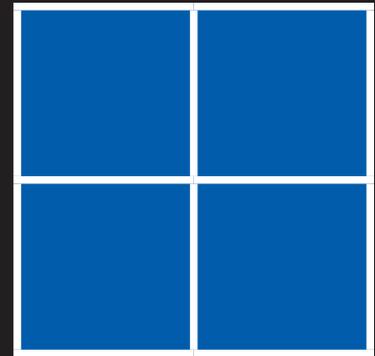


The Inter-generational Effect of Education: Controlling for Bias due to Compulsory Schooling

Richard Dorsett and Martin Weale

LLAKES Research Paper 57



Centre for Learning and Life Chances in Knowledge Economies and Societies

LLAKES is an ESRC-funded Research Centre - grant reference ES/J019135/1.

To request printed copies of this paper or other LLAKES Research Papers, please contact the LLAKES Centre - llakescentre@ioe.ac.uk.

To view this and other LLAKES Research Papers online, please visit www.llakes.ac.uk.

LLAKES Research Papers – Copyright 2017

This paper may be cited or briefly quoted in line with the usual academic conventions, and for personal use. However, this paper must not be published elsewhere (such as mailing lists, bulletin boards etc.) without the authors' explicit permission.

If you copy this paper, you must:

- include this copyright note.
- not use the paper for commercial purposes or gain in any way.
- observe the conventions of academic citation in a version of the following:

Dorsett, R. and Weale, M. (2017) *The Inter-generational Effect of Education: Controlling for Bias due to Compulsory Schooling*, published by the Centre for Learning and Life Chances in Knowledge Economies and Societies at: <http://www.llakes.ac.uk>

The Inter-generational Effect of Education: Controlling for Bias due to Compulsory Schooling

Richard Dorsett and Martin Weale

**Centre for Learning and Life Chances in Knowledge Economies and Societies
(LLAKES),
UCL Institute of Education, University College London**

List of Contents

Abstract.....	1
1 Introduction.....	2
2 Instrumental Variable Estimating and Censoring.....	4
3 The Bias when Variables are Normally Distributed.....	6
4 Empirical Results.....	8
4.1 Data.....	8
4.2 IV Estimates.....	10
4.3 Estimates Corrected for Censoring.....	12
4.4 Allowing for Normality using an Ordered Probit Model.....	15
5. Conclusions.....	20
References.....	21

List of Tables

Table 1	Father's Age of Completing Education and Grandfather's Social Class.....	9
Table 2	Child's Age of Completing Education and Grandfather's Social Class.....	9
Table 3	Determinants of the Probability of an Initial Respondent Remaining in our sample.....	10
Table 4	IV Coefficient Estimates as Functions of the Cut Point for the Dummy Instrumental Variable.....	11
Table 5	Parameter Estimates allowing for Censoring when Child's Age of Completion is influenced by Father's Latent Age of Completion.....	13
Table 6	Parameter Estimates generated by a Censored Normal Distribution.....	15
Table 7	The Parameters of the Ordered Probit Model.....	18
Table 8	Estimates of the Average Marginal Impact of an Extension of Father's Education on that of Children.....	20

The Inter-generational Effect of Education: Controlling for the Bias due to Compulsory Schooling

Richard Dorsett* and Martin Weale†

15th March 2017

Abstract

This paper explores the coefficient linking educational attainment of fathers to that of their children. Standard estimation approaches ignore the left-censoring that arises due to compulsory participation and which affects both generations. This paper formalises the bias introduced by censoring and identifies conditions under which simple estimators may still deliver consistent results. Biases depend on the degree of censoring in both child and parental education as well as the choice of instrument. Empirical results using British data follow the theoretical expectations and reveal linear IV estimates to be substantially upward-biased. In a further extension, we handle the non-normality of the non-censored part of the education distribution using an ordered probit model. This preferred approach delivers an average marginal effect which is smaller still. In addition to this substantive finding, our results also have general implications for the interpretation of instrumental variable estimates and provide a potential explanation for results varying according to choice of instrument that is distinct from the usual attribution to impact heterogeneity.

*University of Westminster; LLAKES Centre for Research on Learning and Life Chances, UCL Institute of Education, University College London

†King's College, London; LLAKES Centre for Research on Learning and Life Chances, UCL Institute of Education, University College London; Centre for Macroeconomics, London School of Economics. This research was supported by the LLAKES Centre for Learning and Life Chances in Knowledge Economies and Societies, an ESRC-funded Research Centre (grant reference ES/J019135/1). The British Household Panel Survey, used in this study, were made available through the ESRC Data Archive. The data were originally collected by the ESRC Research Centre on Micro-social Change at the University of Essex (now incorporated within the Institute for Social and Economic Research). Neither the original collectors of the data nor the Archive bear any responsibility for the analyses or interpretations presented here. Corresponding author: martin.weale@outlook.com.

1 Introduction

Many situations arise in which econometricians analyse censored data. Censoring may arise in two forms. In the first, and perhaps more widely considered case, it arises because the reported data differ from the true outcome. For example data on age of completion of education may not distinguish those who complete their education beyond the age of say twenty-three from those who complete it at the age of twenty-three. In this case any analysis which ignores censoring is complicated by what is in effect measurement error. A second form of censoring arises because one group in the population may be subject to different behavioural influences than another group in the population. For example, in a country with a minimum school-leaving age children may stay at school to the minimum age not because they want to but because they have to. It follows that any analysis of the relationship between the education of parents and that of their children should take account of the distortions which might arise from compulsion.

A substantial survey of work on the connection between parents' and children's education is provided by Holmlund, Lindahl & Plug (2011) following an earlier account by Haveman & Wolfe (1995). They bring together a wide body of work, and discuss at length the issue of identification; how to separate the effects of parents' education on that of their children from other familial influences. They discuss in particular two means of doing this; first, as discussed by Dearden, Machin & Reed (1997) and more fully by Plug (2004), it is possible to study the issue for adopted children. Twou, Liu & Hammitt (2012) also follow this approach which is intended to ensure that the influence of inherited genetic effects is removed. As Holmlund et al. (2011) point out this does, however face the objection that adoption may itself be selective. A second route is to study the children of (ideally identical) twin sisters. In this case the focus is on whether differences in the educational attainment of the twins is connected with differences in the educational attainment of their own children, with the aim of differencing out genetic influences. Of course there remains the question whether the genetic material of the children's fathers is correlated with the educational attainment of their mothers.

Other work (Oreopoulos, Page & Stevens 2006) has looked at changes in the compulsory education of parents on their children. This is obviously a topic of interest in its own right given the importance of compulsory education in advanced economies. At the same time it brings to the fore the question of how best to deal with the effects of compulsion. Right-censoring of data on years of education can arise because the data are collected before some of the respondents have completed their education. This was

explored by de Haan & Plug (2011). Similar, and probably more important issues arise with left-censoring as a result of compulsion.

This is a very material issue; one of the data sets widely used to explore the connection between parents and children's education in the United Kingdom, the British Cohort Survey suggests, before any reweighting to correct for non-response, that 59% of the parents and 45% of the children left school at the school-leaving age rather than obviously at the time of their own choosing. An analytical framework which assumes that the age of completion of education of a child is a linear function of that of its parents plus a random term will mislead if in fact for some this is actually the result of compulsion. Things are even more complicated if, as our data suggest, compulsory schooling is not transmitted across generations in the same way as voluntary schooling. Rigobon & Stoker (2009) point out that, in such circumstances the assumptions required for the IV estimate to be interpretable as a local average treatment effect (Imbens & Angrist 1994) do not hold.

Rigobon & Stoker (2009) discuss the biases which arise from censoring in both OLS and instrumental variable regression (following Austin & Hoch (2004) who looked at OLS regression) when the explanatory variable is censored. In the situation we face both the explanatory variable and the dependent variable are censored. We show in these circumstances that the use of instrumental variables can serve to reduce the biases which arise from censoring, and discuss circumstances in which IV estimates, unlike their OLS counterparts are likely to be reasonably robust. The outcome depends on the nature of the instrument as well as the extent to which the dependent and endogenous explanatory variables are censored.

We proceed as follows. We begin by setting out the effects of censoring on parameter estimates when both explanatory and dependent variable are censored, identifying the way in which biases can offset each other. We explore this further making the assumption of normality. We then present our data drawn from the 1970 British Cohort Survey. This shows the ages at which the fathers of the children studied in the 1970 survey completed their education with similar data for their children. Information on the social class of the child's paternal grandfather provides a set of instruments whose use passes tests for both under- and over-identification. Using the parameters of our model estimated under the assumption of normality, we explore the biases likely to arise from IV estimation. We show that IV methods tend to overstate the influence of parental age of completing education on that of children and that the main source of the bias is likely to be the

fact that a higher proportion of fathers than children completed their education at the statutory minimum age. Our theoretical analysis points to a connection between the choice of instrument and the estimated IV parameter which matches closely what is observed with our data.

