Endogeneity and Discrete Outcomes

Andrew Chesher*
Centre for Microdata Methods and Practice, UCL

April 10th 2007

This paper is dedicated to Tony Lancaster on the occasion of his Festschrift Conference, April 14th-15th 2007, at Brown University

Abstract. This paper studies models for discrete outcomes which permit explanatory variables to be endogenous. Interesting models for discrete outcomes that admit endogeneity necessarily involve a structural function which is non-additive function of a latent variate. In the essentially single equation models considered here this latent variate is restricted to be locally independent of instruments but the models are silent about the nature of dependence between the latent variate and the endogenous variable and the role of the instrument in this relationship. These IV models which, when the outcome is continuous, can have point identifying power, have only set identifying power when the outcome is discrete. Identification regions shrink as the support of a discrete outcome grows. The paper extends the analysis of structural quantile functions with endogenous arguments to cases in which there are discrete outcomes, cases which have so far been excluded from consideration. The results point to a neglected consequence of interval censoring and grouping, namely the loss of point identifying power that can result when endogeneity is present.

Keywords: Partial identification, Nonparametric methods, Nonadditive models, Discrete distributions, Ordered probit, Poisson regression, Binomial regression.

1. Introduction

This paper studies models for discrete outcomes which permit explanatory variables to be endogenous. Outcomes can be binary, or integer valued such as arise when considering counts, or ordered as might be obtained when there is interval censoring of a latent continuous outcome.

The scalar discrete outcome is determined by a structural function

$$Y = h(X, U)$$

where U is a continuously distributed, unobserved, scalar random variable and X is an observable vector random variable. There is endogeneity in the sense that U and

*I am grateful to Simon Lee, Lars Nesheim, Adam Rosen and Richard Spady for helpful discussions. The support of the Leverhulme Trust though a grant to the research project *Evidence Inference and Inquiry* and through a grant to the Centre for Microdata Methods and Practice (cemmap) is gratefully acknowlegded. This is a revised version of cemmap Working Paper CWP 05/07. Some of the results were presented at the Oberwolfach Workshop *Semiparametric and Nonparametric Methods in Econometrics* on March 19th 2007. I am grateful to participants for comments.

X may not be independently distributed. There are instrumental variables, a vector random variable Z, with the property that for some $\tau \in (0,1)$ and all z

$$\Pr[U \leq \tau | Z = z] = \tau \tag{1}$$

which is in the nature of a local (to τ) independence or exclusion restriction. The value, τ, on the right hand side of (1) is a normalization. The function h is restricted to be weakly monotonic (normalized non-decreasing) in its final argument, U.

This paper considers identification of the function $h(x, \tau)$. If h were strictly increasing in U then Y would be continuously distributed and the model is the basis for the identifying models developed in Chernozhukov and Hansen (2005) and Chernozhukov, Imbens and Newey (2007). It is helpful to briefly consider this continuous outcome case and this is the subject of Section 2.

Since a discrete outcome can be very close to continuous if it has many densely packed points of support it seems plausible that there is an identification result for the discrete outcome case. The contribution of this paper is the development of identification results for this case, a case excluded from consideration in the papers just cited.

The identifying power of the model when Y is discrete is the subject of Section 3. Under weak nonparametric restrictions there is only *partial* identification of the structural function h when the outcome it delivers is discrete. As points of support of Y become more dense the sets within which a structural function is identified shrink, approaching point identification results for the continuous Y case under suitable conditions.

The key to analysis of the continuous outcome case is, as shown in Chernozhukov and Hansen (2005), the following condition implied by the model set out above.

$$\text{for all } z: \quad \Pr[Y \leq h(X, \tau) | Z = z] = \tau$$

Under some additional conditions this leads to point identification of the function $h(\cdot, \tau)$.

It is shown in Section 3 that when Y is *discrete* the model implies that $h(\cdot, \tau)$ simultaneously satisfies *two* sets of *inequalities*, as follows.[1]

$$\text{for all } z: \quad \Pr[Y \leq h(X, \tau) | Z = z] \geq \tau$$

$$\text{for all } z: \quad \Pr[Y < h(X, \tau) | Z = z] < \tau$$

This leads to set identification of the structural function $h(\cdot, \tau)$ and can place tight bounds on admissible structural functions when Y has many densely packed points of support.

Succinct general characterization of identifying sets developed in this paper seems difficult so the results are illustrated *via* examples in Section 4. Ordered probit and covariate dependent Poisson and binomial and binary logit models with endogeneity are studied in the examples.

