WESTMINSTER
BUSINESS SCHOOL

WORKING PAPER SERIES

2018/006

Censoring and Instrumental Variable Estimation: Biases in Estimates of the Relationship between Father's and Children's Years of Education

Richard Dorsett
Westminster Business School, University of Westminster, London

Cinzia Rienzo
King's College, London

Martin Weale
King's College, London

UNIVERSITY OF
VISION
STRATEGY
OPPORTUNITY
WESTMINSTER

Censoring and Instrumental Variable Estimation: Biases in Estimates of the Relationship between Fathers' and Children's Years of Education

Richard Dorsett[*], Cinzia Rienzo[†] and Martin Weale[‡]

4th December 2018

Abstract

We explore how censoring biases estimates of the intergenerational impact of education. Compulsory schooling censors the length of time spent in education, both for fathers and their children. We show that this will generally bias linear IV estimates and we identify conditions under which they remain consistent. We propose an IV ordered probit estimator as a flexible means of addressing censoring in the case of a discrete outcome. Our results suggest a substantial bias from ignoring censoring and a smaller bias from assuming normality. Viewing a binary instrument as the dichotomisation of a latent variable, we show how IV estimates are sensitive to the cut-point generating the dummy. This provides a potential explanation for IV estimates varying according to choice of instrument that is distinct from the usual attribution to impact heterogeneity.

Keywords: Instrumental Variables; Censored Regression; Ordered Probit; Fathers' and Children's Education

JEL Classification: C24, C26, C34, C36, I29

[*]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

[‡]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 Cohort Study, used in this study, was made available through the ESRC Data Archive. The data were originally collected by the Centre for Longitudinal Studies. 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

This paper explores the biases which can arise from censoring when relationships are estimated by means of instrumental variables. The particular example we use to illustrate this is the estimation of the relationship between fathers' and children's ages of completing education in the United Kingdom. Very material censoring arises because a high proportion of both fathers and their children left school when their compulsory education was completed rather than at a time obviously of their choice. We show in our application that commonly-used linear IV estimates lead to very serious bias, and suggest that this bias can be avoided by use of an ordered probit approach. We demonstrate that the biases in the linear IV estimates are close to what theory suggests on the assumption of normality and, depending on the choice of instrument, can be large. The bias associated with the assumption of normality is, in contrast, found to be relatively small.

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 discuss at length the issue of identification; how to separate the effects of parents' education on that of their children from other familial influences. One approach that has been used exploits increases in the legal minimum school-leaving age (Oreopoulos, Page & Stevens 2006). Our focus, as stated above, is on a different consequence of a legal minimum school-leaving age – the censoring it introduces to both fathers' and children's ages of completing schooling.

While empirical studies will often choose an analytical approach that takes account of the distribution of the dependent variable, the issue of regressor censoring is often left unaddressed. Rigobon & Stoker (2009), however, do discuss the biases in ordinary least squares (OLS) and linear instrumental variable (IV) regression when the regressor is censored.[1] Frandsen (2015) considers censored outcomes with an endogenous regressor. In this paper we consider the case where the dependent variable and the endogenous regressor are both censored.

We examine the bias of linear IV estimates and also identify conditions under which they will be unbiased. For the special case of normal errors and a continuous instrument, we provide an adjustment factor that can correct for the bias introduced by censoring. In the common case of a binary instrument, correcting linear IV estimates is not generally possible. Nevertheless, despite normality being a potentially restrictive assumption, the

[1] This follows Austin & Hoch (2004) who looked at OLS regression.

bias predicted under normality matches quite closely that seen with the linear IV results. We present as our preferred model a multivariate ordered probit. This specification provides additional flexibility, avoiding the assumption of normality for the dependent variable, the regressor and the instrument.

The estimation results suggest that the common approach of linear IV tends to overstate the degree to which paternal education influences child education. This appears likely to be due to differences between fathers and their children in the degree of censoring of their respective ages of completing education.

Our instrument – grandfathers' social class – is naturally ordered. But it also allows us to experiment with binary instruments constructed by dichotomising the social class variable at different points in its distribution. This is done in recognition of the fact that binary instruments are common in empirical research. The resulting estimates vary in a manner predicted by our theoretical analysis.

This finding has some relevance for the treatment effects literature. In particular, that literature typically allows impacts to vary across individuals and interprets IV estimates as capturing the mean impact on compliers (Imbens & Angrist 1994). Within that conceptual framework, changing the instrument changes the complier set, so variation in estimates is consistent with impact heterogeneity within the population. Our analysis highlights that censoring of the instrument is expected to influence estimates even under the assumption of homogeneous impacts within the population. Hence, censoring provides an alternative potential explanation for results being instrument-specific.

Lastly, we note that censoring arises commonly with economic variables for a variety of reasons. In some cases, the range of values over which a variable is defined is limited. The distribution of hours worked, for instance, is non-negative. In other cases, institutional factors censor distributions (statutory minimum wages are an obvious example). Yet another source of censoring arises from the practice of top-coding, sometimes used as a way of reducing disclosure risk in public-use datasets but also relevant to studies of age of completion of education using data collected before everyone has completed their education (de Haan & Plug 2011). In view of this, our results may find application in other settings.

