Bias in regressions with censoring on both sides: an application to the relationship between fathers' and children's education.

11th June 2017

Abstract

We explore the biases when estimating the relationship between two censored endogenous variables. We show how such censoring affects linear IV estimates but also identify conditions under which they are consistent. We propose an IV ordered probit estimator as a flexible means of addressing censoring in the case of a discrete outcome. We illustrate by estimating the relationship between fathers' and children's education. 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 results 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.

1 Introduction

Under the usual OLS assumptions, a straightforward way of dealing with a censored regressor is to base estimation on the subgroup of the sample for which that regressor is uncensored. When the dependent variable is also censored, however, non-linear models are required. Discarding observations with a censored regressor is no longer valid since the underlying relationship among such observations may differ from that observed among retained observations. Even under the assumption of homogeneity in the latent model, estimates that ignore regressor censoring can be biased.

This paper explores the biases that can arise when estimating the relationship between two censored endogenous variables. Rigobon & Stoker (2009) discuss the biases in ordinary least squares (OLS) and linear instrumental variable (IV) regression when the explanatory variable is censored.[1] Frandsen (2015) considers censored outcomes with an endogenous regressor. In this paper, we consider the case of a censored dependent variable and a censored endogenous regressor. We examine the bias of linear IV estimates and also identify conditions under which they will be unbiased. We show that if errors are normal and the instrument is continuous, IV estimates will be biased by a factor that is a function of the degrees of censoring of both variables and can therefore be straightforwardly corrected. If the degree of censoring is the same on both sides of the regression, the IV estimate will be unbiased without adjustment. Hence, there are conditions under which an IV estimate can be more robust than its OLS counterpart. In the case of a binary instrument no simple adjustment is available to correct IV estimates for bias, although we again derive conditions under which such estimates will be unbiased. Lastly, since the assumption of censored normality will not always be appropriate, we suggest an ordered probit IV model as a more flexible alternative.

We illustrate these points through an empirical analysis of the relationship between fathers' and children's ages of completing education. Censoring in this case arises due to the

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

minimum school-leaving age. Among both fathers and children, a proportion were required to remain in school longer than they would have freely chosen. An examination of the relationship therefore involves censoring of both the dependent variable and the regressor. To address endogeneity of the regressor, we use a dummy variable for the social class of the child's paternal grandfather as an instrument. Diagnostic tests based on a polychotomous social class indicator suggest that the exclusion restriction is valid, allowing our estimates to be interpreted as capturing a causal relationship.

We show how, under the assumption that the underlying variables are normally distributed, the linear IV estimate depends on the cut point of the dummy variable. With our data, linear IV methods tend to overstate the influence of paternal age of completing education. This appears likely to be due to a higher proportion of fathers than children completing their education at the statutory minimum age. We find the estimated regression coefficients are very much in line with theoretical expectations, suggesting that the sensitivity of the IV coefficient to the choice of instrument is indeed explained by the interaction of censoring and the cut point of the instrument. In the case of heterogeneous effects, Imbens & Angrist (1994) show how IV estimates can be interpreted as capturing the local average treatment effect (LATE); that is, the average effect of treatment for compliers. Varying the instrument changes the complier set, possibly resulting in a different LATE estimate. Our results show that, even in the case of homogeneous effects, estimates vary with the cut point of the instrument. This provides a potential explanation for results being instrument-specific that is distinct from the usual attribution to impact heterogeneity.

Our empirical analysis, while illustrative, draws attention to a very material issue; one of the data sets widely used to explore the connection between fathers' and children's education in the United Kingdom, the British Cohort Survey suggests that 59% of the fathers and 45% of the children[2] 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

[2] 63% of fathers and 47% of children after reweighting for attrition. See section 4.

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.

The remainder of the paper has the following structure. Section 2 describes the bias in the linear IV case. Section 3 considers the case when the data are censored normal or are generated by a latent normal model. Detailed derivations are provided in the Appendix. The empirical analysis is presented in Section 4. Section 5 concludes.