[1] The 2001 working paper version of Chernozhukov and Hansen (2005) gives a derivation of these inequalities. The paper does not consider their role in the partial identifcation of structural functions.

The results of this paper shed light on the impact of endogeneity in situations where outcomes are by their nature discrete, for example where they are counts of events. Classical instrumental variables attacks fail because the restrictions of the IV model do not lead to point identification when outcomes are discrete. There are many econometric applications of models for discrete outcomes - see for example the compendious survey in Cameron and Trivedi (1998) - but there is little attention to endogeneity issues except in fully parametric specifications. There are a few papers which take an instrumental variables approach to endogeneity in parametric count data models basing identification on moment conditions - see the discussion in Section 11.3.2 of Cameron and Trivedi (1988), Mullahy (1997) and Windmeijer and Santos Silva (1997). These GMM based approaches do not take full account of the discrete nature of the outcome.

Progress can be made using a control function approach but this requires stronger restrictions and has the drawback that there is not point identification when endogenous *arguments* of structural functions are discrete. Chesher (2003) and Imbens and Newey (2003) study control function approaches to identification in non-additive error models with a discrete or continuous outcome and continuous endogenous arguments in the structural function. Chesher (2005) obtains set identification results in the case in which the endogenous arguments are discrete.

Roehrig (1988), Benkard and Berry (2006) and Matzkin (2005) study nonparametric identification in non-additive error simultaneous equation models without control function type restrictions but only for cases in which the outcomes are continuous. This paper studies the discrete outcome case in a "single equation" setting.

The results of the paper are informative about the effect of interval censoring and grouping on the identifying power of models.[2] The examples in Section 4 are striking in this regard. Quite small amounts of discretization due to interval censoring can result in significant degradation in the identifying power of models. This is useful information for designers of survey instruments who have control over the amount of interval censoring banded responses induce.

2. Continuous outcomes

Consider the model comprising the following restrictions which lie at the core of the models for continuous outcomes Y considered by Chernozhukov and Hansen (2005) and Chernozhukov, Imbens and Newey (2007).

C1. $Y = h(X, U)$ with U continuously distributed and h strictly monotonic (normalized increasing) in its last argument.

C2. For some $\tau \in (0, 1)$ there exists Z such that for all z $\Pr[U \leq \tau | Z = z] = \tau$.

[2] Manski and Tamer (2002) develop partial identification results for regression functions when there is interval censoring. This paper differs in that it is concerned with identification of discretely varying structural functions rather than the continuously varying counterparts associated with uncensored latent variates and it focuses on identification in the presence of endogeneity.

Let $\mathcal{C}$ denote the model comprising restrictions C1 and C2. The identifying power of this model is now briefly reviewed. To this end define $a(\tau, x, z)$ as follows:

$$a(\tau, x, z) \equiv \Pr[U \leq \tau | X = x, Z = z] \tag{2}$$

and note that by virtue of Restriction C1 there is, for all z:

$$\int a(\tau, x, z) f_{X|Z}(x|z) dx = \tau \tag{3}$$

where $f_{X|Z}$ is the conditional density of X given Z and integration is definite over the support of X which may vary with the value of Z.[3]

Since h is strictly increasing in U, applying the function $h(x, \cdot)$ to both sides of the inequality in (2) gives

$$\Pr[h(X, U) \leq h(X, \tau) | X = x, Z = z] = a(\tau, x, z)$$

and because $Y = h(X, U)$ there is

$$\Pr[Y \leq h(X, \tau) | X = x, Z = z] = a(\tau, x, z)$$

and finally by virtue of (3), for all z, on taking expectations with respect to X given $Z = z$:

$$\Pr[Y \leq h(X, \tau) | Z = z] = \tau. \tag{4}$$

This argument fails at the first step if h is not strictly increasing in U.

Without further restriction there are many functions satisfying (4). Certain additional restrictions result in a model that identifies the function h. In the absence of parametric restrictions these include a requirement that the support of Z be at least as rich as the support of X and that the distribution of Y and X conditional on Z has sufficient variation with Z. These necessary conditions are implied by a completeness condition for local (in the sense of Rothenberg (1970)) identification given in Chernozhukov, Imbens and Newey (2007). When these conditions do not hold in full there can be informative partial identification of the function $h(\cdot, \tau)$ in the sense that the condition (4) along with other maintained conditions limit $h(\cdot,)$ to some class of functions. The results of this paper for the discrete outcome case are of precisely this nature.