The remainder of the paper has the following structure. Section 2 describes the bias that arises with linear IV estimation, what we are able to infer about bias in the special case where variables are censored normal and, lastly, our preferred approach which does not impose an assumption on the distribution of observed education. The empirical analysis is presented in Section 3. Section 4 concludes. Detailed derivations are provided in the Appendix.

2 Econometric issues arising with instrumental variable estimation and censoring

2.1 Linear IV

We denote by X_i the explanatory variable for observation i and Y_i the dependent variable. Z_i^* defines the (continuous) instrument used in estimation. X_i^* and Y_i^* denote the latent variables underlying the observed data. These latent variables have means μ_X, μ_Y and μ_Z respectively. In the example we discuss subsequently, X_i is the father's age of completing full-time education, Y_i is the child's age of completing full-time education and Z_i^* is a latent variable representing grandparental social class.

If Y_c is the censor point for Y_i^*

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

with a similar relationship holding for X_i and X_i^*. In our empirical example X_C and Y_C are compulsory minimum school leaving ages.

We assume that the underlying relationship we want to estimate is between the latent variables

$$Y_i^* = \gamma(X_i^* - \mu_X) + \mu_Y + \varepsilon_i^Y; \qquad \varepsilon_i^Y \text{ are iid}$$

Our interest is in the IV estimator of γ; 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

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

Censoring generates a bias such that γ_{IV}, unlike γ_{IV}^*, is no longer a consistent estimate of γ. The degree of bias varies 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.

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.

$$X_i^* = \delta(Z_i^* - \mu_Z) + \mu_x + \varepsilon_i^X \tag{1}$$

$$Y_i^* = \gamma(X_i^* - \mu_X) + \mu_Y + \varepsilon_i^Y \tag{2}$$

$$Z_i^* = \mu_Z + \varepsilon_i^Z \tag{3}$$

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

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 $\sigma_{XZ} = 0$

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

$$X_i^* = \mu_X + \delta\varepsilon_i^Z + \varepsilon_i^X \tag{5}$$

$$Y_i^* = \mu_Y + \gamma\left(\delta\varepsilon_i^Z + \varepsilon_i^X\right) + \varepsilon_i^Y \tag{6}$$

$$Z_i^* = \mu_Z + \varepsilon_i^Z \tag{7}$$

so that

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

We now establish sufficient conditions for the biases to cancel out. We normalise the variables, setting $s_X = \sqrt{\sigma_X^2 + \delta^2\sigma_Z^2}$, $s_Y = \sqrt{\sigma_Y^2 + \gamma^2\left(\sigma_X^2 + \delta^2\sigma_Z^2\right) + 2\gamma\sigma_{XY}}$ and $s_Z = \sigma_z$ so that $x_i^* = \frac{X_i^* - \mu_X}{s_X}$, $y_i^* = (\frac{Y_i - \mu_Y)^*}{s_Y}$ and $z_i^* = \frac{(Z_i^* - \mu_Z)}{s_Z.}$. We also define, for subsequent use, $\rho_{xy} = \frac{\gamma\left(\sigma_X^2 + \delta^2\sigma_Z^2\right) + \sigma_{XY}}{s_X s_Y}$, $\rho_{xz} = \delta\frac{s_Z}{s_X}$ and $\rho_{yz} = \frac{\gamma\delta s_Z}{s_Y}$.

Suppose that x_i^* and y_i^* are drawn from the same probability distribution, $f()$. Thus

$$f(x_i^*) = f(y_i^*). \tag{9}$$

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 location and scale, so that $x_c = (X_c - \mu_X)/s_X = y_c = (Y_c - \mu_Y)/s_Y$, then it follows immediately that $\frac{Cov(Z^*X^o)}{Cov(Z^*X)}$ $= \frac{Cov(Z^*Y^o)}{Cov(Z^*Y)}$ 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.

2.2 The case of normally distributed variables

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 observations, then

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

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, generated from an unobserved latent variable. It is common for empirical analysis to use binary instruments. For instance, in randomised trials, local average treatment effects [2] are usually

[2] Alternatively, 'complier average causal effects'

estimated using linear IV, with the randomisation outcome as instrument. Suppose that

$$Z_i = 0 \ \text{ if } Z_i^* \leq \ Z_c \tag{11}$$

$$Z_i = 1 \text{ if } \ Z_i^* > Z_c \tag{12}$$

The model is very similar to that of equations (1)-(4), differing only in that the relationship between the instrument and X_i^* is altered,

$$X_i^* = \delta(Z_i - E(Z_i)) + \mu_X + \varepsilon_i^X \tag{13}$$

$$Y_i^* = \gamma(X_i^* - \mu_X) + \mu_Y + \varepsilon_i^Y \tag{14}$$

$$Z_i^* = \mu_Z + \varepsilon_i^Z \tag{15}$$

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

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, the correlation between the normalised values x_i^* and z_i^* in the reduced form is

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

Cov(yz) is again evaluated by substitution. The IV estimator with a discrete instrument is

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

The analysis of section 2.1 remains valid, but the condition for the bias to cancel out has to reflect the change of instrument and becomes $\frac{Cov(ZX^o)}{Cov(ZX)} = \frac{Cov(ZY^o)}{Cov(ZY)}$.

Except in the special case when the censor/cut points are different from zero, it is not possible to de-bias through the application of a simple adjustment term, as was the case with a continuous instrument. The working behind this is shown Appendix A, which also provides a framework within which to explore the practical implications of censoring when variables are normal.