2 Instrumental Variable Estimation and Censoring

We denote by X_i the explanatory variable for observation i and Y_i the dependent variable. Z_i^* defines the instrument used in estimation. X_i^* and Y_i^* denote the latent variables underlying the observed data. These latent variables are all measured relative to their means. In the example we discuss subsequently, X_i is the father's age of leaving school, Y_i is the child's age and the instrument is a 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

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

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

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

Despite impacts being homogeneous, 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.

$$X_i^* = \delta Z_i^* + \varepsilon_i^X \tag{1}$$

$$Y_i^* = \gamma X_i^* + \varepsilon_i^Y \tag{2}$$

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

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

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

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

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

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

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 $x_i^* = \frac{X_i^*}{s_X}$, $y_i^* = \frac{Y_i^*}{s_Y}$ and $z_i^* = \frac{Z_i^*}{s_Z}$. 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 scale, so that $x_c = X_c/s_X = y_c = Y_c/s_Y$ 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. 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 observations, then

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

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, generated from an unobserved latent variable. This is more relevant to our empirical example, because the indicator of social class is a discrete, not a continuous variable. Suppose that

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

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

The model then becomes

$$X^*_i = \delta Z_i + \varepsilon^X_i \tag{13}$$

$$Y^*_i = \gamma X^*_i + \varepsilon^Y_i \tag{14}$$

$$Z^*_i = \varepsilon^Z_i \tag{15}$$

$$E\begin{bmatrix}\varepsilon^X_i \\ \varepsilon^Y_i \\ \varepsilon^Z_i\end{bmatrix} = 0, \quad Cov\begin{bmatrix}\varepsilon^X_i \\ \varepsilon^Y_i \\ \varepsilon^Z_i\end{bmatrix} = \begin{bmatrix}\sigma^2_X & \sigma_{XY} & 0 \\ \sigma_{XY} & \sigma^2_Y & 0 \\ 0 & 0 & \sigma^2_Z\end{bmatrix} \tag{16}$$

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

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

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

$$\gamma_{IV}^D = \frac{Cov(yz)}{Cov(xz)}. \tag{18}$$

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 $\gamma_{IV}^D = \gamma_{IV}^*$. Second, 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.

These calculations set out a framework in which to explore the practical implications of censoring. Before we explore that, however, we set out a means of relaxing the two key assumptions we have made so far. First, it has been assumed that the relationship between the explanatory and the dependent variable is linear and secondly it was assumed that they are 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 related to another endogenous censored variable.

3.1 Allowing for Non-normality using an Ordered Probit Model

An alternative means of estimating the latent variable model, with weaker distributional assumptions, is to treat the data for the explanatory and dependent variables as the cut points in a multivariate ordered probit model. With this model, the cut points are free to vary and can represent an arbitrary distribution. This avoids imposing both of the above assumptions, linearity and normality, and can accommodate both left- and right-censoring should that be necessary. Normality of latent variables is of course required but the flexibility of the cut points means that is not restrictive. In addition to these modifications to the first two equations of our model, the third equation of the model, describing the latent variable which underlies a discrete instrument, can also be set out in ordered probit form.[3]

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

$$X_i^* = \delta Z^* + \varepsilon_i^X \tag{19}$$

$$Y_i^* = \zeta X_i^* + \varepsilon_i^Y \tag{20}$$

$$Z_i^* = \varepsilon_i^Z \tag{21}$$

with

$$\begin{bmatrix} \varepsilon_i^X \\ \varepsilon_i^Y \\ \varepsilon_i^z \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho_{XY} & 0 \\ \rho_{XY} & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \right) \tag{22}$$

so that the latent variables all have zero mean. The parameter relating the latent dependent variable to the latent actual 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

[3]Note that the ordered probit specification does not distinguish between a model where X influences Y and a model where X^* influences Y.

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

$$
\begin{aligned}
X_i^* &\leq X_1^C \text{ if } X_i = X_1;\ X_{j-1}^C < X_i^* \leq X_j^C \text{ if } X_i = X_j,\ (1 < j < k); X_k^C < X_i^* \text{ if } X_i = X_k; \\
Y_i^* &\leq Y_1^C \text{ if } Y_i = Y_1;\ Y_{j-1}^C < Y_i^* \leq Y_j^C \text{ if } Y_i = Y_{j-1}\ , (1 < j < m); Y_m^C < Y_i^* \text{ if } Y_i = Y_m; \\
Z_i^* &\leq Z_1^C \text{ if } Z_i = Z_1;\ Z_{j-1}^C < Z_i^* \leq Z_j^C \text{ if } Z_i = Z_{j-1}\ , (1 < j < n); Z_n^C < Z_i^* \text{ if } Z_i = Z_n.
\end{aligned}
$$

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; we discuss this in the empirical section.