When $Z = X$, that is when Restriction C2 above holds with "$Z = z$" replaced by "$X = x$" (there is local-to-τ independence of U and X) there is for all x

$$\Pr[Y \leq h(x, \tau) | X = x] = \tau \tag{5}$$

and $h(x, \tau)$ is the τ-quantile regression function of Y given $X = x$ and h is therefore identified (Matzkin (2003)). When Y is discrete the conditional τ-quantile function of Y given $X = x$ also identifies $h(x, \tau)$.[4] However when Y is discrete and $Z \neq X$ neither equation (4) nor equation (5) hold and the identifiability of h under a weak restriction like C2 is unclear. The discrete Y case is explored now.

---

[3] And similarly if X is discrete.

[4] In this discrete Y case h is normalized to be caglad for variation in U.

3. Discrete outcomes

3.1. A model for discrete outcomes. The model $\mathcal{C}$ is now amended to permit Y to be discrete. There is the following model: $\mathcal{D}$.

D1. $Y = h(X, U)$ with U continuously distributed and h is weakly monotonic (normalized caglad, non-decreasing) in its last argument with as codomain the ascending sequence $\{y_m\}_{m=1}^{M}$ which is independent of X. M may be unbounded.

D2. For some $\tau \in (0, 1)$ there exists Z such that for all z: $\Pr[U \leq \tau | Z = z] = \tau$.

It is now shown that under the restrictions of the model $\mathcal{D}$ the structural function $h(\cdot, \tau)$ satisfies two inequalities as follows.

$$\text{for all } z: \quad \Pr[Y \leq h(X, \tau) | Z = z] \geq \tau \tag{6}$$

$$\text{for all } z: \quad \Pr[Y < h(X, \tau) | Z = z] < \tau \tag{7}$$

First note that under the Restriction D1 the function $h(x, u)$ can be characterized by functions $\{p_m(x)\}_{m=0}^{M}$ as follows:

$$\text{for } m \in \{1, \ldots, M\}: \qquad h(x, u) = y_m \text{ if } p_{m-1}(x) < u \leq p_m(x)$$

with, for all x, $p_0(x) \equiv 0$ and $p_M(x) \equiv 1$.

Now consider the inequality (6). Conditioning on $X = x$ and $Z = z$ there is, for all x and z the inequality

$$\Pr[Y \leq h(X, \tau) | X = x, Z = z] \geq a(\tau, x, z) \tag{8}$$

and on taking expectations with respect to X given $Z = z$ on the left and right hand sides of (8) and using (3) there is, for all z the inequality (6).

The argument leading to the inequality (8) takes the following steps. $1[\cdot]$ denotes the indicator function equal to 1 if its argument is true and 0 otherwise.

$$\begin{aligned}
\Pr[Y \leq h(X, \tau) | X = x, Z = z] &= \Pr[h(X, U) \leq h(X, \tau) | X = x, Z = z] \\
&= \sum_{m=1}^{M} 1[h(x, \tau) = y_m] \Pr[h(X, U) \leq h(X, p_m(x)) | X = x, Z = z] \\
&= \sum_{m=1}^{M} 1[h(x, \tau) = y_m] \Pr[U \leq p_m(x) | X = x, Z = z] \\
&= \sum_{m=1}^{M} 1[h(x, \tau) = y_m] a(p_m(x), x, z) \\
&\geq a(\tau, x, z)
\end{aligned}$$

The second and final lines follow because $h(x, \tau) = y_m$ if and only if $\tau \in (p_{m-1}(x), p_m(x)]$.

Now consider the inequality (7). Conditioning on $X = x$ and $Z = z$ there is, for all x and z, the inequality

$$\Pr[Y < h(X, \tau) | X = x, Z = z] < a(\tau, x, z) \tag{9}$$

and on taking expectations with respect to X given $Z = z$ on the left and right hand sides of (9) and using (3) there is, for all z the inequality (7). The argument leading to the inequality (9) takes the following steps.

$$
\begin{aligned}
\Pr[Y < h(X,\tau)|X = x, Z = z] &= \sum_{m=1}^{M} 1[h(x,\tau) = y_m] \Pr[h(X,U) < h(x,p_m(x))|X = x, Z = z] \\
&= \sum_{m=2}^{M} 1[h(x,\tau) = y_m] \Pr[h(X,U) \leq h(x,p_{m-1}(x))|X = x, Z = z] \\
&= \sum_{m=2}^{M} 1[h(x,\tau) = y_m] \Pr[U \leq p_{m-1}(x)|X = x, Z = z] \\
&= \sum_{m=2}^{M} 1[h(x,\tau) = y_m] a(p_{m-1}(x), x, z) \\
&< a(\tau, x, z)
\end{aligned}
$$

3.2. Set identification. By construction, the function, say $h_0(x,u)$, which, together with the distribution of (U,X) given Z, say $F^0_{UX|Z}$, generates the distribution of (Y,X) given Z defining the probabilities in (6) and (7) satisfies the inequalities (6) and (7) when $u = \tau$.