2.3 Allowing for non-normal observed distributions: the multivariate ordered probit

We present the multivariate ordered probit as our preferred specification, avoiding the restriction that the observed variables are censored normal. The ordered probit is a familiar model, founded on the idea of a continuous latent variable which determines the observed discrete values. In our application, the (observed) dependent variable, regressor and instrument are discrete realisations of their respective latent variables, with values reflecting the value of the latent variable relative to (empirically-estimated) cut-points. It is important to highlight that this is a different type of latent variable from that discussed so far which could alternatively be labelled 'uncensored'. In previous sections, 'latent' variables are identical to observed variables above the censoring point. In the ordered probit case, latent variables are never observed.

As is well-known, the ordered probit is suited to the case of ordered discrete variables, which need not approximate any particular continuous distribution. The cut-points are free to vary and can therefore map the latent variable, which is assumed normal, to an arbitrary discrete distribution. This provides the flexibility required in our application, where the assumption that the observed variables are censored normal is certainly incorrect (the variables are after all discrete) and potentially biasing.

The model in terms of latent variables is that of equations (1)-(3) but, in line with the discussion above, the nature of the latent variables themselves has changed. They are assumed to have zero mean and unit variance

$$X_i^{**} = \delta Z_i^{**} + \varepsilon_i^X \tag{18}$$

$$Y_i^{**} = \zeta X_i^{**} + \varepsilon_i^Y \tag{19}$$

$$Z_i^{**} = \varepsilon_i^Z \tag{20}$$

with

$$\begin{bmatrix} \varepsilon_i^X \\ \varepsilon_i^Y \\ \varepsilon_i^z \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho_{XY} & 0 \\ \rho_{XY} & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \right) \tag{21}$$

so that the latent variables all have zero mean. The parameter relating the latent dependent variable to the latent regressor variable is referred to as ζ to distinguish it from the parameter γ which related the variables in the linear model. As before, we impose the identifying restrictions, $\rho_{XZ} = 0$ and $\rho_{YZ} = 0$.

We define, with k, m, and n the discrete number of possible values of X_i,Y_i and Z_i respectively, cut points X_1^C to X_k^C, Y_1^C to Y_m^C and Z_1^C to Z_n^C

$$X_i^{**} \leq X_1^C \text{ if } X_i = 1;\ X_{j-1}^C < X_i^{**} \leq X_j^C \text{ if } X_i = j,\ (1 < j < k); X_k^C < X_i^{**} \text{ if } X_i = k;$$

$$Y_i^{**} \leq Y_1^C \text{ if } Y_i = 1;\ Y_{j-1}^C < Y_i^{**} \leq Y_j^C \text{ if } Y_i = j\ , (1 < j < m); Y_m^C < Y_i^{**} \text{ if } Y_i = m;$$

$$Z_i^{**} \leq Z_1^C \text{ if } Z_i = 1;\ Z_{j-1}^C < Z_i^{**} \leq Z_j^C \text{ if } Z_i = j\ , (1 < j < n); Z_n^C < Z_i^{**} \text{ if } Z_i = n.$$

Standard multivariate techniques can then be used to estimate the parameters of the model and, in large samples, these should not be sensitive to the cut points. With only the latter affected by censoring, it is possible to estimate the underlying parameters.

There is, however, a question of the interpretation of ζ. That shows the extent to which X^{**} influences Y^{**}. Unlike the earlier models, these latent variables do not at any point directly represent X and Y. With the ordered probit model, the expected marginal increase in Y associated with a marginal increase in X depends on the latter. Furthermore we can evaluate this only where $X \geq X^C$; the specification does not allow us to draw any implications below this point.

For each observation we can, however, work out the marginal relationships between X_i^{**} and Y_i^{**} and use these to translate ζ into a relationship, γ_i, between X_i and Y_i. The non-linearity means that that will be specific to each individual so we then average across the population to achieve an estimate of the mean marginal impact of X on Y.

We write $T_i^X = E\left(X_i^{**}\right)$ and $T_i^Y = E\left(Y_i^{**}\right)$ and then $\lambda_i^X = dT_i^X/dX_i^{**}$ and $\lambda_i^Y = dT_i^Y/dY_i^{**}$. Since $dY_i^{**}/dX_i^{**} = \zeta$ we can also write

$$\gamma_i = \frac{dT_i^Y}{dT_i^X} = \zeta\frac{\lambda_i^Y}{\lambda_i^X}$$

In appendix C we show, in the context of our empirical estimation, that with $\tau_k^X = \mathbb{1}\left(X_{k-1}^C < X_i^{**} \leq X_k^C\right)$ and $\tau_k^Y = \mathbb{1}\left(Y_{k-1}^C < Y_i^{**} \leq Y_k^C\right)$

$$\begin{aligned}\lambda_i^X(Z_i^{**}) &= \frac{dT_i^X}{dX_i^{**}} = \\ &-\frac{\sum_{k=2}^{N-1}(\phi(X_k^C-\delta Z_i^{**})-\phi(X_{k-1}^C-\delta Z_i^{**}))\tau_k^X-\phi(X_1^C-\delta Z_i^{**})\tau_2^x-\phi(X_{N-1}^C-\delta Z_i^{**})\tau_{N-1}^x}{\left\{\Phi(X_{N-1}^C-\delta Z_i^{**})-\Phi(X_1^C-\delta Z_i^{**})\right\}}\end{aligned}$$