This analysis is, however, subject to the criticism that it assumes educational age is normally distributed above the censoring point. Use of an ordered probit model allows us to relax this assumption. The results from this point to an estimate that is slightly smaller than that obtained under the assumption of normality, and far below what is delivered by conventional IV estimates.

2 Instrumental Variable Estimation and Censoring

We denote by X_i the observed years of education of father i and Y_i the years of education of his child. Z_i^* defines the instrument used in estimation, in this case an indicator of the social class of the child's paternal grandfather. X_i^* and Y_i^* denote the latent variables underlying the observed data. These latent variables are all measured relative to their means.

If Y_c is the compulsory school-leaving age, then

$$\begin{aligned} Y_i &= Y_i^* \text{ if } Y_i^* \geq Y_c \\ Y_i &= Y_c \text{ if } Y_i^* < Y_c \end{aligned}$$

with a similar relationship holding for X_i and X_i^* . We assume that the underlying relationship we want to estimate is between the latent variables

$$Y_i^* = \gamma X_i^* + \varepsilon_i^Y; \quad \varepsilon_i^Y \text{ are iid}$$

Our interest is in the IV estimator; this tells us how far the influence of Z_i^* on X_i^* is transmitted to Y_i^* .

In the absence of censoring the IV estimate would be

$$\gamma_{IV}^* = \frac{Cov(Z^*Y^*)}{Cov(Z^*X^*)}$$

while in the presence of censoring

$$\gamma_{IV} = \frac{Cov(Z^*Y)}{Cov(Z^*X)}$$

Following Rigobon & Stoker (2009) we write

$$Y_i^* = Y_i + Y_i^o$$

where $Y_i^o = 0$ if $Y_i^* > Y_c$ and $Y_i^* - Y_c$ otherwise. Similarly

$$X_i^* = X_i + X_i^o$$

with $X_i^o = 0$ if $X_i^* > X_c$ and $X_i^* - X_c$ otherwise. Then

$$\gamma_{IV}^* = \frac{Cov(Z^*Y) + Cov(Z^*Y^o)}{Cov(Z^*X) + Cov(Z^*X^o)}$$

and

$$\begin{aligned} \gamma_{IV} &= \gamma_{IV}^* \frac{Cov(Z^*Y) Cov(Z^*X) + Cov(Z^*X^o)}{Cov(Z^*X) Cov(Z^*Y) + Cov(Z^*Y^o)} \\ &= \gamma_{IV}^* \frac{1 + \frac{Cov(Z^*X^o)}{Cov(Z^*X)}}{1 + \frac{Cov(Z^*Y^o)}{Cov(Z^*Y)}} \end{aligned}$$

Whether censoring leads to attenuation or expansion of the coefficient depends then on the relative magnitudes of $\frac{Cov(Z^*X^o)}{Cov(Z^*X)}$ and $\frac{Cov(Z^*Y^o)}{Cov(Z^*Y)}$. To explore this further we develop a simple structural model.

$$X_i^* = \delta Z_i^* + \varepsilon_i^X \quad (1)$$

$$Y_i^* = \gamma X_i^* + \varepsilon_i^Y \quad (2)$$

$$Z_i^* = \varepsilon_i^Z \quad (3)$$

$$E \begin{bmatrix} \varepsilon_i^X \\ \varepsilon_i^Y \\ \varepsilon_i^Z \end{bmatrix} = 0, \quad Cov \begin{bmatrix} \varepsilon_i^X \\ \varepsilon_i^Y \\ \varepsilon_i^Z \end{bmatrix} = \begin{bmatrix} \sigma_X^2 & \sigma_{XY} & 0 \\ \sigma_{XY} & \sigma_Y^2 & 0 \\ 0 & 0 & \sigma_Z^2 \end{bmatrix} \quad (4)$$

with the standard identifying assumption $\sigma_{YZ} = 0$ imposed. It is also assumed that δ represents the whole of the interrelationship between X_i^* and Z_i^* so that $\sigma_{XZ} = 0$

A key point to note is that impacts here are homogeneous. This means that any differences in estimates cannot be attributed to impact heterogeneity across complier groups. Rather they arise because the censoring generates biases which vary with the degree of censoring and also, as we subsequently show, with the threshold converting a latent instrumental variable into an (observed) dummy instrumental variable.

If we now consider the reduced form of the model, substituting out X_i^* we can write

$$X_i^* = \delta \varepsilon_i^Z + \varepsilon_i^X \quad (5)$$

$$Y_i^* = \gamma (\delta \varepsilon_i^Z + \varepsilon_i^X) + \varepsilon_i^Y \quad (6)$$

$$Z_i^* = \varepsilon_i^Z \quad (7)$$

so that

$$\mathbf{V} = \text{Cov} \begin{bmatrix} X_i^* \\ Y_i^* \\ Z_i^* \end{bmatrix} = \begin{pmatrix} \sigma_X^2 + \delta^2 \sigma_Z^2 & \gamma (\sigma_X^2 + \delta^2 \sigma_Z^2) + \sigma_{XY} & \delta \sigma_Z^2 \\ \gamma (\sigma_X^2 + \delta^2 \sigma_Z^2) + \sigma_{XY} & \sigma_Y^2 + \gamma^2 (\sigma_X^2 + \delta^2 \sigma_Z^2) + 2\gamma \sigma_{XY} & \gamma \delta \sigma_Z^2 \\ \delta \sigma_Z^2 & \gamma \delta \sigma_Z^2 & \sigma_Z^2 \end{pmatrix} \quad (8)$$

We now establish sufficient conditions for the biases to cancel out. We normalise the variables, setting $s_x = \sqrt{\sigma_X^2 + \delta^2 \sigma_Z^2}$ and $s_y = \sqrt{\sigma_Y^2 + \gamma^2 (\sigma_X^2 + \delta^2 \sigma_Z^2) + 2\gamma \sigma_{XY}}$ so that $x_i^* = \frac{X_i^*}{s_x}$, $y_i^* = \frac{Y_i^*}{s_y}$ and $z_i^* = \frac{Z_i^*}{\sigma_Z}$. Suppose that x_i^* and y_i^* are drawn from the same probability distribution, $f(\cdot)$. Thus

$$f(x_i^*) = f(y_i^*). \quad (9)$$

Such a situation of course, arises if the vector $[\varepsilon_i^X, \varepsilon_i^Y, \varepsilon_i^Z]$ is normally distributed, since then all linear combinations of it with zero mean will also be normally distributed about zero. If they have the same censor point after correcting for scale, so that $x_c = X_c/s_x = y_c = Y_c/s_y$ then it follows immediately that $\frac{\text{Cov}(Z^* X^o)}{\text{Cov}(Z^* X)} = \frac{\text{Cov}(Z^* Y^o)}{\text{Cov}(Z^* Y)}$ so that the estimator is unbiased. In our example such a situation might arise if the same proportions of fathers and children stay at school until the minimum school-leaving age, provided of course that the underlying distribution functions are also the same. More practically, with similar cut points and similar distributions the bias is unlikely to be large. We now explore the bias arising when the variables are normally distributed noting that non-parametric methods (Chernozhukov, Fernandez-Val & Kowalski 2015) have not yet evolved to the point where they can address the effects of censoring when both a dependent and an endogenous explanatory variable are censored.

3 The Bias when Variables are Normally Distributed

We first assume that the specification is as above so the instrument is a continuous variable. In appendix A we show that, if γ_{IV} is the IV estimator calculated from the censored data and γ_{IV}^* is the IV estimator calculated from the uncensored data, then

$$\gamma_{IV} = \gamma_{IV}^* \frac{\Phi(-y_c)}{\Phi(-x_c)} \quad (10)$$

giving us a measure of the bias. Of course the term $\Phi(-y_c)/\Phi(-x_c)$ is simply the ratio of the proportions of Y and X which are uncensored observations. Hence, in the normal case IV estimates can be adjusted to correct for censoring bias.

