true model that generates the data is (6), but the estimation approach is one which does not take heterogeneity into account (BIOP). Results suggest that a failure to take heterogeneity into account – when the underlying data generating process is heterogenous – results in biased estimates, even under perfect identification conditions. Taking

[20] The results in Figure 1 show that, if θ, that measures the direct effect of z on education, is small relatively to the indirect (through truancy) mean effect of z on education $(-0.8 \times 1 = -0.8)$, the bias on the estimated mean of γ_i and ρ is rather small. For example, if θ takes on values within the interval $[-0.1, 0.1]$ that account for $[-12.5\%, 12.5\%]$ of the indirect impact of z on education, the induced biases on the mean of γ_i and ρ are also extremely moderate.

heterogeneity into account can thus be seen as an important component of the identification process.[21]

Figure 2: The simulated effect of imperfect identification under heterogeneous conditions using the naïve BIOP estimator



4. Results

For comparison purposes we present results from a naïve independent ordered probit model (IOP), the naïve bivariate ordered probit (BIOP) and bivariate ordered probit with mixed effects (BIOPmixed). In addition to estimating a mixed effect with a Gaussian distribution we also estimate a mixed effect with a log-normal distribution. Results are presented in Table 4.

Before examining the coefficients and implications of our findings, a quick comparison of the naïve bivariate ordered probit model and the 'mixed' bivariate ordered probit models suggests that the mixed effects model significantly increases the estimated log-likelihood, thereby resulting in a model with higher explanatory power. The log pseudo-likelihood of the mixed (Gaussian) model is -23888.39 whilst the log pseudo-likelihood of the naïve model is -24176.60. Unfortunately, the bivariate models cannot be run in a 'intercept only' mode and we are thus unable to gauge the impact of the mixed effect on Pseudo R^2 indicators. However, an increase in the log likelihood function of over 250 points seems substantial given that we are not adding extra variables into the modelling structure. This suggests that the mixed model substantially improves the explanatory power of estimation at little cost (other than computational).

Examining the coefficients on the determinants of truancy reveals that girls are more likely to be truant, non-whites are less likely to be truant, pupils from selective/independent

[21] Incidentally, such findings are mirrored in our application to real data. Appendix D clearly shows that the BIOP and BIOPmixed estimators consistently produce different values for the mean truancy effect, even though the identification conditions are identical.

schools are less likely to be truant, those living in rented housing are more likely to be truant, disabled pupils are more likely to be truant, those from lower socio-economic backgrounds are more likely to truant and those who have both parents in the household are less likely to commit truancy. The log local truancy rate has a positive impact on the probability to truant. Examining the determinants of educational attainment we find that; girls, non-whites, those with educated parents, those from better schools, those in owned houses, individuals from high socio-economic backgrounds and with both parents do better in school.

Examining the impact of truancy on educational attainment reveals that independent ordered probit estimation result in a negative statistically significant effect of -0.327.[22] The naïve bivariate ordered probit estimator suggests a larger effect of truancy on educational attainment of -0.655. Although this estimate is associated with a higher standard error it remains statistically significant at a 1% level. The estimated value of ρ is 0.404 and is statistically significant at a 10% level suggesting that the use of a bivariate model could be justified.

However, results from the bivariate ordered probit model with mixed effects suggest that the impact of truancy on educational attainment is more complex than assumed by the naïve bivariate ordered probit model – which only estimates a common mean effect, γ, to all individuals in the sample. The mixed model estimates an additional parameter; σ_γ, which represents the standard deviation of a normal/log-normal distribution of the effect of truancy on educational attainment. A statistically significant estimate of 0.569 for σ_γ indicates that there is considerably heterogeneity in the impact of truancy on educational outcomes across the population which a naïve estimator cannot capture. In addition, the estimated value of ρ increases somewhat to 0.499, although the standard error suggests no significant difference if compared with the naïve estimate of ρ. Figure 3 depicts our estimates in graphical form which further highlights the varied effect that truancy can have on educational attainment.

[22] Where the IOP method treats T_i as a continuous regressor.

Table 4: The Impact of Truancy on Educational Attainment at Age 16 (GCSEs) – Bivariate Ordered Probit with Mixed Effects