Analogously, with $\sigma_Y = \sqrt{1+\zeta^2+2\rho_{XY}\zeta}$ being the standard deviation of Y_i^{**}

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

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 father's social class Z_i. We evaluate the effect for someone with a father in social class Z_i as

$$\gamma_i = \zeta\frac{\int_{Z_{i-1}^C}^{Z_i^C}\phi\left(Z_i^{**}\right)\left\{\lambda_i^Y\left(Z_i^{**}\right)/\lambda_i^X\left(Z_i^{**}\right)\right\}dZ_i^{**}}{\Phi\left(Z_i^C\right)-\Phi(Z_{i-1}^C)} \tag{22}$$

The expression is meaningful only for uncensored observations. We denote $\theta_i = 1$ if neither the father nor the child is censored and θ_i=0 otherwise. We then have

$$\gamma_{OP} = \sum_i \theta_i \gamma_i / \sum_i \theta_i. \tag{23}$$

3 The effect of paternal education: estimates from linear regression, censored normal and multivariate ordered probit regressions

The data we use to explore the issues described above are taken from the British Cohort Study, a study of 17,196 children born in one week in 1970. 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 workers are classified to social class I and managerial and technical workers to social class II. 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. Social class IV covers partly-skilled occupations and social class V unskilled occupations. Following convention we refer to a class with a lower number being higher than one with a higher number because it reflects higher social status.

Some of the fathers completed their education before the school-leaving age was increased to fifteen.[3] 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 6,036 observations being dropped. On top of this there is considerable attrition, giving us a final sample of 3,868 father-child pairs. A description of how weights were generated to control for the effects of attrition is provided in Appendix B. These weights were used throughout.

The use of grandparental social class as an instrument deserves some comment. For this to be legitimate requires that it not be correlated with child's education, other than via father's education. Doubts about this have led some researchers to explore, instead the effects of natural experiments arising from, for example, geographic differences in school-leaving ages or changes in the law on compulsory schooling at particular dates. Our main purpose, however, is to illustrate the effects of censoring and the way it can interact with the instrument. Grandparental social class is ideal in this respect.

[3] This was in April 1947 in Great Britain but ten years later in Northern Ireland.

Age at which Father completed Education	Grandfather's Class						
	I	II	III NM	III M	IV	V	All
15	13.9%	41.0%	38.9%	69.9%	74.2%	83.7%	63.4%
16	10.6%	18.1%	25.4%	15.2%	13.4%	9.8%	15.6%
17	14.5%	11.2%	11.1%	4.8%	5.4%	2.0%	6.4%
18	20.9%	9.3%	8.4%	3.7%	3.4%	2.3%	5.2%
19	1.6%	1.6%	1.4%	1.1%	0.2%	0.6%	1.0%
20	4.0%	1.1%	1.0%	0.4%	0.0%	0.3%	0.6%
21	9.5%	5.0%	4.6%	2.1%	0.9%	0.6%	2.6%
22	4.8%	4.4%	3.1%	1.0%	1.1%	0.7%	1.8%
23+	20.2%	8.3%	6.0%	1.7%	1.4%	0.0%	3.4%
Number (unweighted)	117	661	328	1818	650	294	3868

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

It is polychotomous and can therefore be used to construct a range of dichotomous instruments depending on where the cut point is placed. This allows us to show how our empirical results change in line with expectations as the position of the cut changes.

Table 1 shows the cross tabulation of fathers' age of completing education against 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 on fathers' ages of completion 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 right-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. These tables show that, for both children and their fathers, higher grandparental social class is associated with spending longer in education.

The remainder of this section presents estimates from the different modelling approaches. We begin with the linear IV case.

Since we observe six categories of social class, we can incorporate the full information on grandfathers' social class by including five independent dummy variables as instruments. This is different from the set-up in the previous sections, where there is a single binary instrument. Since the dummy variables are ordered, it is possible to consolidate them in order to carry out five different IV regressions, in each of which the instrument is a single dichotomous dummy, constructed by splitting the social class variable at different points in its distribution. Exploring this is informative of the common case where only a single dummy instrument is available and we wish to assess how the distribution

Age at which Child	Grandfather's Class						
completed Education	I	II	III NM	III M	IV	V	All
16	12.0%	30.3%	36.8%	51.5%	54.2%	60.9%	47.2%
17	10.0%	13.8%	12.9%	12.7%	12.9%	12.1%	12.8%
18	11.4%	15.8%	13.1%	13.3%	15.1%	15.6%	14.1%
19	9.7%	4.7%	3.9%	3.2%	3.2%	2.3%	3.6%
20	2.3%	2.7%	1.7%	1.8%	1.9%	0.0%	1.8%
21	13.5%	7.4%	6.6%	5.4%	3.6%	3.1%	5.5%
22	17.6%	11.9%	11.6%	4.6%	4.0%	1.8%	6.3%
23+	23.5%	13.4%	13.4%	7.5%	5.1%	4.2%	8.6%
Number (unweighted)	117	661	328	1818	650	294	3868

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

of that variable may influence estimates.

The linear IV results are summarised in the first row of table 3 (full results are given in table 5). 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. The table also shows, at the bottom, the proportion of respondents in each category.

The results show a clear tendency for the estimated coefficient to rise in line with the point at which the social class distribution is split to provide a dummy variable. 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?