We now turn to the case where the instrument is a dummy variable, with the underlying latent variable unobserved. This is more relevant to the question we face, because the indicator of social class is a discrete, not a continuous variable. Suppose that

$$Z_i = 0 \text{ if } Z_i^* \leq Z_c \quad (11)$$

$$Z_i = 1 \text{ if } Z_i^* > Z_c \quad (12)$$

The model then becomes

$$X_i^* = \delta Z_i + \varepsilon_i^X \quad (13)$$

$$Y_i^* = \gamma X_i^* + \varepsilon_i^Y \quad (14)$$

$$Z_i^* = \varepsilon_i^Z \quad (15)$$

$$E \begin{bmatrix} \varepsilon_i^X \\ \varepsilon_i^Y \\ \varepsilon_i^Z \end{bmatrix} = 0, \quad Cov \begin{bmatrix} \varepsilon_i^X \\ \varepsilon_i^Y \\ \varepsilon_i^Z \end{bmatrix} = \begin{bmatrix} \sigma_X^2 & \sigma_{XY} & 0 \\ \sigma_{XY} & \sigma_Y^2 & 0 \\ 0 & 0 & \sigma_Z^2 \end{bmatrix} \quad (16)$$

We show in Appendix A that, when the underlying disturbances driving the latent variables are normal, with z_c the normalised value of Z_c and $\rho_{xz} = \delta/s_x$ the correlation between the normalised values x_i^* and z_i^* in the reduced form

$$\begin{aligned} Cov(xz) &= \phi(x_c)\Phi\left(\frac{\rho_{xz}x_c - z_c}{\sqrt{1 - \rho_{xz}^2}}\right) + \rho_{xz}\phi(z_c)\Phi\left(\frac{\rho_{xz}z_c - x_c}{\sqrt{1 - \rho_{xz}^2}}\right) \\ &\quad + x_c\Phi(x_c, -z_c, -\rho_{xz}) - \Phi(-z_c)\{\Phi(x_c)x_c + \phi(x_c)\} \end{aligned} \quad (17)$$

$Cov(yz)$ is again evaluated by substitution. The IV estimator is

$$\gamma_{IV}^D = \frac{Cov(yz)}{Cov(xz)}.$$

The analysis of section 2 remains valid, but the condition for the bias to cancel out has to reflect the change of instrument and becomes $\frac{Cov(ZX^o)}{Cov(ZX)} = \frac{Cov(ZY^o)}{Cov(ZY)}$.

The results in Appendix A identify two cases where γ_{IV}^D is an unbiased estimator of γ_{IV}^* . First, and not surprisingly, if the x_i and y_i are uncensored, so that $x_c = y_c = -\infty$, then

$$Cov(xz) = \rho_{xz}\phi(z_c) \text{ and } Cov(yz) = \rho_{yz}\phi(z_c)$$

and $\gamma_{IV}^D = \gamma_{IV}^*$. Secondly, if the censor/cut points are all zero, then

$$Cov(xz) = \rho_{xz}\phi(0)/2 \text{ and } Cov(yz) = \rho_{yz}\phi(0)/2$$

so that $\gamma_{IV}^D = \gamma_{IV}^*$. Beyond this it is necessary to calculate γ_{IV}^D in order to establish how large the biases are when the censor/cut points are different from zero. A particular case of interest arises when the two censor points are the same while the cut point, Z_c , varies. Rather than explore this for a wide range of possible parameters, we now present our data set. We can then estimate the model, both by IV and on the assumption of normality. This allows us to explore the implications of the model parameters for the biases that are generated in the IV estimators,

4 Empirical results

4.1 Data

The data we use are taken from the British Cohort Study. They present father-child pairs giving the age at which each completed their full-time education. They also show the occupation of the child's paternal grandfather at the time when the child's father left school. This occupational status is used to provide an indicator of grandparental social class, with six categories being identified. Professional and managerial workers are classified to social class 1, while social class V covers elementary occupations. Social class III is split between non-manual (III NM) and manual (III M) workers with the former regarded as having higher social status than the latter.

Table 1 shows the cross tabulation of fathers' age of completing education against the grandfathers' social class. The table consolidates those fathers who completed their education at the age of twenty-three or older into a single category. This is done purely for convenience; the data we use are not top-coded. Table 2 shows the analogous data for the children; since these data were observed when the children were aged twenty-six, there is an element of top-censoring, but its impact is unlikely to be large; only 0.2% of the sample were still receiving education at the age of twenty-six. Some of the fathers completed their education before the school-leaving age was increased to fifteen in April 1947 in Great Britain but ten years later in Northern Ireland. We exclude those father-child pairs whose fathers were born in 1932 or earlier in Great Britain or who were born in 1942 or from Northern Ireland, as well as those whose fathers were born abroad. This exclusion results in 6036 observations being dropped out an initial 17196 children. On top of this there is considerable attrition, giving us a final sample of 3868 father-child pairs.

The data are unweighted. It is, however, possible to relate the probability of dropping

Age at which Father completed Education	Grandfather's Class						
	I	II	III NM	III M	IV	V	All
15	11.1%	37.5%	36.3%	66.7%	71.1%	81.6%	59.3%
16	10.3%	17.7%	26.2%	16.3%	14.6%	10.5%	16.5%
17	14.5%	11.5%	12.2%	5.3%	5.8%	2.4%	7.1%
18	21.4%	9.8%	9.1%	4.1%	4.0%	2.7%	5.9%
19	1.7%	1.7%	1.5%	1.2%	0.3%	0.7%	1.1%
20	4.3%	1.2%	1.2%	0.5%	0.0%	0.3%	0.7%
21	10.3%	5.6%	5.2%	2.5%	1.1%	0.7%	3.1%
22	5.1%	4.8%	3.7%	1.2%	1.4%	1.0%	2.2%
23+	21.4%	10.1%	4.6%	2.2%	1.7%	0.0%	4.1%
Number	117	661	328	1818	650	294	3868

Table 1: Father's Age of Completing Education and Grandfather's Social Class (column percentages)

Age at which Child completed Education	Grandfather's Class						
	I	II	III NM	III M	IV	V	Total
16	10.3%	28.5%	35.1%	49.2%	52.5%	58.5%	44.6%
17	10.3%	13.1%	12.5%	13.0%	12.9%	12.2%	12.8%
18	11.1%	16.0%	13.4%	13.5%	15.2%	16.0%	14.3%
19	9.4%	4.7%	4.0%	3.2%	3.4%	2.7%	3.7%
20	2.6%	2.7%	1.8%	1.8%	1.8%	0.0%	1.8%
21	13.7%	7.8%	7.3%	6.0%	4.2%	3.4%	6.1%
22	18.8%	12.7%	12.8%	5.0%	4.3%	2.0%	7.1%
23+	23.9%	14.4%	13.1%	8.3%	5.7%	5.1%	9.5%
Number	117	655	328	1824	650	294	3868

Table 2: Child's Age of Completing Education and Grandfather's Social Class (column percentages)

Probit	Coeff. (s.e.)
Father married	0.532*** (0.110)
Social Class I	0.661*** (0.078)
Social Class II	0.661*** (0.067)
Social Class III NM	0.601*** (0.066)
Social Class III M	0.392*** (0.059)
Social Class IV	0.227*** (0.065)
Constant	-1.280*** (0.119)
N	10,494
Log-likelihood	-6,795.2

Table 3: Determinants of the Probability of an Initial Respondent remaining in our Sample

out of the survey to characteristics reported in the original 1970 survey, at least for the vast majority of respondents. We report in table 3 a probit equation used to explain the probability of dropping out. We find that the probability of a child remaining in our sample is increasing in the social status of the father and is higher if the parents were married than if they were not. After excluding observations of fathers who could leave school at fourteen, these data are available for 10494 respondents out of the total initial sample of 17196 children. While some covariates are available for all the children, we judge that the benefits of using reasonably powerful covariates to account for non-response outweighs the costs of losing those children for whom the covariates are not available. We use the probit equation to provide weights with which we correct our sample for the effects of attrition. The subsequent calculations are carried out using weighted data.

4.2 IV Estimates

The first stage in assessing the importance of bias is to examine IV estimates. As is clear from section 4.1, we can observe six categories of social class. This gives rise to five independent dummy variables which can be used as instruments, while the simple model set out above has only one dummy variable. At the same time, because the dummy variables are ordered, it is possible to consolidate them in order to carry out

	Five Social		Grandfather's Social Class			
	Class Dummies	I	\geq II	\geq IIINM	\geq IIIM	\geq IV
γ_D^{IV}	0.844*** (0.058)	0.786*** (0.096)	0.807*** (0.067)	0.858*** (0.064)	1.000*** (0.106)	1.034*** (0.140)
Constant	4.323*** (0.924)	5.255*** (1.551)	4.918*** (1.082)	4.086*** (1.031)	1.811 (1.697)	1.261 (2.254)
N	3868	3868	3868	3868	3868	3868
Kleinbergen-Paap	310	58.9	190	256.6	159.8	107.2
Sargan	$\chi_4^2=4.35$					
Percentage Dummy=1		2.6%	18%	26%	74.3%	91.7%

Table 4: IV Coefficient Estimates as Functions of the Cut Point for the Dummy Instrumental Variable

five possible IV regressions, in each of which the instrument is a single dichotomous dummy. This allows us to explore the effect of moving the cut point for the dummy, Z_c , on the resulting estimate of γ_{IV}^D . Equation (17) suggests that that should influence the regression coefficient. A further benefit of the presence of five independent dummies is that it is possible to carry out Sargan's (1958) test for over-identification and thus provide a degree of reassurance that the restriction $\sigma_{YZ} = 0$ is acceptable and hence that the statistical analysis is valid.