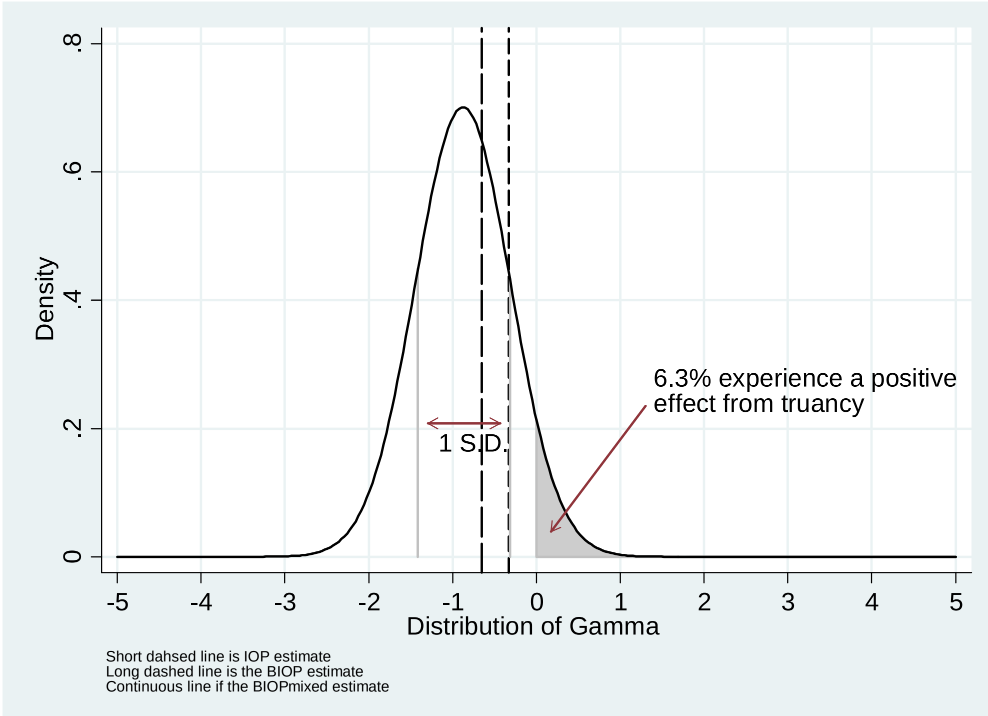
Variables		IOP			BIOP			BIOPmixed (Gaussian)			BIOPmixed (Log-normal)		
Dependent	Independent	Coef	S.E	P-val	Coef	S.E	P-val	Coef	S.E	P-val	Coef	S.E	P-val
Truancy	Female	0.053	0.026	0.042	0.052	0.026	0.041	0.056	0.026	0.031	0.093	0.023	0.000
	Non-white	-0.123	0.044	0.006	-0.117	0.043	0.006	-0.111	0.044	0.012	-0.082	0.037	0.026
	Neither parents employed	-0.119	0.044	0.008	-0.119	0.044	0.007	-0.112	0.044	0.010	-0.102	0.037	0.006
	At least one parent with A-level (ref: degree level)	0.041	0.042	0.333	0.042	0.041	0.306	0.049	0.042	0.246	0.036	0.039	0.352
	Neither parent with A-level (ref: degree level)	-0.027	0.036	0.451	-0.024	0.036	0.504	-0.026	0.036	0.468	-0.024	0.032	0.444
	Selective school (ref: comprehensive)	-0.329	0.058	0.000	-0.327	0.058	0.000	-0.318	0.061	0.000	-0.260	0.056	0.000
	Independent school (ref: comprehensive)	-0.441	0.048	0.000	-0.433	0.047	0.000	-0.478	0.053	0.000	-0.339	0.049	0.000
	House: rented (ref: owned)	0.290	0.038	0.000	0.292	0.037	0.000	0.299	0.037	0.000	0.317	0.032	0.000
	House: other (ref: owned)	0.122	0.091	0.181	0.113	0.090	0.210	0.128	0.089	0.150	0.115	0.083	0.168
	Disabled	0.089	0.064	0.160	0.103	0.062	0.098	0.064	0.063	0.312	0.150	0.052	0.004
	SEC: lower professional/higher technical (ref: high prof.)	0.046	0.038	0.230	0.045	0.037	0.231	0.031	0.038	0.422	0.007	0.036	0.845
	SEC: intermediate occupations (ref: high prof.)	0.136	0.043	0.001	0.140	0.042	0.001	0.113	0.042	0.008	0.103	0.039	0.009
	SEC: lower supervisory occupations (ref: high prof.)	0.233	0.054	0.000	0.238	0.053	0.000	0.230	0.053	0.000	0.183	0.047	0.000
	SEC: semi-routine routine/routine (ref: high prof.)	0.222	0.052	0.000	0.229	0.051	0.000	0.217	0.051	0.000	0.179	0.046	0.000
	SEC: other	0.279	0.063	0.000	0.283	0.062	0.000	0.263	0.062	0.000	0.193	0.054	0.000
	Both parents in household	-0.308	0.033	0.000	-0.306	0.032	0.000	-0.311	0.032	0.000	-0.314	0.029	0.000
	Log local education rate (% 5+ABCs)	-0.468	0.276	0.090	-0.468	0.272	0.085	-0.988	0.225	0.000	-0.944	0.201	0.000
	Log local unemployment rate 1999	-0.037	0.054	0.496	-0.040	0.054	0.453	-0.040	0.013	0.003	-0.054	0.024	0.000
	Log local part-time hourly pay rate 1999	0.269	0.176	0.126	0.255	0.172	0.139	0.255	0.124	0.140	0.212	0.234	0.003
	Log local full-time hourly pay rate 1999	-0.017	0.185	0.928	-0.006	0.182	0.976	-0.022	0.016	0.161	-0.018	0.019	0.040
	Local ASBOs per 1000 inhabitants 1999-2003	0.628	0.973	0.519	0.604	0.952	0.526	0.620	0.948	0.513	0.540	0.846	0.161
	Local ASBO breaches per 1000 inhabitants 1999-2003	-0.893	0.766	0.244	-0.901	0.748	0.229	-1.155	0.740	0.119	-0.985	0.752	0.513
	Local urbanicity % 2001	0.001	0.001	0.570	0.001	0.001	0.541	0.000	0.001	0.666	0.000	0.001	0.119
	Log local truancy rate (unauthorised absence)	0.098	0.028	0.000	0.098	0.027	0.000	0.084	0.028	0.002	0.086	0.023	0.000
	Government Office Region dummies		yes			yes			yes			yes	
	Behavioural proxies		yes			yes			yes			yes	
	j11	0.750	0.504	0.137	0.757	0.496	0.137	-0.315	0.103	0.002	0.355	0.045	0.000
	j12	1.690	0.505	0.001	1.697	0.496	0.001	0.638	0.105	0.000	1.363	0.047	0.000
	j13	2.220	0.505	0.000	2.238	0.497	0.000	1.205	0.103	0.000	1.922	0.052	0.000
	j14	2.556	0.506	0.000	2.585	0.497	0.000	1.611	0.105	0.000	2.334	0.061	0.000

continued

Variables		IOP			BIOP			BIOPmixed (Gaussian)			BIOPmixed (Log-normal)		
Dependent	Independent	Coef	S.E	P-val	Coef	S.E	P-val	Coef	S.E	P-val	Coef	S.E	P-val
Education	Female	0.333	0.025	0.000	0.340	0.033	0.000	0.389	0.061	0.000	0.332	0.033	0.000
	Non-white	-0.015	0.046	0.741	-0.062	0.055	0.264	-0.043	0.062	0.485	0.054	0.046	0.244
	Neither parents employed	-0.058	0.039	0.138	-0.107	0.048	0.026	-0.130	0.051	0.011	-0.114	0.042	0.007
	At least one parent with A-level (ref: degree level)	-0.218	0.048	0.000	-0.185	0.058	0.001	-0.206	0.079	0.009	-0.180	0.059	0.002
	Neither parent with A-level (ref: degree level)	-0.345	0.039	0.000	-0.335	0.048	0.000	-0.373	0.075	0.000	-0.335	0.058	0.000
	Selective school (ref: comprehensive)	1.165	0.123	0.000	0.970	0.221	0.000	1.161	0.353	0.001	0.986	0.251	0.000
	Independent school (ref: comprehensive)	0.568	0.089	0.000	0.340	0.187	0.068	0.490	0.292	0.093	0.577	0.202	0.004
	House: rented (ref: owned)	-0.429	0.032	0.000	-0.292	0.117	0.012	-0.286	0.169	0.090	-0.192	0.131	0.144
	House: other (ref: owned)	-0.210	0.076	0.006	-0.155	0.090	0.084	-0.164	0.118	0.163	-0.152	0.111	0.170
	Disabled	-0.406	0.058	0.000	-0.345	0.091	0.000	-0.380	0.119	0.001	-0.269	0.103	0.009
	SEC: lower professional/higher technical (ref: high prof.)	-0.233	0.047	0.000	-0.207	0.059	0.000	-0.251	0.079	0.001	-0.170	0.049	0.001
	SEC: intermediate occupations (ref: high prof.)	-0.405	0.049	0.000	-0.325	0.089	0.000	-0.374	0.126	0.003	-0.272	0.084	0.001
	SEC: lower supervisory occupations (ref: high prof.)	-0.612	0.054	0.000	-0.479	0.129	0.000	-0.524	0.195	0.007	-0.399	0.128	0.002
	SEC: semi-routine routine/routine (ref: high prof.)	-0.722	0.052	0.000	-0.586	0.138	0.000	-0.633	0.211	0.003	-0.561	0.149	0.000
	SEC: other	-0.694	0.059	0.000	-0.544	0.151	0.000	-0.601	0.224	0.007	-0.517	0.151	0.001
	Both parents in household	0.031	0.032	0.332	-0.099	0.083	0.234	-0.135	0.098	0.168	-0.157	0.079	0.047
	Log local education rate (% 5+ABCs)	1.148	0.257	0.000	0.909	0.374	0.015	0.725	0.625	0.246	0.746	0.049	0.247
	Log local unemployment rate 1999	0.069	0.053	0.189	0.059	0.057	0.304	0.001	0.022	0.964	0.005	0.012	0.964
	Log local part-time hourly pay rate 1999	0.360	0.168	0.032	0.492	0.182	0.007	0.156	0.128	0.045	0.179	0.140	0.223
	Log local full-time hourly pay rate 1999	0.244	0.181	0.177	0.229	0.191	0.231	0.117	0.122	0.231	0.154	0.154	0.340
	Local ASBOs per 1000 inhabitants 1999-2003	-0.535	0.965	0.579	-0.306	1.053	0.772	0.102	1.149	0.772	0.099	1.041	0.929
	Local ASBO breaches per 1000 inhabitants 1999-2003	-0.077	0.716	0.914	-0.266	0.778	0.733	-0.581	0.884	0.511	-0.411	0.774	0.511
	Local urbanicity % 2001	0.000	0.001	0.708	0.001	0.001	0.613	0.000	0.001	0.949	0.000	0.001	0.949
	Government Office Region dummies		yes			yes			yes			yes	
	Behavioural proxies		yes			yes			yes			yes	
	j21	-1.269	0.483	0.009	-0.788	0.478	0.100	-1.414	0.388	0.000	-1.743	0.261	0.000
	j22	-0.920	0.483	0.057	-0.462	0.472	0.328	-1.027	0.344	0.003	-1.578	0.238	0.000
	j23	0.031	0.483	0.949	0.433	0.467	0.355	-0.021	0.292	0.941	-1.045	0.166	0.000
	j24	0.887	0.483	0.066	1.246	0.477	0.009	0.878	0.340	0.010	-0.239	0.110	0.030
	$\mu_\gamma(\gamma)$	-0.327	0.014	0.000	-0.655	0.220	0.003	-0.876	0.212	0.000	-0.245	0.235	0.297
	σ_γ	n/a	n/a	n/a	n/a	n/a	n/a	0.569	0.117	0.000	0.039	0.049	0.426
	ρ	n/a	n/a	n/a	0.409	0.231	0.077	0.499	0.282	0.077	0.503	0.259	0.052
	Log pseudolikelihood		-11101.958/-13045.934			-24176.609			-23888.397			-23899.745	
	N		12175			12175			12175			12175	