We explore this initially under the assumption of normality, before presenting our preferred results where this is relaxed. We use Stata's *cmp* command to estimate the model set out in equations (1)-(4), but in a way which corrects for the effects of censoring. The continuous variable underlying social class, Z_i^* is not observed, but we assume that observed social class, Z_i is defined according a sequence of cut points, $Z_1^C \dots Z_5^C$, in the usual ordered probit formulation.

These are the assumptions of a multivariate censored normal model; the parameter estimates, γ_C, are shown in the second row of table 3 (see table 6 for full results). Once again they can be calculated with five social class dummies or with dichotomous instruments. The parameters are much more stable than in the linear IV case.

These censored normal results are useful to the extent that they allow us to apply

	Five Social Class Dummies	I	Grandfather's Social Class $\geqslant$II	$\geqslant$IIINM	$\geqslant$IIIM	$\geqslant$IV
γ_{IV}^{D}	0.844	0.786	0.807	0.858	1.000	1.034
	(0.058)	(0.096)	(0.067)	(0.064)	(0.106)	(0.140)
γ_C	0.604	0.706	0.635	0.608	0.592	0.554
	(0.041)	(0.069)	(0.048)	(0.043)	(0.070)	(0.091)
$\tilde{\gamma}_{IV}$	0.871^A	0.731	0.800	0.822	0.950	1.035
	(0.058)	(0.059)	(0.060)	(0.060)	(0.060)	(0.059)
$\tilde{\gamma}_{IV}^{0}$ if $x_c = y_c = 0$	0.604^A	0.558	0.577	0.584	0.629	0.662
	(0.041)	(0.043)	(0.041)	(0.041)	(0.038)	(0.036)
γ_{OP}	0.556	B	0.582	0.570	0.649	0.659
	(0.03)		(0.032)	(0.033)	(0.050)	(0.066)
Cumulative Proportion (weighted)		2.6%	18%	26%	74%	92%

Table 3: Parameter Estimates from Different Models

Notes:
Standard errors are shown in parentheses.
A. These parameters are calculated on the assumption that the instrument is continuous.
B γ_{OP} could not be estimated in the case where the instrument was a dummy indicating whether the grandfather was in social class I. This was probably because only 2.6% of the sample lay above the cut point.
Social Class I covers professional occupations, II managerial and technical occupations, IIINM skilled non-manual occupations, IIIM skilled manual occupations, IV partly-skilled occupations and V unskilled occupations

the analysis of section 2.2, to calculate how linear IV results would look when both children's and fathers' education are censored-normal and the instrument is continuous. Appendix D.2 describes the evaluation of equation (10) with a continuous instrument under normality. The bias then comes from the ratio of the proportion of children's observations that are uncensored to the proportion of fathers' observations that are uncensored (see sections A.1 and D.2), $\Phi\left(-y^C\right)/\Phi(-x^C) = 0.528/0.366$. Multiplying this by the estimate of the underlying relationship from the censored normal model, 0.604, yields an estimate for the parameter value associated with a continuous instrument of 0.871.[4]

This parameter, shown in first column of the third row of table 3 and labelled $\tilde{\gamma}_{IV}$, is not very far from the empirical estimate of 0.844 found using five social class dummies as instruments (first column, first row of table 3). While we do not offer any proof, we suggest that the two parameters are close because using five categorical dummies

[4]Since the parameters are not normally distributed the value depends on whether, as here, we use the observed probabilities to calculate the adjustment or whether we calculate the normalised censor points and then assume that the distribution of the underlying variables is normal. In the latter case we obtain a value of 0.907.

delivers a reasonable approximation to a single continuous instrument. Comparing this with the uncensored estimate of 0.604 shows the extent to which censoring biases the linear IV estimates when the instruments are linear.

We can perform an analogous exercise for the single-instrument specifications. The cut-points, Z^C, are derived from the cumulative probabilities shown at the end of table 3. In the third row of the table we also show the linear IV parameters generated by simulation under the assumption of normality and without any adjustment for censoring. Comparing these with the actual linear IV estimates in the first row reveals a close match, which might be taken to suggest that, with our data, the assumption of normality is a reasonable approximation, despite the discrete nature of the variables. We investigate this further in the next section. For now, we note simply the connection between the choice of instrument (i.e. Z_c) and the estimated parameter value.

The table includes in its fourth row and denoted by $\tilde{\gamma}^0_{IV}$ the estimates of the parameters which would be generated from an underlying parameter value of 0.604 if $x_c = y_c = 0$, i.e. if half of the fathers and children had completed their education at the statutory minimum age. When the instrument is a continuous variable, as is assumed here, the two proportions are equal and so the bias disappears; in the first column of table 3 $\gamma_C = \tilde{\gamma}^0_{IV}$. As the calculations of section A.2 make clear, that is no longer true when the instrument is dichotomous; the bias depends on the cut point of the dichotomous instrument. This effect is, however, much smaller than that arising because the censor points of the distributions of fathers' and children's ages of completion are very different from the estimated means of the underlying uncensored distributions. The results suggest that the bias arises primarily from the difference in the proportions of fathers and children completing their education at the minimum age, rather than from 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 fifth row of table 3 shows the results derived from the estimates of the multivariate ordered probit model of section 2.3. These estimates were obtained in Stata using the multivariate ordered probit procedure available in routine *cmp*. As discussed in subsection 2.3, the ordered probit parameter captures the relationship between latent variables and so must be transformed using equations (22) and (23) in order to be directly comparable to the linear IV or censored normal results.These results are calculated only for the population of fathers and children who continued their education beyond the legal minimum age because the expression cannot be evaluated for people affected by censoring.[5] We show results both with our model estimated using the full range of social classes and also with dichotomous classifications. We find it is not possible to estimate the model when the only distinction is made between those with a grandfather from social class I and those with grandfathers from other social classes. We assume that this is because only a small proportion of children fall into the first category.