In table 4 the results of these IV regressions are shown. The first column shows the estimates when all five social class dummies are used as instruments. The subsequent five columns show the estimates produced by dummies indicating social class of at least the value indicated.¹ The table also shows the proportion of respondents in each category, and the cut point calculated on the assumption that the latent variable underlying social class is normally distributed.

The results with five dummies suggest that the Sargan test is easily met, while the Kleinbergen-Paap statistic does not point to any concerns that the instruments are weak; in statistical terms the instruments seem valid. The IV estimates also show a clear tendency for the coefficient to rise with the cut point. The question we now wish to address is whether this is a natural feature of the interaction between the cut point of the instrument and the censored nature of the data on age of completing education. In other words, does this relationship between the IV coefficient and the definition of the instrument reflect the bias arising from censoring?

¹Following convention, we refer to social class I being higher than social class II. It indicates higher status even if a lower class number.

4.3 Estimates Corrected for Censoring

The first step in examining this is to estimate the counterparts to the models of table 4 using the structure of equations (1)-(4), but in a way which corrects for the effects of censoring. Once again, it is possible to do it for five different definitions of the instrument. The five single instrument models can be estimated using the *cmp* command in STATA despite the fact that only the dummies are observed, given the assumption that the underlying variables are normally distributed. We can also set up a model in which all five categories of social class are used to delineate the latent variable assumed to underlie social class. The five single instrument models provide a valuable comparison with table 4 while the model which exploits the information on all categories of social class offers the most obvious set of parameters with which to explore how closely the empirical findings of table 4 match the theoretical implications conditional on normality.

The empirical analogue to the model set out by equations 1-4 is specified as follows:-

$$X_i^* = \mu_X + \delta Z_i^* + \varepsilon_i^X \quad (18)$$

$$Y_i^* = \mu_y + \gamma X_i^* + \varepsilon_i^Y \quad (19)$$

$$Z_i^* = \varepsilon_i^Z \quad (20)$$

where the observed values, X_i and Y_i are defined as in section 2.

The continuous variable underlying social class is not observed, but we define a sequence of cut points

$$Z_i^* \leq Z_1^C \text{ if } Z_i = 1, \quad Z_n^C < Z_i^* \leq Z_{n+1}^C \text{ if } Z_i = n + 1 \text{ and } Z_i^* > Z_5^C \text{ if } Z_i = 6$$

By analogy with the earlier models, we can estimate the system using an ordered probit model for equation (20) or we can specify it with a dichotomous variable defined with reference to a single cut point. The parameters are identified by setting the variance of ε_i^Z is set to 1 and the covariances σ_{XZ} and σ_{YZ} to zero. The results of this are shown in table 5. It can be seen that the parameter γ is much more stable across the different specifications than in table 4; it is falling slightly, rather than rising in the cut point.

It should be noted that a closely related specification is provided by replacing equation (19) by

$$Y_i^* = \mu_y + \gamma X_i + \varepsilon_i^Y \quad (21)$$

Here it is the actual age at which the father completes his education, rather than his latent age of completion, which influences the age of completion of the child. The two

	Grandfather's Class					
	I	II	III NM	III M	IV	V
Child's Age of Completion						
Constant	8.344*** (0.583)	6.995*** (0.950)	7.940*** (0.674)	8.307*** (0.604)	8.515*** (0.954)	9.021*** (1.233)
γ	0.604*** (0.041)	0.706*** (0.069)	0.635*** (0.048)	0.608*** (0.043)	0.592*** (0.070)	0.554*** (0.091)
Father's Age of Completion						
Constant	13.340*** (0.110)	13.310*** (0.112)	13.318*** (0.112)	13.301*** (0.112)	13.324*** (0.111)	13.321*** (0.111)
δ	-1.835*** (0.093)	-2.289*** (0.156)	-2.052*** (0.117)	-2.195*** (0.110)	-1.437*** (0.121)	-1.528*** (0.168)
Cut Points						
Cut 1	-1.964*** (0.040)	-1.945*** (0.040)				
Cut 2	-0.916*** (0.023)		-0.914*** (0.023)			
Cut 3	-0.640*** (0.022)			-0.644*** (0.021)		
Cut 4	0.656*** (0.022)				0.654*** (0.023)	
Cut 5	1.374*** (0.030)					1.385*** (0.031)
Variance-covariance						
$\log \sigma_X$	1.447*** (0.024)	1.403*** (0.029)	1.431*** (0.025)	1.417*** (0.025)	1.487*** (0.024)	1.481*** (0.026)
$\log \sigma_Y$	1.327*** (0.020)	1.357*** (0.031)	1.334*** (0.022)	1.328*** (0.021)	1.321*** (0.024)	1.313*** (0.024)
$\tanh^{-1}\sigma_{XY}/(\sigma_X\sigma_Y)$	-0.228*** (0.049)	-0.383*** (0.093)	-0.272*** (0.060)	-0.244*** (0.054)	-0.200* (0.088)	-0.151 (0.119)
N	3868	3868	3868	3868	3868	3868
Log-Lik.	-14934	-10758.6	-11820.6	-12106.9	-12216.8	-11322.4
Log-Lik. (eq 21)	-14996	-10788	-11873	-12164.6	-12245.2	-11338.5

Table 5: Parameter Estimates allowing for Censoring when Child's Age of Completion is influenced by Father's Latent Age of Completion

models have the same number of parameters, so it is reasonable to discriminate between them on the basis of the log likelihoods associated with them. The log-likelihoods of this second group of models are shown in the final row of table 5. These log-likelihoods suggest strongly that the latent variable model of equation (19) should be preferred to the actual variable model of equation (21).

The estimation of equations (18) to (20) provides us with the parameters of the system described by equations (1) to (4) of section 3; it is natural to choose the parameters found with multiple cut points for the social class variable, since these are the estimators which make most use of the available information. With the parameters of the first column of table 5 and the underlying assumption of joint normality, we can calculate the values of the IV estimator which the theoretical analysis of section 3 suggests should be found with a dichotomous dummy instrument.

The model parameters imply the following values for the elements of the covariance matrix of the uncensored data. \mathbf{V} , defined by equation (8), and its normalised equivalent, Σ

$$\mathbf{V} = \begin{bmatrix} 21.44 & 9.24 & -1.84 \\ 9.24 & 17.55 & -1.11 \\ -1.84 & -1.11 & 1 \end{bmatrix}; \quad \Sigma = \begin{bmatrix} 1 & 0.48 & -0.40 \\ 0.48 & 1 & -0.26 \\ -0.40 & -0.26 & 1 \end{bmatrix}$$

Using standard notation to refer to the elements of \mathbf{V} and Σ ,

$$\gamma_{IV}^* = \frac{\mathbf{V}_{2,3}}{\mathbf{V}_{1,3}} = \frac{\Sigma_{2,3}}{\Sigma_{1,3}} \sqrt{\frac{\mathbf{V}_{2,2}}{\mathbf{V}_{1,1}}} = 0.60$$

In order to explore the biases arising from censoring we work from matrix Σ , so as to exploit the analysis of section 3. We then multiply the results by $\sqrt{\mathbf{V}_{2,2}/\mathbf{V}_{1,1}}$ in order to express them in terms of a relationship between ages of completion of education of fathers and children.

With the scaled and weighted data $x_c = 0.36$ and $y_c = -0.10$ corresponding to proportions of fathers and children completing their education at the statutory minimum age of 64% and 47% respectively. Equation (10) implies that, if the latent instrument were observed, it would deliver an estimate of the parameter, $\gamma_{IV}^l = 0.89$ in contrast to the true parameter of 0.60. In table 6 we show the cut points for the latent variable underlying the five dichotomous instruments of table 4 together with the estimates of the IV parameter, γ_{IV}^D , which would be generated by using these dummies, with no correction for the effects of censoring. These are compared with the estimates from table 4. Finally we show in the table the estimates of the parameters which would be

	Grandfather's Social Class				
	I	\geq II	\geq IIINM	\geq IIIM	\geq IV
Instrument Cut Point (table 5)	-1.964	-0.916	-0.640	0.657	1.374
Theoretical γ_{IV}^D	0.751	0.826	0.850	0.988	1.078
Estimate (table 4)	0.790	0.810	0.860	1.000	1.030
γ_{IV}^D if $x_c = y_c = 0$	0.558	0.577	0.584	0.629	0.662

Table 6: Parameter Estimates generated by a Censored Normal Distribution

generated if $x_c = y_c = 0$, i.e. if half of the fathers and children had completed their education at the statutory minimum age. This sheds further light on the effects of censoring and its interaction with the cut point of the instrument.