Source: YCS 10 (surveys those eligible to leave school in 1998-1999), weighted

Figure 3: The Impact of Truancy on Educational Attainment on GCSE outcomes



Results from the mixed estimator suggest that the average effect of truancy on educational attainment is negative (-0.876) and statistically significant and that there exists a significant amount of heterogeneity around this average. For example, 68% of individuals experience a negative effect of truancy in the range of -1.444 and -0.307. Alternatively, 6.3% of individuals experience a positive (greater than 0) effect of truancy on educational attainment. This result is likely an artefact of the distributional assumption we make as by definition a small but finite percent of individuals will experience a positive effect if we impose an unrestricted normal distribution.[23] Indeed to test the distributional assumption we make by using a normal distribution, we also impose log-normal distributed mixed effect which has the advantage of being asymmetrical (perhaps a bulk of individuals experience a small effect and a few individuals experience very large effects) and bounded between $-\infty$ and 0.[24]

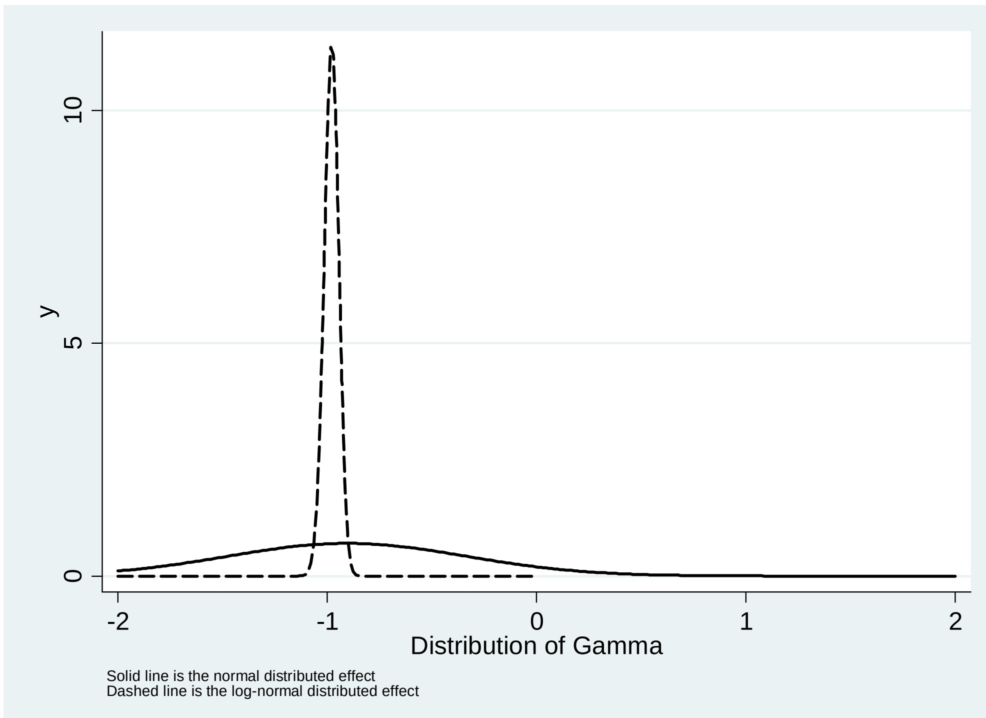
Regression results indicate that a log-normal functional form for γ_i does not improve the pseudo-likelihood and that the moments of the underlying normal distribution are statistically insignificant from zero (μ_γ= -0.245 with a *p*-value of 0.297 and σ_γ = 0.039 with a *p*-value of 0.426). However, it is prudent to plot the log-normal distribution anyway as the nature of the log-normal distribution can involve the estimation of rather small effects (as these are later exponentiated) which may not be easily seen. We plot both the normal and log-normal distribution in Figure 4 for comparative purposes and find that both estimators estimate

[23] An alternative hypothesis for the positive impact of truancy on educational attainment is that we are unable to examine 'why' students truant. One could imagine that youths who truant in the 'last hour of school' would not have received much added educational value anyway. Moreover, if they truant because they participate in extra-curricular activities or hold part-time jobs then the marginal benefit of an hour outside school to educational skills may higher than inside school.

[24] The alternative bounding is between 0 and ∞ which is unlikely to be correct in the case of truancy on attainment.

approximately the same mean effect, μ_γ. However, it appears that the log-normal estimator attempted to impose a symmetrically distributed effect around this mean – which in the case of log-normal distribution can only be done with a very 'tight' standard deviation. Such results therefore confirm out initial hypothesis that a symmetric distribution is the right choice and we accept the normally distributed mixed effect.