Our results suggest, with a parameter estimate of 0.556, a smaller impact of paternal education than that suggested when assuming normality. However, there is closer agreement with those results than with the linear IV results, which take no account of censoring. This suggests that, in this case, the bias arising from assuming normality is considerably less than the bias arising from censoring. As with the censored normal model we see that, when dichotomous instruments are used, the coefficients show much less variation than when linear IV was used. Our preferred specification provides additional flexibility and generality and is strongly to the censored normal model on statistical grounds. The AIC and BIC for the ordered probit model are 31,318.6 and 31,540.9 respectively, compared to 33,212.2 and 33,288.4 for the censored normal model.

4 Conclusions

Using the relationship between fathers' and children's ages of completion of education as an example, we have shown, both analytically and empirically, the distortions which can arise when parameter estimates are produced by instrumental variables using data that are censored. In our application, 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 upward bias. Making the assumption that age of completing education is censored normally distributed, we provide expressions for the expected bias caused by censoring.

We find a close match between linear IV estimates and the values predicted under the assumption of normality. These show an upward bias compared to the underlying parameter estimate. While linear IV regression suggests a child's age of completing education rises by 0.844 years for each extra year that their father underwent full-time education, censored normal regression points to a coefficient of only 0.604. Our preferred model allows for censoring and non-normality and suggests a smaller impact still, of 0.556

[5] In applying this formula we set the upper cut point to that for age 29 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.

years. In this application the bias arising from the use of IV estimates with censored data is much greater than any bias arising from the assumption of normality.

Our results highlight the need to pay adequate regard to the issue of censoring. Furthermore, it not just censoring of dependent variables and regressors that is relevant, but also the related issue of how dichotomous instruments divide the population. Viewing such instruments as being generated by a latent variable crossing a threshold, the similarity to censoring is clear. We have shown how the bias depends on the distribution of the binary instrument. This finding has some relevance for the treatment effects literature. In that literature, impacts are allowed to vary across individuals and IV estimates are usually interpreted as capturing the mean impact of a treatment on compliers (Imbens & Angrist 1994). Altering the instrument changes the group of compliers such that differences in the resulting estimates are informative of differences between groups of compliers in their response to the treatment. Our analytical results indicate that, even if we assume the effect of paternal education is the same for everyone, different instruments will be expected to yield different estimates. This provides a potential explanation for results being instrument-specific that differs from the impact heterogeneity interpretation.

References

Austin, P. & Hoch, J. (2004), ‘Estimating Linear Regression Models in the Presence of a Censored Independent Variable’, *Statistics in Medicine* **23**, 411–429.

de Haan, M. & Plug, E. (2011), ‘Estimating Intergenerational Schooling Mobility on Censored Samples: Consequences and Remedies’, *Journal of Applied Econometrics* **26**, 151–166.

Frandsen, B. (2015), ‘Treatment effects with censoring and endogeneity’, *Journal of the American Statistical Association* **110**, 1745–1752.

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.

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.

A A Statistical Analysis of Censoring with Bivariate Normality

Write

$$X_i^* = \mu_X + \delta\varepsilon_i^Z + \varepsilon_i^X \tag{24}$$

$$Y_i^* = \mu_Y + \gamma\left(\delta\varepsilon_i^Z + \varepsilon_i^X\right) + \varepsilon_i^Y \tag{25}$$

$$Z_i^* = \mu_Z + \varepsilon_i^Z \tag{26}$$

where

$$\begin{bmatrix} \varepsilon_i^X \\ \varepsilon_i^Y \\ \varepsilon_i^Z \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_X^2 & \sigma_{XY} & 0 \\ \sigma_{XY} & \sigma_Y^2 & 0 \\ 0 & 0 & \sigma_Z^2 \end{bmatrix}\right)$$

so that

$$\mathbf{V} = Cov \begin{bmatrix} X_i^* \\ Y_i^* \\ Z_i^* \end{bmatrix} = \begin{pmatrix} \sigma_X^2 + \delta^2\sigma_Z^2 & \gamma\left(\sigma_X^2 + \delta^2\sigma_Z^2\right) + \sigma_{XY} & \delta\sigma_Z^2 \\ \gamma\left(\sigma_X^2 + \delta^2\sigma_Z^2\right) + \sigma_{XY} & \sigma_Y^2 + \gamma^2\left(\sigma_X^2 + \delta^2\sigma_Z^2\right) + 2\gamma\sigma_{XY} & \gamma\delta\sigma_Z^2 \\ \delta\sigma_Z^2 & \gamma\delta\sigma_Z^2 & \sigma_Z^2 \end{pmatrix} \tag{27}$$