This table shows the connection between the choice of instrument (i.e. with Z_c) and the estimated parameter value. The theoretical model shows this ranging from 0.75 to 1.08 and the empirical estimates match the theoretical values closely. The theoretical results found when the two censor points are set to zero suggests that the bias arises primarily from the difference in the proportions of fathers and children completing their education at the minimum age, rather than the interaction of this with the instrument. Further simulations with other values of the censor point confirm this, at least given the assumption of normality.

The close match between the estimates of table 4 and the theoretical results might be taken to suggest that, in this particular case, the assumption of normality is not too far from the mark. It is, however, possible to investigate this further, and we do that in the next section, using an ordered probit model.

4.4 Allowing for Non-normality using an Ordered Probit Model

The analysis so far has made two important assumptions. First, it has been assumed that the relationship between father's education and child's education is linear. The British education system identifies a number of important thresholds, with school exams typically taken at ages sixteen and eighteen, followed, for those who proceeded to university, by graduation from university three years later. If attaining these thresholds is valuable and the benefits can be inherited, then one would expect the relationship between father's and child's age of completing education to be non-linear. Secondly, it was assumed that the latent ages of completion were, along with the latent variable representing grandparental social class, jointly normally distributed. While, as noted earlier, some progress has been made with non-parametric methods requiring much weaker assumptions, these techniques do not make it possible to estimate the model we have here,

in which a censored variable is explained by another endogenous censored variable.

An alternative means of estimating the latent variable model, with weaker distributional assumptions, is to treat the ages of completion of education of the fathers and children as the cut points in a multivariate ordered probit model. With this model, the cut points on the latent education variables are free to vary and can represent an arbitrary distribution. This avoids the assumption of normality of outcomes. Normality of latent variables is of course required but the flexibility of the cut points means that that is not restrictive. Grandfather's social class is the third equation in the model; also of ordered probit form. With this instrument, the model captures the causal relationship between father's and child's latent education.

The model also addresses two other, more minor, features of our data. First, the data themselves are interval censored, reporting ages of completion in complete years². Secondly, as we noted earlier, the data for child's age of completion are top-censored at twenty-six because they were collected when the children were aged twenty-six. As we argued earlier, there is no reason to believe that this top-censoring has, despite de Haan (2011), a material effect on the coefficients; it nevertheless does no harm to remove it.

The model in terms of latent variables is that of equations (1)- (3) but the latent variables themselves have changed.

$$X_i^* = \delta Z^* + \varepsilon_i^X \quad (22)$$

$$Y_i^* = \zeta X_i^* + \varepsilon_i^Y$$

$$Z_i^* = \varepsilon_i^Z \quad (23)$$

with

$$\begin{bmatrix} \varepsilon_i^X \\ \varepsilon_i^Y \\ \varepsilon_i^Z \end{bmatrix} \sim N \left(0, \begin{bmatrix} 1 & \sigma_{XY} & 0 \\ \sigma_{XY} & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \right) \quad (24)$$

so that the latent variables all have zero mean. The parameter relating father's and child's latent variables is referred to as ζ to distinguish it from the parameter γ which related ages of completing education. As before, we impose the identifying restrictions, $\sigma_{XZ} = 0$ and $\sigma_{YZ} = 0$.

²This is in fact a more complicated issue than simple interval censoring. People tend to complete their education at fixed points such as the end of an academic year. Their precise age on completion thus depends partly on the date in the year at which they were born.

We define cut points X_1^C to X_{16}^C , Y_1^C to Y_{10}^C and Z_1^C to Z_5^C

$$\begin{aligned}
X_i^* &\leq X_1^C \text{ if } X_i = 15, X_n^C < X_i^* \leq X_{n+1}^C \text{ if } X_i = 15 + n \text{ with } 1 \leq n \leq 13 \\
X_{14}^C &< X_i^* \leq X_{15}^C \text{ if } X_i = 29; X_{15}^C < X_i^* \leq X_{16}^C \text{ if } X_i = 32; X_i^* > X_{16}^C \text{ if } X_i \geq 33 \\
Y_i^* &\leq Y_1^C \text{ if } Y_i = 16, Y_n^C < Y_i^* \leq Y_{n+1}^C \text{ if } Y_i = 16 + n \text{ and } Y_i^* > Y_{10}^C \text{ if } Y_i \geq 26 \\
Z_i^* &\leq Z_1^C \text{ if } Z_i = 1, Z_n^C < Z_i^* \leq Z_{n+1}^C \text{ if } Z_i = n + 1 \text{ and } Z_i^* > Z_5^C \text{ if } Z_i = 6
\end{aligned}$$

It should be noted that there are no observations with $X_i = 30$ or 31 . The parameters of the model can then be estimated using the multivariate ordered probit procedure in *cmp*. The results are shown in table 7.

There are a number of issues raised by the table. First of all, the log-likelihood of -14170 compares with that of -14934 for the censored linear model of table 5. There are twenty-three more parameters in the ordered probit model, but even allowing for this, the log-likelihood suggests that the ordered probit model should be strongly preferred to the censored linear model.³ A counterpart of this is that the cut points shown in table 7 are very unevenly placed.

This in turn raises issues over the interpretation of the coefficient ζ . That shows the extent to which the latent variable determining father's age of completing education influences the latent variable determining the age at which the child leaves education. Unlike the situation with the earlier models, the latent variables do not directly represent ages of completing education. With the ordered probit model, the expected marginal increase in the child's age of completion associated with a marginal increase in the father's age of completion depends on the latter. Furthermore we can evaluate this only for ages beyond the father's compulsory schooling because the specification does not allow us to draw any implications about the relationship between latent ages of completion below the limit set by the statutory minimum school leaving age.

For each observation we can, however, work out the marginal relationships between the latent variables and use these to translate ζ into a relationship between ages of completion of the father and the child. The non-linearity means that that will be specific to each individual. Averaging across the population, however, provides an estimate of the average marginal impact of father's education on that of his child.

We denote by T_i^x the expected age of completion of the father conditional on the latent variable for social class of z_i^* , and T_i^y the expected age of completion of the child

³The AIC and BIC for the ordered probit model are 31,318.6 and 31,540.882 respectively. These are both lower than for the model of table 5 (AIC and BIC of 33,212.2 and 33,288.411 respectively).

Child's age of completion	Father's age of completion	Grandfather's social class			
ζ	0.654*** (0.040)	δ -0.438*** (0.022)			
Cut Points					
	16	0.368*** (0.023)	Class I	-1.964*** (0.040)	
17	-0.082*** (0.023)	17	0.877*** (0.024)	Class II	-0.916*** (0.023)
18	0.279*** (0.023)	18	1.154*** (0.027)	Class III NM	-0.640*** (0.022)
19	0.727*** (0.026)	19	1.449*** (0.030)	Class III M	0.657*** (0.022)
20	0.859*** (0.027)	20	1.519*** (0.031)	Class IV	1.374*** (0.030)
21	0.932*** (0.028)	21	1.563*** (0.032)		
22	1.177*** (0.031)	22	1.798*** (0.035)		
23	1.535*** (0.038)	23	2.021*** (0.040)		
24	1.901*** (0.046)	24	2.250*** (0.047)		
25	2.299*** (0.059)	25	2.457*** (0.055)		
26	3.085*** (0.126)	26	2.774*** (0.075)		
		27	3.030*** (0.099)		
		28	3.123*** (0.112)		
		29	3.194*** (0.122)		
		32	3.523*** (0.189)		
		33	3.691*** (0.244)		
$\tanh^{-1}\sigma_{XY}$	-0.212*** (0.049)				
N	3,868				
Log-likelihood	-14,169.9				

Table 7: The Parameters of the Ordered Probit Model

conditional on z_i^* . Write $\lambda_i^x = dT_i^x/dx_i^*$ and $\lambda_i^y = dT_i^y/dy_i^*$. Since $dy_i^*/dx_i^* = \zeta$ we can then write

$$\gamma_i = \frac{dT_i^y}{dT_i^x} = \zeta \frac{\lambda_i^y}{\lambda_i^x}$$

We show in appendix B that, with τ_k^x being the age of completion of education associated with individuals whose latent variables lie between cut point $k - 1$ and cut point k . Then conditional on a given value of the social class latent variable, z_i^*

$$\lambda_i^x(z_i^*) = \frac{dT_i^x}{dx_i^*} = -\frac{\sum_{k=2}^{N-1} (\phi(X_k - \delta z_i^*) - \phi(X_{k-1} - \delta z_i^*)) \tau_k^x - \phi(X_1 - \delta z_i^*) \tau_2^x - \phi(X_{N-1} - \delta z_i^*) \tau_{N-1}^x}{\{\Phi(X_{N-1} - \delta z_i^*) - \Phi(X_1 - \delta z_i^*)\}}$$

In applying this formula we set the upper cut point to that for age 29 (so that $\tau_2^x = 16$ and $\tau_{N-1}^x = 28$) because the next cut point is at age 32. This has negligible effect because the proportion of fathers reporting completing their education after age 29 is minimal.