Figure 4: Normal vs. log-normal distributed effect



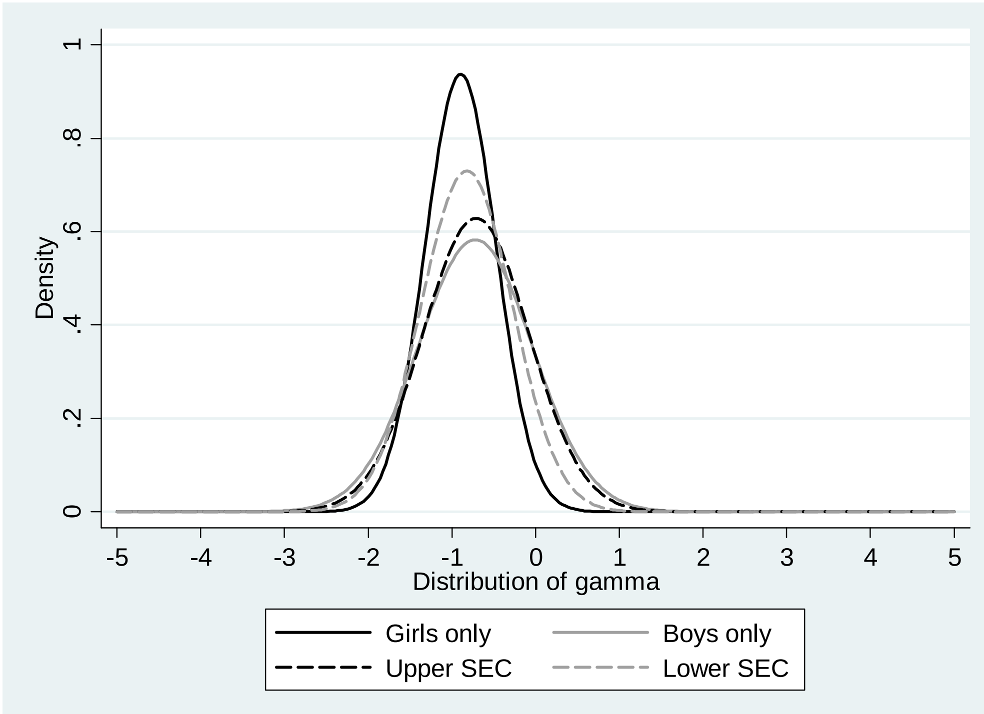
Our estimator has shown that the effect of truancy on educational outcomes at 16 is more nuanced than originally supposed. We have estimated a distributional effect of truancy over the schooling population (aged 16) and found that substantial heterogeneity exists. However, our mixed effect estimator is best used when multiple distributional effects of sub-groups are compared with each other in order to examine how the effect of truancy on attainment differs by different groups of individuals. We therefore also estimate the effect of truancy on educational attainment by gender and socio-economic class. For convenience we have suppressed the full regression output for these results and present our results in a graphical format in Figure 5.[25]

Results from Figure 5 suggest that there is little difference in the impact of truancy on educational outcomes at 16 by socio-economic class. Both individuals from poor and rich socio-economic background experience the same detrimental effect of truancy on education; in terms of the estimated average effect and how the effect is distributed around this average. However, results for gender indicates that the estimated distribution for boys is different compared to the estimated distribution for girls. In particular, girls experience a higher negative mean effect -0.894 for girls compared to -0.720 for boys) and the distribution for girls is much 'tighter' than

25 Regression results are available upon request.

that for boys (0.425 for girls compared to 0.684 for boys). The impact of truancy on educational attainment for boys is therefore more heterogeneous than for girls.

Figure 5: The Impact of Truancy on GCSE outcomes by Gender and Socio-Economic Class



5. Discussion and Conclusion

In this paper we examined the effect that truancy has on the GCSE attainment of 15/16 year old pupils in the UK. Using a modified bivariate ordered probit estimator with mixed effects we found that the causal effect of truancy on GCSE outcomes is a) more detrimental to educational attainment than it would appear if estimated by a naïve estimator (which does not take the endogeneity of truancy into account); b) truancy is a distributional effect as opposed to a common mean effect;[26] c) this distributional effect is likely to be a normally distributed effect and d) the distributional effect appears to differ by gender.

That truancy is found to be an endogenous variable and that its effect on educational attainment is highly negative is perhaps not surprising – after all, one would hardly expect a positive effect. However, to what extent can we be confident that our estimate is the *causal* impact that truancy has on educational outcomes? Simulation evidence suggests that, even if our exclusion restriction is partially correlated with education unobservables, we can infer the direction of bias on the estimates of the structural parameters of system (1), that is μ_γ, σ_γ and ρ. In our case, a more cautious interpretation would suggest that the true mean effect of truancy on education lies between the most negative estimate of the bivariate mixed estimator (-0.876) and

[26] These findings mirror research done by Arulampalam *et al.* (2008) who use quantile regression methodology to argue that a standard conditional mean estimation is an oversimplification and that the impact of truancy (at university) on attainment is more complex than originally assumed.

the least negative estimate from the naïve ordered estimator (-0.327). Moreover, our simulation results suggest that failure to account for heterogeneous treatment effects may induce additional bias into naïve estimation procedures.

No study has so far attempted a causal estimation of the effect of truancy on educational attainment at age 16 in the UK and our result further strengthens the urgent need for policy measures which prove *effective* at dealing with the truancy problem. In addition, such policy measures need to take into consideration the varying heterogeneity around the impact of truancy on education. A uniform policy targeting all truants may be inefficient as many pupils appear to suffer little detrimental effects to their GCSE scores. Conversely, other pupils experience large negative effects to their education outcomes. Any policy wishing to tackle the 'truancy problem' should possibly be flexible enough to cope with such heterogeneous treatment effects if it wishes to succeed.

What such policy measures could entail, why recent policy measure have been deemed to fail and what the determinants of truancy might be are important questions for further research which we strongly encourage. Given the importance of education and the gravity of the decision that 16 year old pupils face in the UK, it is surprising that so little research has been carried out in examining one of the most important determinants of educational attainment.

In addition to contributing to the education literature this paper also offers an important empirical contribution by outlining a bivariate ordered probit estimator with mixed effects. Such an estimator can have many different types of applications and is valid whenever both dependent variable and a suspected endogenous variable take an ordered form (including binary). In particular our estimator allows one to recover a sense of the heterogeneity surrounding the point estimates. Such estimation can be particularly useful when the number of explanatory variables is low or there is a danger of over-complicating the regression specification with too many, possibly endogenous, variables being included as additional regressors. Removing some and adding a mixed effect could be a viable option in such a case.[27] Moreover, the difficult interpretation of non-linear models (in terms of interaction terms and marginal effects) can make the mixed effects estimator an attractive proposition by visualising differing heterogeneities among sub-groups.[28]

[27] Although this comes at the cost of additional assumptions about the distributional properties of γ_i.

[28] Interested readers are encouraged to contact the authors who are happy to share and explain the Stata code behind this paper.

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

Angrist, J.D. and Pischke, J.S. (2008) *Mostly Harmless Econometrics: An Empiricist's Companion*, Princeton University Press, Princeton, New Jersey

Arulampalam, W., Naylor, R. and Smith, J. (2008) "Am I Missing Something? The Effects of Absence from Class on Student Performance", IZA Discussion Paper No. 3749

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 C. D'Ippoliti (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

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

Chen, J. and Lin, T. (2008) "Class Attendance and Exam Performance: A Randomized Experiment", *Journal of Economic Education*, Vol. 39, No. 3, pp. 213-227

Conte, A., Hey, J.D. and Moffatt, P.G. (2011) "Mixture models of choice under risk", *Journal of Econometrics*, Vol. 162, Issue 1, pp. 79-88.

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

Dobkin, C., Gil, R. and Marion, J. (2009) "Skipping Class in College and Exam Performance: Evidence from a Regression Discontinuity Classroom Experiment" *Economics of Education Review*, In Press, Corrected Proof, Available online 23rd September 2009

Durden, G. and Ellis, L. (1995) "The Effects of Attendance on Student Learning in Principles of Economics", *American Economic Review*, Vol. 85, No. 2, pp. 343–46.

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

Farrington, D. (1996). "Later Life Outcomes of Truants in the Cambridge Study", in *Unwillingly to School 4th edition*, (eds: Ian Berg & Jean Nursten). Gaskell Press, pp. 96-118.