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

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

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

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

We set

$$\begin{aligned} s_X &= \sqrt{\sigma_X^2 + \delta^2\sigma_Z^2};\ s_Y = \sqrt{\sigma_Y^2 + \gamma^2\left(\sigma_X^2 + \delta^2\sigma_Z^2\right) + 2\gamma\sigma_{XY}};\ s_Z = \sigma_Z \\ \rho_{xy} &= \frac{\gamma\left(\sigma_X^2 + \delta^2\sigma_Z^2\right) + \sigma_{XY}}{s_X s_Y};\ \rho_{xz} = \frac{\delta s_Z}{s_X};\ \rho_{yz} = \frac{\gamma\delta s_Z}{s_Y} \end{aligned}$$

so that

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

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 $Z_i^* \geq Z_c + \mu_Z$. 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

$$\begin{aligned}
x_i^* &= \frac{X_i^* - \mu_X}{s_X};\ x_i = \frac{X_i - \mu_X}{s_X} \text{ and } x_c = \frac{X_c - \mu_X}{s_X}. \\
y_i^* &= \frac{Y_i^* - \mu_Y}{s_Y};\ y_i = \frac{Y_i - \mu_Y}{s_Y} \text{ and } y_c = \frac{Y_c - \mu_Y}{s_Y} \\
z_i^* &= \frac{Z_i^* - \mu_Z}{s_Z};\ z_i = \frac{Z_i - \mu_Z}{s_Z} \text{ and } z_c = \frac{Z_C - \mu_Z}{s_Z};
\end{aligned}$$

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 bias from censoring when the instrument is fully-observed

We set out here the bias arising when $Cov(xz^*)$ is used in place of the covariance of the uncensored data, $Cov(x^*z^*)$. The bias associated with $Cov(yz^*)$ can then be evaluated simply by substituting y for x in the resulting formulae, and the impact on the IV estimator can then be calculated.

We consider separately the cases where $x_i > x_c$ (equivalently, $x_i^* > x_c$) and $x_i{=}x_c$ ($x_i^* \leq x_c$).

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)[6]

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

2. $x_i = x_c$

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

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

Since the first moment of $z_i^* = 0$, r_{xz} =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, simply by substituting y for x we have

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

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 bias from censoring when the instrument is a dichotomised latent variable

Once again, it is adequate to focus on the $Cov(xz)$ with $Cov(yz)$ evaluated by substitution. 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

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

so the estimate of the covariance of the normalised variables is

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

Similarly

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

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

$$\gamma_{IV}^{D} = \frac{\hat{s}_{yz}}{\hat{s}_{xz}} \frac{s_Y}{s_X}$$

showing a clear bias, if one which is less straightforwardly represented than with the continuous 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 Adjusting for non-response

Our empirical analysis uses data weighted to adjust for survey non-response. We relate the probability of dropping out of the survey to characteristics reported in the original 1970 survey, at least for the vast majority of respondents. Table 4 shows the results of a probit model of drop-out. The probability of a child remaining in the 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 10,494 respondents out of the total initial sample of 17,196 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.

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

C Interpretation of the Ordered Probit Model

A general model relaxes the assumption of censored normality and instead allows outcomes to be generated by an ordered probit model. The cut-points of this model allow it to approximate an arbitrary discrete distribution of observed outcomes. The model is motivated by the empirical analysis of the relationship between the years of education of fathers and their children so we present it in this section with this in mind.

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

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

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

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

$$\gamma_i = \frac{dT_i^Y}{dT_i^X} = \zeta \frac{\lambda_i^Y}{\lambda_i^X}$$

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

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

We are interested in the effect that a small increase, h in $X_i^{**} = \delta Z_i^{**}$ 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, $T_i^{X^o}$ for the expected age of completion of education for someone with a latent variable of $X_i^{**} = \delta Z_i^{**}$, and $T_i^{X^{oo}}$ for someone with a latent variable $X_i^{**} = \delta Z_i^{**} + h$. We then evaluate

$$\frac{dT_i^X}{dX_i^{**}} = \begin{array}{c} Limit \\ h \to 0 \end{array} \frac{T_i^{X^{oo}} - T_i^{X^o}}{h}$$

First,

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

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 $T_i^{X^{oo}}$ the ranges are shifted by h. The upper limit is, however, X_{N-1}.

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

Applying Taylor's theorem to each expression, we have

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

and

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

Using Taylor's theorem further

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

and

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

Taking the difference between $T_i^{X^{oo}}$ and $T_i^{X^o}$ and letting h tend to zero

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

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

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

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

	Five Social Class Dummies	Grandfather's Social Class I	$\geqslant$II	$\geqslant$IIINM	$\geqslant$IIIM	$\geqslant$IV
γ_{IV}^{D}	0.844***	0.786***	0.807***	0.858***	1.000***	1.034***
	(0.058)	(0.096)	(0.067)	(0.064)	(0.106)	(0.140)
Constant	4.323***	5.255***	4.918***	4.086***	1.811	1.261
	(0.924)	(1.551)	(1.082)	(1.031)	(1.697)	(2.254)
N	3868	3868	3868	3868	3868	3868
Kleinbergen-Paap	310	58.9	190	256.6	159.8	107.2
Sargan	χ_4^2=4.35					
Percentage Dummy=1		2.6%	18%	26%	74.3%	91.7%

Table 5: IV Coefficient Estimates as Functions of the Cut Point for the Dummy Instrumental Variable

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 Z_i of observation i being observed. We evaluate the effect for someone with a father in social class Z_i as

$$\gamma_i = \zeta \frac{\int_{Z_{i-1}^C}^{Z_i^C} \phi\left(Z_i^{**}\right) \left\{\lambda_i^Y\left(Z_i^{**}\right) / \lambda_i^X\left(Z_i^{**}\right)\right\} dZ_i^{**}}{\Phi\left(Z_i^C\right) - \Phi(Z_{i-1}^C)}$$

as the expected marginal impact conditional on a father from social class Z_i.