4 Illustration: the Influence of Compulsory Schooling on Estimates of the Relationship between Fathers' and Children's Education

4.1 Background

We now turn to our example, looking at the relationship between fathers' and children's ages of completing education, and the way in which IV estimates of the coefficient relating them are influenced by the proportion of the population affected by the compulsory minimum school-leaving age.

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. 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. Their approach delivers unbiased estimates of the average effect of extended parental education as a result of changed compulsion on the education of their children. At the same time it brings to the fore the question of how best to deal with the effects of compulsion.

4.2 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 I, 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.

Some of the fathers completed their education before the school-leaving age was increased to fifteen.[4] 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 out an initial 17,196 children. 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

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

Age at which Father	Grandfather's Class						
completed Education	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)

to control for the effects of attrition is provided in section 1 of the supplementary material. These weights were used throughout.

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 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.

4.3 Linear IV Estimates

The first stage in assessing the importance of bias is to examine linear IV estimates. As is clear from section 4.2, 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 possible IV

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)

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 the stnadard test for over-identification (Sargan 1958) 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 3 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.[5] 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 (P=0.36), 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

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

	Five Social Class Dummies	Grandfather's Social Class I	⩾II	⩾IIINM	⩾IIIM	⩾IV
γ_D^{IV}	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^2_4=4.35					
Percentage Dummy=1		2.6%	18%	26%	74.3%	91.7%

Table 3: IV Coefficient Estimates as Functions of the Cut Point for the Dummy Instrumental Variable

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?

4.4 Estimates Corrected for Censoring

The first step in examining this is to estimate the counterparts to the models of table 3 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 3 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 3 match the theoretical implications conditional on normality.

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

$$X_i^* = \mu_X + \delta Z_i^* + \varepsilon_i^X \tag{23}$$

$$Y_i^* = \mu_Y + \gamma X_i^* + \varepsilon_i^Y \tag{24}$$

$$Z_i^* = \varepsilon_i^Z \tag{25}$$

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

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

By analogy with the earlier models, we can estimate the system using an ordered probit model for equation (25) 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 to 1 and the covariances σ_{XZ} and σ_{YZ} to 0. The results of this are shown in table 4. It can be seen that the parameter γ is much more stable across the different specifications than in table 3; 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 (24) by

$$Y_i^* = \mu_Y + \gamma X_i + \varepsilon_i^Y \tag{26}$$

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 4. These log-likelihoods suggest strongly that the latent variable model of equation (24) should be preferred to the actual variable model of equation (26).

The estimation of equations (23) to (25) 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

	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	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 σ_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 σ_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. Eqn (26)	-14996	-10788	-11873	-12164.6	-12245.2	-11338.5

Table 4: Parameter Estimates allowing for Censoring when Child's Age of Completion is influenced by Father's Latent Age of Completion

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 4 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, $\mathbf{\Sigma}$

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

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

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

In order to explore the biases arising from censoring we work from matrix $\mathbf{\Sigma}$, 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 $x_c = 0.34$ and $y_c = -0.07$ corresponding to weighted proportions of fathers and children completing their education at the statutory minimum age of 63.4% and 47.2% respectively. Equation (10) implies that, if the latent instrument were observed, it would deliver an estimate of the parameter, $\gamma^l_{IV} = 0.87$ in contrast to the true parameter of 0.60. In table 5 we show the cut points for the latent variable underlying the five dichotomous instruments of table 3 together with the theoretical estimates of the IV parameter, γ^D_{IV}, which would be generated by using these dummies, with no correction for the effects of censoring. These are calculated using equation (18) and compared with the estimates from table 3. Finally we show in the table the estimates of the parameters which would be 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.

	Grandfather's Social Class				
	I	$\geqslant$II	$\geqslant$IIINM	$\geqslant$IIIM	$\geqslant$IV
Instrument Cut Point (table 4)	-1.964	-0.916	-0.640	0.657	1.374
Theoretical γ^D_{IV}	0.731	0.800	0.823	0.949	1.033
	(0.058)	(0.059)	(0.059)	(0.058)	(0.057)
Estimate (table 3)	0.786	0.807	0.858	1.000	1.034
	(0.096)	(0.067)	(0.064)	(0.106)	(0.140)
γ^D_{IV} if $x_c = y_c = 0$	0.558	0.577	0.584	0.629	0.662
	(0.043)	(0.041)	(0.041)	(0.038)	(0.036)

Standard errors are shown in brackets. For γ^D_{IV} these are calculated by making five thousand draws of the theoretical parameters assuming that the latter are jointly normally distributed around the values shown in table 3 with the estimated covariance matrix.