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)

Harmon, C., Oosterbeek, H. and Walker, I. (2003) "The Returns to Education: Microeconomics", *Journal of Economic Surveys*, Vol. 17, No. 2, pp. 115-155

Heilbrunn, J.Z. (2007) "Piece of the Truancy Jigsaw: A Literature Review", an initiative by the Colorado Foundation for Families and Children, available at www.schoolengagement.org

Hibbert, A. and Fogelman, K. (1990), "Future Lives of Truants: Family Formation and Health Related Behaviour" *British Journal of Educational Psychology*, Vol. 60, No. 2, pp. 171-179.

Marburger, D.R. (2001) "Absenteeism and Undergraduate Exam Performance", *Journal of Economic Education*, Vol. 32, No. 2, pp. 99-110

Kirby, A. and McElroy, B. (2003). "The effect of attendance on grade for first year Economics students in University College Cork". *The Economic and Social Review*, Vol. 34, No. 3, pp. 311-326.

Manksi, C.F. (1993) "Identification of Endogenous Social Effects: The Reflection Problem", *Review of Economic Studies*, Vol. 60, pp. 531-542

Nevo, A. and Rosen, A. (2008) "Identification with Imperfect Instruments", cemmap working paper CWP16/18, The Institute for Fiscal Studies, London

Payne, J. (2001) "Patterns of Participation in Full-Time Education after 16: An Analysis of the England and Wales Youth Cohort Study", Department for Education and Skills Research Report, RR307

Prior, D. and Paris, A. (2005) *Preventing Children's Involvement in Crime and Anti-social Behaviour: A Literature Review*. Birmingham: DfES

Raffe, D. (1986) "Unemployment and School Motivation: The Case of Truancy", *Educational Review*, Vol. 38, pp. 11-19

Romer, D. (1993) "Do Students Go To Class? Should They?" *Journal of Economic Perspectives*, Vol. 7, pp. 167-174

Sajaia, Z. (2008) "Maximum Likelihood Estimation of a Bivariate Ordered Probit Model: Implementation and Monte Carlo Simulations", World Bank Applied Economic Research Papers

Stanca, L. (2006) "The Effects of Attendance on Academic Performance: Panel Data Evidence for Introductory Microeconomics", *Journal of Economic Education,* Vol. 37, No. 3, pp. 251-266

Select Committee on Public Accounts (2006) *Progress in improving school attendance* House of Commons Public Accounts – Eighteenth Report

Appendix A – Derivation of the reduced form

Let us consider the following transformation of the recursive system in eq. (1), where $y_{1i}^* = T_i^*$ and $y_{2i}^* = E_i^*$:

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

that in matrix form is:

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

where:

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

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

$$Y^* = \Lambda X + \Upsilon\,, \quad \text{(A4)}$$

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:

$$\Pi Y^* = \Pi\Lambda X + \Pi\Upsilon\,. \quad \text{(A5)}$$

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

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

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:

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

Appendix B – Monte Carlo Simulations

To examine the small sample properties of our estimator we implement Monte Carlo simulations. The latent variables y_{1i}^* and y_{2i}^* for the bivariate ordered probit with mixed effects model are generated by the following data generating process:

$$\begin{cases} y_{1i}^* = 1x_{1i} + 1z_{1i} + \varepsilon_{1i} \\ y_{2i}^* = \gamma_i y_{1i}^* - 2.5x_{1i} + \varepsilon_{2i} \end{cases}, \qquad \text{(B1)}$$

where $\gamma_i \sim N(\mu_\gamma = 0.5, \sigma_\gamma = 0.5)$. We generate x_{1i} and z_{1i} as independent standard normal random variables and ε_{1i} and ε_{2i} as bivariate standard normal random variables with correlation ρ. The observational rule is:

$$y_{1i} = \begin{cases} 1 & if\ y_{1i}^* \leq -2 \\ 2 & if -2 < y_{1i}^* \leq 0.5 \\ 3 & if\ 0.5 < y_{1i}^* \end{cases} \qquad y_{2i} = \begin{cases} 1 & if\ y_{2i}^* \leq -3 \\ 2 & if -3 < y_{2i}^* \leq -1 \\ 3 & if -1 < y_{2i}^* \leq 1 \\ 4 & if\ 1 < y_{2i}^* \end{cases}, \qquad \text{(B2)}$$

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 authors are happy to share and explain the associated Stata .do files in order to replicate the simulation results and/or to modify our estimator for personal use.

Table C1: Simulation results

Bivariate ordered probit with mixed effects

			Obs = 200								*Obs = 500*			
Rho	μ_γ	RMSE	σ_γ	RMSE	ρ	RMSE		Rho	μ_γ	RMSE	σ_γ	RMSE	ρ	RMSE
-0.9	0.4971	0.0729	0.4723	0.1082	-0.8943	0.0502		-0.9	0.4997	0.0436	0.4994	0.0647	-0.9021	0.0324
-0.5	0.5092	0.1000	0.5121	0.1599	-0.5015	0.1166		-0.5	0.5043	0.0587	0.5068	0.0915	-0.5029	0.0721
0	0.5346	0.1455	0.5647	0.2184	0.0124	0.1518		0	0.5067	0.0757	0.5084	0.1091	0.0015	0.0964
0.5	0.5256	0.1489	0.5254	0.1781	0.5259	0.1387		0.5	0.5116	0.0924	0.5104	0.1066	0.5081	0.0785
0.9	0.5006	0.1504	0.4589	0.1387	0.8823	0.0591		0.9	0.4965	0.0928	0.4858	0.0845	0.9008	0.0404
			Obs = 1000								*Obs = 5000*			
Rho	μ_γ	RMSE	σ_γ	RMSE	ρ	RMSE		Rho	μ_γ	RMSE	σ_γ	RMSE	ρ	RMSE
-0.9	0.5007	0.0324	0.5018	0.0468	-0.9033	0.0234		-0.9	0.4996	0.0140	0.4979	0.0198	-0.8997	0.0098
-0.5	0.5022	0.0434	0.4997	0.0642	-0.5027	0.0517		-0.5	0.4995	0.0186	0.4976	0.0281	-0.4996	0.0230
0	0.5043	0.0528	0.4998	0.0760	-0.0030	0.0656		0	0.4998	0.0243	0.4980	0.0341	-0.0005	0.0288
0.5	0.5047	0.0590	0.4994	0.0750	0.5001	0.0575		0.5	0.5008	0.0278	0.4989	0.0327	0.4991	0.0248
0.9	0.5066	0.0666	0.5004	0.0610	0.9029	0.0321		0.9	0.5009	0.0292	0.5000	0.0276	0.9016	0.0138