The expression is meaningful only for uncensored observations. We denote $\theta_i = 1$ if neither the father nor the child is censored and θ_i=0 otherwise. We then have

$$\gamma_{OP} = \sum_i \theta_i \gamma_i / \sum_i \theta_i.$$

D Full estimation results

D.1 Linear IV results

The linear IV results are presented in table 5.

The Kleinbergen-Paap statistic does not point to any concerns that the instruments are weak in any of the regressions. In the first column, the five dummies allow Sargan's over-identification test to be carried out, with an acceptable result.

D.2 Censored normal results

The censored normal results are presented in table 6.

It should be noted that a closely related specification is provided by replacing equation (2) by

$$Y_i^* = \gamma X_i + \varepsilon_i^Y \tag{28}$$

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 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 6. These log-likelihoods suggest strongly that the latent variable model of equation (2) should be preferred to the actual variable model of equation (28).

The estimated 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, $\mathbf{\Sigma}$

$$\mathbf{V} = \begin{bmatrix} 21.442 & 9.365 & -1.835 \\ 9.365 & 17.700 & -1.108 \\ -1.835 & -1.108 & 1 \end{bmatrix}; \quad \mathbf{\Sigma} = \begin{bmatrix} 1 & 0.404 & -0.396 \\ 0.404 & 1 & -0.263 \\ -0.396 & -0.263 & 1 \end{bmatrix}$$

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

$$\gamma_{IV}^* = \frac{\mathbf{V}_{2,3}}{\mathbf{V}_{1,3}} = \frac{\mathbf{\Sigma}_{2,3}}{\mathbf{\Sigma}_{1,3}} \sqrt{\frac{\mathbf{V}_{2,2}}{\mathbf{V}_{1,1}}} = 0.604$$

In order to explore the biases arising from censoring we work from matrix $\mathbf{\Sigma}$, so as to exploit the analysis of section 2.2. 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.

We can see from tables 1 and 2 that 63.4% of fathers and 47.2% of children left school at the legal minimum age. These imply that $x^C = \Phi^{-1}(0.472) = -0.0702$ and $y^C = \Phi^{-1}(0.634) = 0.343$ where Φ^{-1} is the inverse cumulative normal distribution. These values are used in the application of the results of section 2.2 and appendix A to give the values for $\tilde{\gamma}_{IV}$ in table 3.

	Grandfather's Class					
	I	II	III NM	III M	IV	V
			Child's Age of Completion			
Constant	8.344***	6.995***	7.940***	8.307***	8.515***	9.021***
	(0.583)	(0.950)	(0.674)	(0.604)	(0.954)	(1.233)
γ_C	0.604***	0.706***	0.635***	0.608***	0.592***	0.554***
	(0.041)	(0.069)	(0.048)	(0.043)	(0.070)	(0.091)
			Father's Age of Completion			
Constant	13.340***	13.310***	13.318***	13.301***	13.324***	13.321***
	(0.110)	(0.112)	(0.112)	(0.112)	(0.111)	(0.111)
δ	-1.835***	-2.289***	-2.052***	-2.195***	-1.437***	-1.528***
	(0.093)	(0.156)	(0.117)	(0.110)	(0.121)	(0.168)
			Cut Points			
Cut 1	-1.964***	-1.945***				
	(0.040)	(0.040)				
Cut 2	-0.916***		-0.914***			
	(0.023)		(0.023)			
Cut 3	-0.640***			-0.644***		
	(0.022)			(0.021)		
Cut 4	0.656***				0.654***	
	(0.022)				(0.023)	
Cut 5	1.374***					1.385***
	(0.030)					(0.031)
			Variance-covariance			
$\log \sigma_X$	1.447***	1.403***	1.431***	1.417***	1.487***	1.481***
	(0.024)	(0.029)	(0.025)	(0.025)	(0.024)	(0.026)
$\log \sigma_Y$	1.327***	1.357***	1.334***	1.328***	1.321***	1.313***
	(0.020)	(0.031)	(0.022)	(0.021)	(0.024)	(0.024)
$tanh^{-1}\sigma_{XY}/(\sigma_X\sigma_Y)$	-0.228***	-0.383***	-0.272***	-0.244***	-0.200*	-0.151
	(0.049)	(0.093)	(0.060)	(0.054)	(0.088)	(0.119)
N	3868	3868	3868	3868	3868	3868
Log-Lik.	-14934	-10758.6	-11820.6	-12106.9	-12216.8	-11322.4
Log-Lik. (28)	-14996	-10788	-11873	-12164.6	-12245.2	-11338.5

The parameters are identified by setting the variance of ε_i^Z to 1 and the covariances σ_{XZ} and σ_{YZ} to 0.

Table 6: Parameter Estimates allowing for Censoring when Child's Age of Completion is influenced by Father's Latent Age of Completion

D.3 Multivariate ordered probit results

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

Note that there are no observations with $X_i = 30$ or 31.

Table 7: The Parameters of the Ordered Probit Model