For children this complication is not present; with $\tau_2^y = 17$ and $\tau_{N-1}^y = 25$ we have

$$\lambda_i^y(z_i^*) = \frac{dT_i^y}{dy_i^*} = -\frac{\sum_{k=2}^{N-1} (\phi(\frac{Y_k - \delta \zeta z_i^*}{\sigma_y}) - \phi(\frac{Y_{k-1} - \delta \zeta z_i^*}{\sigma_y})) \tau_k^y - \phi(\frac{Y_1 - \delta \zeta z_i^*}{\sigma_y}) \tau_2^y - \phi(\frac{Y_{N-1} - \delta \zeta z_i^*}{\sigma_y}) \tau_{N-1}^y}{\left\{ \Phi\left(\frac{Y_{N-1} - \delta \zeta z_i^*}{\sigma_y}\right) - \Phi\left(\frac{Y_1 - \delta \zeta z_i^*}{\sigma_y}\right) \right\}}$$

Both λ_i^x and λ_i^y and thus ζ_i are functions of z_i^* which is of course unobserved. We may, however, calculate their expected values conditional on social class n_i^z being observed.

We evaluate

$$\gamma_{n_i^z} = \zeta \frac{\int_{Z_{n_i^z-1}^{z_i^*}}^{Z_{n_i^z}^{z_i^*}} \{\lambda^y(z_i^*) / \lambda^x(z_i^*)\} \phi(z_i^*) dz_i^*}{\Phi(Z_{n_i^z}) - \Phi(Z_{n_i^z-1})}$$

as the expected marginal impact conditional on a grandfather from social class n_i^z . The average marginal effect is then given as

$$\gamma_{OP} = \sum_i w_i \gamma_{n_i^z} / \sum_i w_i \quad (25)$$

where n_i^z is the social class of observation i .

We can evaluate γ_{OP} either for the whole sample or, perhaps more appropriately, only for the restricted sample of 1,166 observations for which both the father and the child have completed their education when older than the minimum school-leaving age. We show in table 8 estimates of γ_{OP} for these two populations and also for father/child pairs as a function of the social class of the grandfather.

The nonlinearities imply that the marginal transmission of educational advantage is greater for those with grandfathers from the high social classes than from the low social

	Whole	Restricted	Grandfather's Class					
	Sample	Sample	I	II	III NM	III M	IV	V
γ_{OP}	0.54	0.56	0.68	0.59	0.55	0.53	0.49	0.50

Table 8: Estimates of the Average Marginal Impact of an Extension of Father's Education on that of Children

classes. The average marginal value for the restricted sample of 0.56 can be compared with the value of 0.60 found using the censored linear model (table 5) and 0.84 estimated by IV (table 4).

5 Conclusions

We have shown here, in a very practical example, the sort of distortions which can arise when parameter estimates are produced by instrumental variables using data that are censored. In our application – an investigation of the relationship between fathers' and children's ages of completing their education – the fact that more than half of the fathers and nearly half of the children left school at the compulsory school leaving age generates a substantial bias upward bias in the estimate of the magnitude of the relationship. Making the assumptions that the underlying variables are normally distributed, and that the structural relationship is between the unobserved latent ages of completion of education, we are able quantify the biases.

We find strong evidence to support the belief that the relationship is indeed between the latent variables, rather than influenced by the actual experience of the fathers. The instrument available to us, grandparental social class, is hexachotomous, allowing us to identify five different dichotomous dummy variables. We find a close match between the IV parameter estimates using these dummy variables and the values predicted by our theoretical analysis on the assumption of normality. All of these values show an upward bias compared to the underlying parameter estimate. The estimate produced using all five dummy variables as instruments suggests that a child's age of completing education rises by 0.84 years for each extra year that their father underwent full-time education, while methods which correct for the effects of censoring point to a coefficient of only 0.60. Use of a multivariate ordered probit model allows us to relax the assumption of normally distributed education and points to an average marginal impact of father's age of completion on that of his child of only 0.56 years. This suggests that the bias arising from the use of IV estimates with censored data is much greater than the bias arising from the assumption of normality.

Our results cast doubt on the idea that the intergenerational transmission of educational advantage is as powerful as is suggested by conventional linear models. Methodologically, they highlight the need to pay adequate regard to the issue of censoring. Furthermore, they caution against attributing variation in IV results to impact heterogeneity across instrument-specific complier populations. In the common case of dummy instruments, such variation can, when data are censored, equally be due to the choice of threshold according to which latent instrumental variables are dichotomised; this can itself be seen as an extreme form of censoring.

References

- Austin, P. & Hoch, J. (2004), ‘Estimating Linear Regression Models in the Presence of a Censored Independent Variable’, *Statistics in Medicine* **23**, 411–429.
- Chernozhukov, V., Fernandez-Val, I. & Kowalski, A. (2015), ‘Quantile Regression with Censoring and Endogeneity’, *Journal of Econometrics* **186**, 201–221.
- de Haan, M. (2011), ‘The Effect of Parents’ Schooling on Child’s Schooling: A Nonparametric Bounds Analysis’, *Journal of Labor Economics* **29**, 859–892.
- de Haan, M. & Plug, E. (2011), ‘Estimating Intergenerational Schooling Mobility on Censored Samples: Consequences and Remedies’, *Journal of Applied Econometrics* **26**, 151–166.
- Dearden, L., Machin, S. & Reed, H. (1997), ‘Intergenerational Mobility in Britain’, *Economic Journal* **107**, 47–66.
- Haveman, R. & Wolfe, B. (1995), ‘The Determinants of Children’s Attainments: a Review of Methods and Findings’, *Journal of Economic Literature* **33**, 1829–1878.
- Holmlund, H., Lindahl, M. & Plug, E. (2011), ‘The Causal Effects of Parents’ Schooling on Children’s Schooling: A Comparison of Estimation Methods’, *Journal of Economic Literature* **49**, 615–651.
- Imbens, J. & Angrist, J. (1994), ‘Identification and estimation of local average treatment effects’, *Econometrica* **62**, 467–475.
- Muthen, B. (1990), ‘Moments of the Censored and Truncated Normal Distribution’, *British Journal of Mathematical and Statistical Psychology* **43**.

Oreopoulos, P., Page, M. & Stevens, A. (2006), ‘The Intergenerational Effects of Compulsory Schooling’, *Journal of Labor Economics* **24**, 729–760.

Plug, E. (2004), ‘Estimating the Effect of Mother’s Schooling on Children’s Schooling using a Sample of Adoptees’, *American Economic Review* **94**, 358–368.

Rigobon, R. & Stoker, T. (2009), ‘Bias from Censored Regressors’, *Journal of Business and Economic Statistics* pp. 340–353.

Rosenbaum, S. (1961), ‘Moments of a Truncated Bivariate Normal Distribution’, *Journal of the Royal Statistical Society, Series B* **23**, 223–229.

Twou, M.-W., Liu, J.-T. & Hammitt, J. (2012), ‘The Intergenerational Transmission of Education: Evidence from Taiwanese Adoptions’, *Economics Letters* **115**, 134–136.