Monte Carlo Simulations: 1000 Replications

Appendix C – Descriptive statistics

Table B1: Descriptive Statistics

Variable	Type		Obs.	Freq	Wght. Freq
Gender	Binary	0 = boys	5,890	45.64	50.66
		1 = girls	7,016	54.36	49.34
Ethnicity	Binary	0 = white	11,190	86.70	85.46
		1 = non-white	1,717	13.30	14.54
Parental Employment	Binary	0 = otherwise	10,780	83.53	80.71
		1 = neither parent empl.	2,126	16.47	19.29
Parent Education	Categorical	0 = 1 parent has degree level	3,299	25.56	21.80
		1 = 1 parent has a-level	1,998	15.48	14.76
		2 = both parents have no a-level	7,609	58.96	63.44
School Type	Categorical	0 = comprehensive school	10,941	84.77	88.81
		1 = selective/grammar school	746	5.78	4.08
		2 = independent school	1,219	9.45	7.10
Housing	Categorical	0 = owned by parents	10,288	79.71	75.37
		1 = rented	2,178	16.88	20.82
		2 = other	440	3.41	3.80
Disability	Binary	0 = non-disabled	12,087	95.7	94.88
		1 = disabled	543	4.30	5.12
Socio-Economic Class	Categorical	0 = higher professional	2,640	20.46	17.22
		1 = lower professional	3,629	28.12	25.89
		2 = intermediate	2,629	20.37	20.4
		3 = lower supervisory	1,210	9.38	10.5
		4 = semi-routine and routine	1,488	11.53	13.57
		5 = other	1,310	10.105	12.41
Household Composition	Binary	0 = otherwise	2,768	21.45	23.25
		1 = both parents in household	10,139	78.55	76.75
Parental Involvement in School	Categorical	0 = never	861	6.67	5.68
		1 = not often	916	7.10	8.12
		2 = sometimes	3,656	28.33	28.45
		3 = often	7,473	57.90	57.75
Sweep 1 response type	Categorical	0 = 1st questionnaire	7,415	57.45	54.99
		1 = 2nd questionnaire	2,553	19.78	20.82
		2 = 3rd questionnaire	1,199	9.29	10.07
		3 = Telephone	1,739	13.47	14.12
			Mean	*Std.*	*Min / Max*
Local Truancy Rate	Continuous	148 LEA averages	1.068	0.599	0.30 / 3.90
Local Education Rate (5ABCs)	Continuous	148 LEA averages	0.472	0.076	0.23 / 0.61
Local Part-time Median Hourly Pay	Continuous	148 LEA averages	5.203	0.548	4.33 / 7.83
Local Full-time Median Hourly Pay	Continuous	148 LEA averages	8.396	1.049	6.70 / 13.71
Local Unemployment Rate	Continuous	148 LEA averages	3.280	1.573	0.70 / 9.30
ASBO per 1000 inhabitants 1999-2003	Continuous	148 LEA averages	0.047	0.031	0.01 / 0.14
ASBO breaches per 1000 inhabitants 1999-2003	Continuous	148 LEA averages	0.048	0.039	0.00 / 0.19
Urban Share of LEA	Continuous	148 LEA averages	93.25	11.36	0.00 / 100

YCS 10, sweep 1, England sample only.

Appendix D – Naive identification approach

Variables		BIOP							BIOPmixed						
Dependen	Independent	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Truancy	*Possible exclusion restrictions*														
	Log local truancy rate (unauthorised absence) 1999	0.109***							0.092***						
		(0.027)							(0.025)						
	Log local part-time median hourly pay rate 1999		0.313*							0.200*					
			(0.131)							(0.100)					
	Log local full-time median hourly pay rate 1999			0.028							-0.010				
				(0.059)							(0.010)				
	Log unemployment rate 1999				-0.064							-0.021			
					(0.046)							(0.030)			
	ASBO per 1000 inhabitants 1999-2003					0.539							0.543		
						(0.689)							(0.581)		
	ASBO breach per 1000 inhabitants 1999-2003						0.668							0.699	
							(0.481)							(0.474)	
	Urban share of LEA 2005							0.013							-0.002
								(0.013)							(0.001)
	Output for controls ommitted														
Education															
	μ_γ	-0.664**	0.371	0.941***	-0.563	-0.766	-0.756	-0.155	-0.824***	0.597*	-1.004***	0.751	-1.044***	-0.941***	0.497
		(0.211)	(0.313)	(0.133)	(0.622)	(0.853)	(0.461)	(0.974)	(0.209)	(0.287)	(0.217)	(0.460)	(0.248)	(0.347)	(0.403)
	σ_γ	n/a	n/a	n/a	n/a	n/a	n/a	n/a	0.579***	0.353*	0.201	0.225	0.406	0.510*	0.358*
		n/a	n/a	n/a	n/a	n/a	n/a	n/a	(0.097)	(0.180)	(0.288)	(0.317)	(0.308)	(0.239)	(0.145)
	ρ	0.345	-0.641***	-0.990***	0.247	0.474	0.462	-0.169	0.431	-0.836***	0.875***	-0.929***	0.912	0.586	-0.792***
		(0.271)	(0.241)	(0.037)	(0.720)	(1.039)	(0.557)	(0.950)	(0.260)	(0.172)	(0.304)	(0.206)	(0.604)	(0.509)	(0.247)
	N	12537	12537	12537	12537	12537	12537	12537	12537	12537	12537	12537	12537	12537	12537