Table 5: Parameter Estimates generated by a Censored Normal Distribution

This table shows the connection between the choice of instrument (i.e. 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 3 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.

4.5 An Ordered Probit Model

We fit the ordered probit model of section 3.1. 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

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

The parameters of the model can then be estimated in Stata using the multivariate ordered probit procedure available in routine *cmp*. It should be noted that there are no observations with $X_i = 30$ or 31. The results are shown in table 6.

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 4. 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.[6] A counterpart of this is that the cut points shown in table 6 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

[6] The 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 4 (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***	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				

Table 6: The Parameters of the Ordered Probit Model

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 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 section 2 of the supplementary material that, with τ_k^X being the age of completion of education associated with fathers whose latent variables lie between cut point $k-1$ and cut point k, and τ_k^Y the equivalent for their children, 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, with $\sigma_Y = \sqrt{1 + \zeta^2 + 2\rho_{XY}\zeta}$ being the standard deviation of Y_i^* conditional on Z_i^*,

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

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 social class n_i^Z being observed. We

	Whole Sample	Restricted Sample	Grandfather's Class 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
	(0.03)	(0.03)	(0.02)	(0.03)	(0.03)	(0.03)	(0.04)	(0.04)

Standard errors are shown in brackets. These are calculated by making five thousand draws of the theoretical parameters assuming that the latter are jointly normally distributed around the values shown in table 6 with the estimated covariance matrix.

Table 7: Estimates of the Average Marginal Impact of an Extension of Father's Education on that of Children

evaluate

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

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

$$\gamma_{OP} = \sum_i w_i \gamma_{n_i^Z} / \sum_i w_i \tag{27}$$

where n_i^Z is the grandpaternal 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 7 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 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 normal model (table 4) and 0.84 estimated by linear IV (table 3).

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 upward bias. 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 bias.

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 linear IV estimates using these dummy variables and the values predicted by our theoretical analysis under 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 any bias arising from the assumption of normality.

Our results highlight the need to pay adequate regard to the issue of censoring. Furthermore, where IV results are sensitive to the choice of instrument, they offer an alternative to the explanation of 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 at which latent instrumental variables are dichotomised; which 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.

Dearden, L., Machin, S. & Reed, H. (1997), 'Intergenerational Mobility in Britain', *Economic Journal* **107**, 47–66.

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.

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.

Sargan, J. (1958), 'The Estimation of Economic Relationahips using Instrumental Variables', *Econometrica* **28**, 393–415.

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

A Appendix: A Statistical Analysis of Censoring with Bivariate Normality

Write the three variables of interest as

$$X_i^* = \delta\varepsilon_i^Z + \varepsilon_i^X \tag{28}$$

$$Y_i^* = \gamma\left(\delta\varepsilon_i^Z + \varepsilon_i^X\right) + \varepsilon_i^Y \tag{29}$$

$$Z_i^* = \varepsilon_i^Z \tag{30}$$

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

$$\mathbf{V} = Cov\begin{bmatrix} X_i^* \\ Y_i^* \\ Z_i^* \end{bmatrix} = \begin{pmatrix} \sigma_X^2 + \delta^2\sigma_Z^2 & \gamma\left(\sigma_X^2 + \delta^2\sigma_Z^2\right) + \sigma_{XY} & \delta\sigma_Z^2 \\ \gamma\left(\sigma_X^2 + \delta^2\sigma_Z^2\right) + \sigma_{XY} & \sigma_Y^2 + \gamma^2\left(\sigma_X^2 + \delta^2\sigma_Z^2\right) + 2\gamma\sigma_{XY} & \gamma\delta\sigma_Z^2 \\ \delta\sigma_Z^2 & \gamma\delta\sigma_Z^2 & \sigma_Z^2 \end{pmatrix} \tag{31}$$

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

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

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 $Z_i^* \geq Z_c + \mu_Z$; similarly, we write $z_i = 1$ if $z_i^* \geq z_c$. 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

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

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 $x_i^* > x_c$ and $x_i^* \leq x_c$.

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)[7]

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

2. $x_i = x_c$

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

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

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

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 dichotomous 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)

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

[7]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].

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.