A A Statistical Analysis of Censoring with Bivariate Normality

The model we set out here a reduced form of three jointly normally distributed variables. Two of the variables, X_i^* and Y_i^* are assumed to be censored, so that the observed values X_i and Y_i are defined as

$$X_i = X_i^* \text{ if } X_i^* > X_C \text{ while } X_i = X_C \text{ if } X_i^* < X_C \text{ and}$$

$$Y_i = Y_i^* \text{ if } Y_i^* > Y_C \text{ while } Y_i = Y_C \text{ if } Y_i^* < Y_C$$

The identifying conditions of section 2 are assumed to be met.

$$\begin{bmatrix} X_i^* \\ Y_i^* \\ Z_i^* \end{bmatrix} \sim N \left(\begin{bmatrix} \mu_X \\ \mu_Y \\ \mu_Z \end{bmatrix}, \begin{bmatrix} \sigma_X^2 & \rho_{xy}\sigma_X\sigma_Y & \rho_{xz}\sigma_X\sigma_Z \\ \rho_{xy}\sigma_X\sigma_Y & \sigma_Y^2 & \rho_{yz}\sigma_Y\sigma_Z \\ \rho_{xz}\sigma_X\sigma_Z & \rho_{yz}\sigma_Y\sigma_Z & \sigma_Z^2 \end{bmatrix} \right).$$

We examine two cases. In the first Z_i^* is observed, while in the second case Z_i^* is not observed. Instead we observe a dummy variable, Z_i with $Z_i = 0$ if $Z_i^* < Z_c + \mu_Z$ and $Z_i = 1$ if $Z_i^* > Z_c + \mu_Z$. Since the instrumental variable estimator of the regression coefficient is the ratio of two covariances, we evaluate the effect of censoring on the estimate of the correlation, r_{xz} calculated from observations on normalised censored data. The first step is to normalise the variables. We set

$$x_i^* = \frac{X_i^* - \mu_X}{\sigma_X}; \quad x_i = \frac{X_i - \mu_X}{\sigma_X} \text{ and } x_c = \frac{X_c - \mu_X}{\sigma_X}.$$

with similar definitions of y_i^* , y_i , y_c , z_i^* , z_i and z_c ,

We use $\phi()$ and $\Phi()$ to represent the density function and cumulative distribution of the standard normal distribution respectively. One argument indicates that the function relates to the univariate normal distribution, while three arguments (the two ordinates and the correlation) are used to indicate the bivariate normal distribution. The subsequent analysis draws heavily on the results quoted by Rosenbaum (1961) and Muthen (1990) for the moments of truncated and censored bivariate normal distributions.

A.1 The Biasing Effect of Compulsion when the Instrument is Fully Observed

We consider separately the cases depending on whether the $x_i^* > x_c$ or not.

There are two cases where parents are educated beyond the compulsory age. In the first their children are also educated beyond the compulsory age, while in the second, they are educated only up to the compulsory age. The probabilities of these are given as

1. $x_i > x_c$ with $P(x_i > x_c) = \Phi(-x_c)$
2. $x_i = x_c$ with $P(x_i > x_c) = \Phi(x_c)$

The product moment needs to be evaluated in two components, one for each of the two cases above

1. $x_i > x_c$ (Rosenbaum 1961)⁴

$$m_{xz}^1 = (\rho_{xz}\Phi(-x_c) + \rho_{xz}x_c\phi(x_c))/\Phi(-x_c)$$

2. $x_i = x_c$

$$m_{xz}^2 = -x_c\rho_{xz}\phi(x_c)/\Phi(x_c)$$

Since the first moment of $z_i^* = 0$, $r_{xz} = \text{Cov}(xz^*)$ estimated from the censored data is

$$r_{xz} = \Phi(-x_c)m_{xz}^1 + \Phi(x_c)m_{xz}^2 = \rho_{xz}\Phi(-x_c)$$

Similarly,

$$r_{yz} = \rho_{yz}\Phi(-y_c)$$

⁴Rosenbaum (1961) uses the function $Q(x)$ to refer to the probability mass of the normal distribution in the range $[x, \infty]$ rather than the range $[-\infty, x]$.

and the IV estimator from the censored data is therefore

$$\gamma_{IV} = \frac{\rho_{yz}\Phi(-y_c)\sigma_Y}{\rho_{xz}\Phi(-x_c)\sigma_X}$$

in contrast to the estimator from the uncensored data

$$\gamma_{IV}^* = \frac{\rho_{yz}\sigma_Y}{\rho_{xz}\sigma_X}$$

so that

$$\gamma_{IV} = \gamma_{IV}^* \frac{\Phi(-y_c)}{\Phi(-x_c)}$$

A.2 The Biasing Effect with a Dummy Instrument

When we observe z_i rather than z_i^* the covariance is the expected value of x_i conditional on $z_i = 1$. The expected value of the second moment around zero is given as Muthen (1990)

$$\phi(x_c)\Phi\left(\frac{\rho_{xz}x_c - z_c}{\sqrt{1 - \rho_{xz}^2}}\right) + \rho_{xz}\phi(z_c)\Phi\left(\frac{\rho_{xz}z_c - x_c}{\sqrt{1 - \rho_{xz}^2}}\right) + x_c\Phi(x_c, -z_c, -\rho_{xz})$$

and the product of the two means is given as

$$\Phi(-z_c) \{ \Phi(x_c)x_c + \phi(x_c) \}$$

so the estimate of the covariance is

$$\begin{aligned} \hat{\sigma}_{xz} &= \phi(x_c)\Phi\left(\frac{\rho_{xz}x_c - z_c}{\sqrt{1 - \rho_{xz}^2}}\right) + \rho_{xz}\phi(z_c)\Phi\left(\frac{\rho_{xz}z_c - x_c}{\sqrt{1 - \rho_{xz}^2}}\right) \\ &\quad + x_c\Phi(x_c, -z_c, -\rho_{xz}) - \Phi(-z_c) \{ \Phi(x_c)x_c + \phi(x_c) \} \end{aligned}$$

Similarly

$$\begin{aligned} \hat{\sigma}_{yz} &= \phi(y_c)\Phi\left(\frac{\rho_{yz}y_c - z_c}{\sqrt{1 - \rho_{yz}^2}}\right) + \rho_{yz}\phi(z_c)\Phi\left(\frac{\rho_{yz}z_c - y_c}{\sqrt{1 - \rho_{yz}^2}}\right) \\ &\quad + y_c\Phi(y_c, -z_c, -\rho_{yz}) - \Phi(-z_c) \{ \Phi(y_c)y_c + \phi(y_c) \} \end{aligned}$$

so the parameter estimated from the censored data using a dummy variable as instrument is

$$\gamma_{IV}^D = \frac{\hat{\sigma}_{yz}}{\hat{\sigma}_{xz}} \frac{\sigma_Y}{\sigma_X}$$

showing a clear bias, if one which is less straightforwardly represented than with the linear instrument.

It should be noted that, in the absence of censoring ($x_c = -\infty$), then

$$\hat{\sigma}_{xz} = \rho_{xz}\phi(z_c)$$

while if $x_c = z_c = 0$

$$\hat{\sigma}_{xz} = \frac{(1 + \rho_{xz})\phi(0) - \phi(0)}{2} = \rho_{xz} \frac{\phi(0)}{2}$$

It follows that if $x_c = y_c = z_c = 0$ then γ_{IV}^D is unbiased.

B Interpretation of the Ordered Probit Model

A general model explores the relationship between the years of education of fathers and children using the parametric structure of an ordered probit model. Since the ordered values for years of education are simply cut points, no assumption is made about the distribution of years of education. We assume that educational attainment is represented by latent variables, Y_i^* and X_i^* for the respondent and the respondent's father respectively. These latent variables are explained by the following system of equations

$$\begin{aligned} X_i^* &= \delta Z_i^* + \varepsilon_i^x \\ Y_i^* &= \zeta X_i^* + \varepsilon_i^y \\ Z_i^* &= \varepsilon_i^z \end{aligned}$$

$$\begin{bmatrix} \varepsilon_i^x \\ \varepsilon_i^y \\ \varepsilon_i^z \end{bmatrix} \sim N(0, \Sigma) \text{ with } \Sigma = \begin{bmatrix} 1 & \rho_{xy} & \rho_{xz} \\ \rho_{xy} & 1 & 0 \\ \rho_{xz} & 0 & 1 \end{bmatrix}$$

Actual age of completion of education is observed as N ordinal variables, We denote a sequence of age thresholds, $X_1..X_{N-1}$ with $X_0 = -\infty$ and $X_N = \infty$ as the thresholds for the father, with the corresponding thresholds for the child being $Y_0...Y_N$. $Z_0...Z_N$ are the thresholds which locate values of Z_i^* to observed social classes. These are estimated together with the parameters of the equations above, again using the STATA routine *cmp*.

Here considerable care is needed over the interpretation of ζ . It shows the marginal impact of the father's latent variable on that of the child; since neither latent variable represents age of completion of education it is not directly interpretable in terms of the influence of the father's age of completion on that of the child. If the thresholds are evenly spaced there is a simple linear relationship between the latent variable and the age of completing education. That is, however, unlikely to be the case; the point of

estimating an ordered probit model is to allow for the possibility of non-linearity. In turn that implies that the relationship between father's age of completion and child's age of completion will be non-linear. For each observation we can, however work out the marginal relationships between the latent variable and the age of completion of education. These can then be used to translate ζ into a relationship between ages of completion of the father and the child. The non-linearity means that that will be specific to each individual. Averaging across the population, however, provides an estimate of the average marginal impact of father's age of completion on child's age of completion.

We denote by T_i^x the expected age of completion of the father conditional on the latent variable for social class of Z_i^* , and T_i^y the expected age of completion of the child conditional on z_i^* . With $\lambda_i^x = dT_i^x/dx_i^*$ and $\lambda_i^y = dT_i^y/dy_i^*$. Since $dy_i^*/dx_i^* = \zeta$ we can then write

$$\gamma_i = \frac{dT_i^y}{dT_i^x} = \zeta \frac{\lambda_i^y}{\lambda_i^x}$$

With w_i the weight attached to observation i

$$\gamma_{OP} = \frac{\sum_i w_i \gamma_i}{\sum w_i}$$

provides our estimate of the average marginal relationship between age of completing education of the father and that of the child.