Standard errors in parentheses: * p<0.05, ** p<0.01, *** p<0.001

Appendix E – STATA Programme

```
clear
set seed 1234

*Simulate data
scalar rho=0.5

matrix C = (1, rho\ rho, 1)
matrix m = (0,0)
drawnorm e1 e2, n(1000) corr(C) means(m)

gen x1 = invnorm(uniform())
gen x2 = invnorm(uniform())
gen x3 = invnorm(uniform())
gen x4 = invnorm(uniform())
gen z1 = invnorm(uniform())

gen gamma=0.8+0.5*invnorm(uniform())

gen y1star =                         x1 + x2 + x3 + x4 + z1 + e1
gen y2star = gamma*y1star + x1 + x2 + x3 + x4      + e2

gen y1 = 1
replace y1=2 if y1star>-2 & y1star<=-1
replace y1=3 if y1star>-1 & y1star<=0
replace y1=4 if y1star>0 & y1star<=1
replace y1=5 if y1star>1

gen y2 = 1
replace y2=2 if y2star>-2 & y2star<=-1
replace y2=3 if y2star>-1 & y2star<=0
replace y2=4 if y2star>0 & y2star<=1
replace y2=5 if y2star>1

*Bivariate ordered-ordered probit with mixed gaussian effect

gen m0=(y1==1&y2==1)
gen m1=(y1==1&y2==2)
gen m2=(y1==1&y2==3)
gen m3=(y1==1&y2==4)
gen m4=(y1==1&y2==5)

gen m5=(y1==2&y2==1)
gen m6=(y1==2&y2==2)
gen m7=(y1==2&y2==3)
gen m8=(y1==2&y2==4)
gen m9=(y1==2&y2==5)

gen m10=(y1==3&y2==1)
gen m11=(y1==3&y2==2)
gen m12=(y1==3&y2==3)
gen m13=(y1==3&y2==4)
gen m14=(y1==3&y2==5)

gen m15=(y1==4&y2==1)
gen m16=(y1==4&y2==2)
gen m17=(y1==4&y2==3)
gen m18=(y1==4&y2==4)
gen m19=(y1==4&y2==5)

gen m20=(y1==5&y2==1)
gen m21=(y1==5&y2==2)
gen m22=(y1==5&y2==3)
gen m23=(y1==5&y2==4)
gen m24=(y1==5&y2==5)
```

```
gen double a1=.
replace a1=-2.453407083009012499e-01
gen double a2=.
replace a2=-1.234076215395323007e+00
gen double a3=.
replace a3=-2.254974002089275523e+00
gen double a4=.
replace a4=-3.347854567383216326e+00
gen double a5=.
replace a5=-4.603682449550744272e+00
gen double a6=.
replace a6=-7.374737285453943587e-01
gen double a7=.
replace a7=-1.738537712116586206e+00
gen double a8=.
replace a8=-2.788806058428130480e+00
gen double a9=.
replace a9=-3.944764040115625210e+00
gen double a10=.
replace a10=-5.387480890011232861e+00
gen double a11=.
replace a11=2.453407083009012499e-01
gen double a12=.
replace a12=1.234076215395323007e+00
gen double a13=.
replace a13=2.254974002089275523e+00
gen double a14=.
replace a14=3.347854567383216326e+00
gen double a15=.
replace a15=4.603682449550744272e+00
gen double a16=.
replace a16=7.374737285453943587e-01
gen double a17=.
replace a17=1.738537712116586206e+00
gen double a18=.
replace a18=2.788806058428130480e+00
gen double a19=.
replace a19=3.944764040115625210e+00
gen double a20=.
replace a20=5.387480890011232861e+00


gen double w1=.
replace w1=4.622436696006100896e-01
gen double w2=.
replace w2=1.090172060200233200e-01
gen double w3=.
replace w3=3.243773342237861832e-03
gen double w4=.
replace w4=7.802556478532063693e-06
gen double w5=.
replace w5=4.399340992273180553e-10
gen double w6=.
replace w6=2.866755053628341297e-01
gen double w7=.
replace w7=2.481052088746361088e-02
gen double w8=.
replace w8=2.283386360163539672e-04
gen double w9=.
replace w9=1.086069370769281693e-07
gen double w10=.
replace w10=2.229393645534151292e-13
gen double w11=.
replace w11=4.622436696006100896e-01
gen double w12=.
replace w12=1.090172060200233200e-01
gen double w13=.
replace w13=3.243773342237861832e-03
gen double w14=.
replace w14=7.802556478532063693e-06
gen double w15=.
replace w15=4.399340992273180553e-10
```

```
gen double w16=.
replace w16=2.866755053628341297e-01
gen double w17=.
replace w17=2.481052088746361088e-02
gen double w18=.
replace w18=2.283386360163539672e-04
gen double w19=.
replace w19=1.086069370769281693e-07
gen double w20=.
replace w20=2.229393645534151292e-13


global draws "20"
cap drop l logl r
gen l=0
gen logl = 0
gen r = 0




program drop _all

program sm_ml_lf
        args lnf xb1 c11 c12 c13 c14 xb2 c21 c22 c23 c24  mg sg r

        tempvar ll Y1 Y2 lj rho eta rt sigmag g L1 L2

        quietly{
        /*components to be indirectly estimated*/
                        gen double `rho'= tanh(`r')
                        gen double `sigmag'= exp(`sg')

                        gen double `Y1'=0
                        gen double `Y2'=0
                        gen double `eta'=0
                        gen double `rt'=0
                        gen double `g'=0

                        /*partial likelihoods*/
                        gen double `ll'=0
                        gen double `L1'=0
                        gen double `L2'=0


                forvalues r=1/$draws{
                quietly {
                replace `g'=`mg'+sqrt(2)*`sigmag'*a`r'
                replace `eta'=1/sqrt(1+2*`g'*`rho'+`g'^2)
                replace `rt'=`eta'*(`g'+`rho')

                replace `Y1'=`xb1'
                replace `Y2'=`g'*`Y1'+`xb2'

                replace `ll' = (binormal((`c11'-`Y1'),`eta'*(`c21'-`Y2'),`rt')) if m0==1
                replace `ll' = (binormal((`c11'-`Y1'),`eta'*(`c22'-`Y2'),`rt')-binormal((`c11'-
`Y1'),`eta'*(`c21'-`Y2'),`rt')) if m1==1
                replace `ll' = (binormal((`c11'-`Y1'),`eta'*(`c23'-`Y2'),`rt')-binormal((`c11'-
`Y1'),`eta'*(`c22'-`Y2'),`rt')) if m2==1
                replace `ll' = (binormal((`c11'-`Y1'),`eta'*(`c24'-`Y2'),`rt')-binormal((`c11'-
`Y1'),`eta'*(`c23'-`Y2'),`rt')) if m3==1
                replace `ll' = ((normal((`c11'-`Y1'))-binormal((`c11'-`Y1'),`eta'*(`c24'-
`Y2'),`rt'))) if m4==1

                replace `ll' = (binormal((`c12'-`Y1'),`eta'*(`c21'-`Y2'),`rt')-binormal((`c11'-
`Y1'),`eta'*(`c21'-`Y2'),`rt')) if m5==1
                replace `ll' = (binormal((`c12'-`Y1'),`eta'*(`c22'-`Y2'),`rt')-binormal((`c11'-
`Y1'),`eta'*(`c22'-`Y2'),`rt')-binormal((`c12'-`Y1'),`eta'*(`c21'-`Y2'),`rt')+binormal((`c11'-
`Y1'),`eta'*(`c21'-`Y2'),`rt')) if m6==1
                replace `ll' = (binormal((`c12'-`Y1'),`eta'*(`c23'-`Y2'),`rt')-binormal((`c11'-
`Y1'),`eta'*(`c23'-`Y2'),`rt')-binormal((`c12'-`Y1'),`eta'*(`c22'-`Y2'),`rt')+binormal((`c11'-
`Y1'),`eta'*(`c22'-`Y2'),`rt')) if m7==1
```

```