We proceed using $\Phi()$ to represent the cumulative normal distribution and ϕ to represent the density function of the normal distribution. Given z_i^* and conditional on the age at which father i completed his education being within the range $X_1..X_{N-1}$ his expected age of completion is, with τ_k^x the age of completion associated with threshold X_k

$$T_i^x = \frac{\sum_{k=2}^{N-1} (\Phi(X_k - \delta Z_i^*) - \Phi(X_{k-1} - \delta Z_i^*)) \tau_k^x}{\Phi(X_{N-1} - \delta Z_i^*) - \Phi(X_1 - \delta Z_i^*)}$$

We are interested in the effect that a small increase, h in δZ_i^* has on T_i^x . We can, however, only evaluate this for the population for which both $X_1 < \delta Z_i^* < X_{N-1}$ and $X_1 < \delta Z_i^* + h < X_{N-1}$ since it is only for this population that we can evaluate the expected age of completion both before and after a disturbance, h . This means that the derivative of T_i^x will not provide what we need; we have to evaluate two terms, T_i^{x*} for the expected age of completion of education for someone with a latent variable of δZ_i^* , and T_i^{x**} for someone with a latent variable $\delta Z_i^* + h$.

First,

$$T_i^{x*} = \frac{\sum_{k=2}^{N-1} (\Phi(X_k - \delta Z_i^*) - \Phi(X_{k-1} - \delta Z_i^*)) \tau_k^x + \{\Phi(X_{N-1} - \delta Z_i^* - h) - \Phi(X_{N-1} - \delta Z_i^*)\} \tau_{N-1}^x}{\{\Phi(X_{N-1} - \delta Z_i^*) - \Phi(X_1 - \delta Z_i^*)\} + \{\Phi(X_{N-1} - \delta Z_i^* - h) - \Phi(X_{N-1} - \delta Z_i^*)\}}$$

Here the second term in the numerator is an adjustment to recognise that the upper limit of integration has to be $X_{N-1} - h$ so that after the increment of h the latent variable remains within the permitted range; a similar adjustment to the denominator is needed.

For T_i^{x**} the ranges are shifted by h . The upper limit is, however, X_{N-1} .

$$T_i^{x**} = \frac{\sum_{k=2}^{N-1} (\Phi(X_k - \delta Z_i^* - h) - \Phi(X_{k-1} - \delta Z_i^* - h))\tau_k^x - \{\Phi(X_1 - \delta Z_i^*) - \Phi(X_1 - \delta Z_i^* - h)\}\tau_2^x}{\{\Phi(X_{N-1} - \delta Z_i^* - h) - \Phi(X_1 - \delta Z_i^* - h)\} - \{\Phi(X_1 - \delta Z_i^*) - \Phi(X_1 - \delta Z_i^* - h)\}}$$

Applying Taylor's theorem to each expression, we have

$$T_i^{x*} = \frac{\sum_{k=2}^{N-1} (\Phi(X_k - \delta Z_i^*) - \Phi(X_{k-1} - \delta Z_i^*))\tau_k^x - h\phi(X_{N-1} - \delta Z_i^*)\tau_{N-1}^x}{\{\Phi(X_{N-1} - \delta Z_i^*) - \Phi(X_1 - \delta Z_i^*)\} - h\phi(X_{N-1} - \delta Z_i^*)}$$

and

$$T_i^{x**} = \frac{\sum_{k=2}^{N-1} (\Phi(X_k - \delta Z_i^*) - \Phi(X_{k-1} - \delta Z_i^*))\tau_k^x - h\sum_{k=2}^{N-1} (\phi(X_k - \delta Z_i^*) - \phi(X_{k-1} - \delta Z_i^*))\tau_k^x}{\{\Phi(X_{N-1} - \delta Z_i^*) - \Phi(X_1 - \delta Z_i^*)\} - h\phi(X_1 - \delta Z_i^*)} \\ - \frac{h\phi(X_1 - \delta Z_i^*)\tau_2^x}{\{\Phi(X_{N-1} - \delta Z_i^*) - \Phi(X_1 - \delta Z_i^*)\} - h\phi(X_{N-1} - \delta Z_i^*)}$$

Using Taylor's theorem further

$$T_i^{x*} = T_i^x - h \frac{\phi(X_{N-1} - \delta Z_i^*)\tau_{N-1}^x}{\{\Phi(X_{N-1} - \delta Z_i^*) - \Phi(X_1 - \delta Z_i^*)\}} + h \frac{T_i^x \phi(X_{N-1} - \delta Z_i^*)}{\{\Phi(X_{N-1} - \delta Z_i^*) - \Phi(X_1 - \delta Z_i^*)\}^2}$$

and

$$T_i^{x**} = T_i^x - h \frac{\sum_{k=2}^{N-1} (\phi(X_k - \delta Z_i^*) - \phi(X_{k-1} - \delta Z_i^*))\tau_k^x - \phi(X_1 - \delta Z_i^*)\tau_2^x}{\{\Phi(X_{N-1} - \delta Z_i^*) - \Phi(X_1 - \delta Z_i^*)\}} \\ - h \frac{T_i^x \phi(X_{N-1} - \delta Z_i^*)}{\{\Phi(X_{N-1} - \delta Z_i^*) - \Phi(X_1 - \delta Z_i^*)\}^2}$$

Taking the difference between T_i^{x**} and T_i^* and letting h tend to zero

$$\lambda_i^x = \frac{dT_i^x}{dx_i^*} = - \frac{\sum_{k=2}^{N-1} (\phi(X_k - \delta Z_i^*) - \phi(X_{k-1} - \delta Z_i^*))\tau_k^x - \phi(X_1 - \delta Z_i^*)\tau_2^x - \phi(X_{N-1} - \delta Z_i^*)\tau_{N-1}^x}{\{\Phi(X_{N-1} - \delta Z_i^*) - \Phi(X_1 - \delta Z_i^*)\}}$$

Here the first term shows the effect of shunting some of the probability range across the thresholds. The second term corrects for the fact that the people who cross the upper threshold, X_{N-1} are excluded from the analysis, and the third term adjusts for the fact that the range is those observations lying between X_1 and X_{N-1} both before and after the increment.

To perform a similar calculation for children, we substitute out the fathers' latent variable, so that

$$Y_i^* = \delta\zeta Z_i^* + \zeta\varepsilon_i^x + \varepsilon_i^y$$

We need to take account of the fact that, while x_i^* is distributed with unit variance, the variance of y_i^* conditional on Z_i^* is $\sigma_y^2 = 1 + \zeta^2 + 2\rho_{xy}\zeta$. This implies that

$$\lambda_i^y = \frac{dT_i^y}{dy_i^*} = - \frac{\sum_{k=2}^{N-1} (\phi(\frac{Y_k - \delta\zeta Z_i^*}{\sigma_y}) - \phi(\frac{Y_{k-1} - \delta\zeta Z_i^*}{\sigma_y})) \tau_k^y - \phi(\frac{Y_1 - \delta\zeta Z_i^*}{\sigma_y}) \tau_2^y - \phi(\frac{Y_{N-1} - \delta\zeta Z_i^*}{\sigma_y}) \tau_{N-1}^y}{\left\{ \Phi(\frac{Y_{N-1} - \delta\zeta Z_i^*}{\sigma_y}) - \Phi(\frac{Y_1 - \delta\zeta Z_i^*}{\sigma_y}) \right\}}$$

allowing γ_i and thus γ_{OP} to be evaluated.

Both λ_i^x and λ_i^y and thus γ_i are functions of Z_i^* which is of course unobserved. We may, however, calculate their expected values conditional on social class n_i^Z being observed. We evaluate

$$\gamma_{n_i^Z} = \zeta \frac{\int_{Z_{n-1}}^{Z_n} \{\lambda^y(Z^*) / \lambda_i^x(Z^*)\} \phi(Z^*) dZ^*}{\Phi(Z_{n_i^Z}) - \Phi(Z_{n_i^Z-1})}$$

as the expected marginal impact conditional on a father from social class n_i^Z . The average marginal effect is then given as

$$\gamma_{OP} = \sum_i w_i \gamma_{n_i^Z} / \sum_i w_i$$

where n_i^Z is the social class of observation i .

For more information, please contact
ioe.llakescentre@ucl.ac.uk

LLAKES Centre
UCL Institute of Education
University College London
20 Bedford Way
London WC1H 0AL
U.K.



UNIVERSITY OF
Southampton
School of Education