		replace `ll' = (binormal((`c12'-`Y1'),`eta'*(`c24'-`Y2'),`rt')-binormal((`c11'-
`Y1'),`eta'*(`c24'-`Y2'),`rt')-binormal((`c12'-`Y1'),`eta'*(`c23'-`Y2'),`rt')+binormal((`c11'-
`Y1'),`eta'*(`c23'-`Y2'),`rt')) if m8==1
		replace `ll' = (normal((`c12'-`Y1'))-normal((`c11'-`Y1'))-binormal((`c12'-
`Y1'),`eta'*(`c24'-`Y2'),`rt')+binormal((`c11'-`Y1'),`eta'*(`c24'-`Y2'),`rt')) if m9==1

		replace `ll' = (binormal((`c13'-`Y1'),`eta'*(`c21'-`Y2'),`rt')-binormal((`c12'-
`Y1'),`eta'*(`c21'-`Y2'),`rt')) if m10==1
		replace `ll' = (binormal((`c13'-`Y1'),`eta'*(`c22'-`Y2'),`rt')-binormal((`c12'-
`Y1'),`eta'*(`c22'-`Y2'),`rt')-binormal((`c13'-`Y1'),`eta'*(`c21'-`Y2'),`rt')+binormal((`c12'-
`Y1'),`eta'*(`c21'-`Y2'),`rt')) if m11==1
		replace `ll' = (binormal((`c13'-`Y1'),`eta'*(`c23'-`Y2'),`rt')-binormal((`c12'-
`Y1'),`eta'*(`c23'-`Y2'),`rt')-binormal((`c13'-`Y1'),`eta'*(`c22'-`Y2'),`rt')+binormal((`c12'-
`Y1'),`eta'*(`c22'-`Y2'),`rt')) if m12==1
		replace `ll' = (binormal((`c13'-`Y1'),`eta'*(`c24'-`Y2'),`rt')-binormal((`c12'-
`Y1'),`eta'*(`c24'-`Y2'),`rt')-binormal((`c13'-`Y1'),`eta'*(`c23'-`Y2'),`rt')+binormal((`c12'-
`Y1'),`eta'*(`c23'-`Y2'),`rt')) if m13==1
		replace `ll' = (normal((`c13'-`Y1'))-normal((`c12'-`Y1'))-binormal((`c13'-
`Y1'),`eta'*(`c24'-`Y2'),`rt')+binormal((`c12'-`Y1'),`eta'*(`c24'-`Y2'),`rt')) if m14==1

		replace `ll' = (binormal((`c14'-`Y1'),`eta'*(`c21'-`Y2'),`rt')-binormal((`c13'-
`Y1'),`eta'*(`c21'-`Y2'),`rt')) if m15==1
		replace `ll' = (binormal((`c14'-`Y1'),`eta'*(`c22'-`Y2'),`rt')-binormal((`c13'-
`Y1'),`eta'*(`c22'-`Y2'),`rt')-binormal((`c14'-`Y1'),`eta'*(`c21'-`Y2'),`rt')+binormal((`c13'-
`Y1'),`eta'*(`c21'-`Y2'),`rt')) if m16==1
		replace `ll' = (binormal((`c14'-`Y1'),`eta'*(`c23'-`Y2'),`rt')-binormal((`c13'-
`Y1'),`eta'*(`c23'-`Y2'),`rt')-binormal((`c14'-`Y1'),`eta'*(`c22'-`Y2'),`rt')+binormal((`c13'-
`Y1'),`eta'*(`c22'-`Y2'),`rt')) if m17==1
		replace `ll' = (binormal((`c14'-`Y1'),`eta'*(`c24'-`Y2'),`rt')-binormal((`c13'-
`Y1'),`eta'*(`c24'-`Y2'),`rt')-binormal((`c14'-`Y1'),`eta'*(`c23'-`Y2'),`rt')+binormal((`c13'-
`Y1'),`eta'*(`c23'-`Y2'),`rt')) if m18==1
		replace `ll' = (normal((`c14'-`Y1'))-normal((`c13'-`Y1'))-binormal((`c14'-
`Y1'),`eta'*(`c24'-`Y2'),`rt')+binormal((`c13'-`Y1'),`eta'*(`c24'-`Y2'),`rt')) if m19==1

		replace `ll' = (normal(`eta'*(`c21'-`Y2'))- binormal((`c14'-`Y1'),`eta'*(`c21'-
`Y2'),`rt')) if m20==1
		replace `ll' = (normal(`eta'*(`c22'-`Y2'))-binormal((`c14'-`Y1'),`eta'*(`c22'-
`Y2'),`rt')-normal(`eta'*(`c21'-`Y2'))+binormal((`c14'-`Y1'),`eta'*(`c21'-`Y2'),`rt')) if
m21==1
		replace `ll' = (normal(`eta'*(`c23'-`Y2'))-binormal((`c14'-`Y1'),`eta'*(`c23'-
`Y2'),`rt')-normal(`eta'*(`c22'-`Y2'))+binormal((`c14'-`Y1'),`eta'*(`c22'-`Y2'),`rt')) if
m22==1
		replace `ll' = (normal(`eta'*(`c24'-`Y2'))-binormal((`c14'-`Y1'),`eta'*(`c24'-
`Y2'),`rt')-normal(`eta'*(`c23'-`Y2'))+binormal((`c14'-`Y1'),`eta'*(`c24'-`Y2'),`rt')) if
m23==1
		replace `ll' = (1-normal((`c14'-`Y1'))-normal(`eta'*(`c24'-
`Y2'))+binormal((`c14'-`Y1'),`eta'*(`c24'-`Y2'),`rt')) if m24==1

		replace `L1'=(1/sqrt(2*asin(1)))*w`r'*`ll'
		replace `L2'=`L2'+`L1'
				}
					}
		}


		quietly replace `lnf'=ln(`L2') if $ML_y1 ==1 & $ML_y2 == 1
		quietly replace `lnf'=ln(`L2') if $ML_y1 ==1 & $ML_y2 == 2
		quietly replace `lnf'=ln(`L2') if $ML_y1 ==1 & $ML_y2 == 3
		quietly replace `lnf'=ln(`L2') if $ML_y1 ==1 & $ML_y2 == 4
		quietly replace `lnf'=ln(`L2') if $ML_y1 ==1 & $ML_y2 == 5

		quietly replace `lnf'=ln(`L2') if $ML_y1 ==2 & $ML_y2 == 1
		quietly replace `lnf'=ln(`L2') if $ML_y1 ==2 & $ML_y2 == 2
		quietly replace `lnf'=ln(`L2') if $ML_y1 ==2 & $ML_y2 == 3
		quietly replace `lnf'=ln(`L2') if $ML_y1 ==2 & $ML_y2 == 4
		quietly replace `lnf'=ln(`L2') if $ML_y1 ==2 & $ML_y2 == 5

		quietly replace `lnf'=ln(`L2') if $ML_y1 ==3 & $ML_y2 == 1
		quietly replace `lnf'=ln(`L2') if $ML_y1 ==3 & $ML_y2 == 2
		quietly replace `lnf'=ln(`L2') if $ML_y1 ==3 & $ML_y2 == 3
		quietly replace `lnf'=ln(`L2') if $ML_y1 ==3 & $ML_y2 == 4
		quietly replace `lnf'=ln(`L2') if $ML_y1 ==3 & $ML_y2 == 5
```

```
                quietly replace `lnf'=ln(`L2') if $ML_y1 ==4 & $ML_y2 == 1
                quietly replace `lnf'=ln(`L2') if $ML_y1 ==4 & $ML_y2 == 2
                quietly replace `lnf'=ln(`L2') if $ML_y1 ==4 & $ML_y2 == 3
                quietly replace `lnf'=ln(`L2') if $ML_y1 ==4 & $ML_y2 == 4
                quietly replace `lnf'=ln(`L2') if $ML_y1 ==4 & $ML_y2 == 5

                quietly replace `lnf'=ln(`L2') if $ML_y1 ==5 & $ML_y2 == 1
                quietly replace `lnf'=ln(`L2') if $ML_y1 ==5 & $ML_y2 == 2
                quietly replace `lnf'=ln(`L2') if $ML_y1 ==5 & $ML_y2 == 3
                quietly replace `lnf'=ln(`L2') if $ML_y1 ==5 & $ML_y2 == 4
                quietly replace `lnf'=ln(`L2') if $ML_y1 ==5 & $ML_y2 == 5

exit

end


*Starting values
        set more off
        set seed 12347
        tab y1
        tab y2

        mat drop _all
        xi: oprobit y1 x1 x2 x3 x4 z1
        mat b1 = e(b)
        mat coleq b1 = y1
        xi: oprobit y2 y1 x1 x2 x3 x4
        mat c2 = e(b)
        mat list c2
        local last  = colsof(c2)
        mat b2 = c2[1,2.. `last']
        mat coleq b2 = y2
        mat b3 = -1,-1,0.5
        mat coleq b3 = mg/sg/rho
        mat start = b1, b2, b3
        mat list start

*Estimate
xi: ml model lf sm_ml_lf  (y1 = x1 x2 x3 x4 z1, noconstant ) (c11:) (c12:) (c13:) (c14:) (y2 =
x1 x2 x3 x4 , noconstant) (c21:) (c22:) (c23:) (c24:) (mg:) (sg:) (r:) , technique(bhhh)
ml init start, copy
ml max, iterate(200) difficult trace

nlcom (meang: ([mg]_b[_cons]))
nlcom (sigmag: exp([sg]_b[_cons]))
nlcom (rho: tanh([r]_b[_cons]))

scalar mg0 =  ([mg]_b[_cons])
scalar sg0 = exp([sg]_b[_cons])
scalar rho0= tanh([r]_b[_cons])

local mg0 = mg0
tw (fn y =normalden(x,mg0,sg0), range(-5 5) xline(`mg0') subtitle("The impact of y1 on y2")
ytitle("Density") xtitle("Distribution of Gamma " "Mean =" `mg0' ) xlabel(-5(1)5,grid)
ylabel(,grid))
```