If there are other functions which, for a value of τ of interest, satisfy (6) and (7) for all z in the support of Z then the model $\mathcal{D}$ may not *point identify* $h_0(\cdot,\tau)$. The discreteness of the outcome Y leads one to expect that such functions exist. Because Y is discrete it is not possible to identify $F^0_{U|XZ}$. Because Y is discrete there may be a structural function h_1 and a distribution $F^1_{UX|Z}$ which do not violate the restrictions of the model or the bounding inequalities and which combine to give the distribution, $F^0_{YX|Z}$ say, that is produced by h_0 and $F^0_{UX|Z}$.

A distribution of (Y,X) given Z, $F^0_{YX|Z}$, the support of Z and the inequalities (6) and (7) calculated using $F^0_{YX|Z}$ in general define a *set* of functions which the inequalities (6) and (7) implied by the model $\mathcal{D}$ does not rule inadmissible as generators of $F^0_{YX|Z}$. In this sense the model $\mathcal{D}$ can have *set identifying* power for the structural function $h(\cdot,\tau)$. There seems to be no additional information conveyed in the model $\mathcal{D}$ which can reduce the extent of this set. Section 4 gives examples of particular classes of structure for which the model has set identifying power.

Faced with a particular distribution $F^0_{YX|Z}$ and support of Z one can enumerate or otherwise characterize the set of potential structural functions that satisfy the inequalities (6) and (7). One can examine how additional restrictions can reduce the extent of the identification set implied by the model.

Armed with an estimate $\hat{F}_{YX|Z}$ and perhaps some additional restrictions, one can attempt various sorts of inference about the structural function, for example testing hypotheses about particular aspects of the data generating structural function. There are challenging problems here.

The distance between the functions on the left hand sides of (6) and (7) is $\Delta_\tau(z)$, the probability that Y falls *on* the τ- structural function conditional on $Z = z$.

$$\begin{aligned}\Delta_\tau(z) &\equiv \Pr[Y = h(X,\tau)|Z = z] \\ &= \int (a(p_m(x), x, z) - a(p_{m-1}(x), x, z))\, f_{X|Z}(x|z)dx\end{aligned}$$

As the support of Y grows more dense then, if a continuous limit is approached, the maximal probability mass (conditional on X and Z) on any point of support of Y, and so $\Delta_\tau(z)$, will converge to zero and the upper and lower bounds will come to coincide.

Even when the bounds coincide there can remain a set of structural functions admitted by the model. This is always the case when Z has no variation at all and more generally when the support of Z is less rich than the support of X.

The next Section illustrates using as vehicles: an ordered probit model, a Poisson regression model and a binomial regression model, in each case with an endogenous explanatory variable.

4. ILLUSTRATIONS

Particular structures obtained from a generic model, set out below, are considered. For particular completely specified structures admitted by this model the nature of the set of structural functions admitted by the inequalities (6) and (7) is explored. This is done by considering structural functions which are candidates for inclusion in the identification set and admitting those which do not lead to violations of the inequalities (6) and (7) for any value of the instrument z. The effects on the identified set of structural functions of varying amounts of discreteness and of the strength and support of instruments are examined.

4.1. A generic model.

Here is the model used to generate the structures used in the illustrative examples. Φ^{-1} is the standard normal quantile function.

$$Y = h(X,U) \qquad U \in (0,1) \qquad W \equiv \Phi^{-1}(U)$$

$$\text{for } m \in \{1,\ldots,M\}: \qquad h(x,u) = y_m \text{ if } p_{m-1}(x) < u \leq p_m(x) \tag{10}$$

$$p_0(x) \equiv 0 \qquad p_M(x) \equiv 1$$

$$X = \beta_0 + \beta_1 Z + V$$

$$\left[\begin{array}{c} W \\ V \end{array}\right] |Z = z \sim N\left(\left[\begin{array}{c} 0 \\ 0 \end{array}\right], \left[\begin{array}{cc} 1 & \sigma_{wv} \\ \sigma_{wv} & \sigma_{vv} \end{array}\right]\right) \tag{11}$$

$\{y_m\}_{m=1}^M$ is an ascending sequence.

The model embodies the restrictions of the model $\mathcal{D}$, in particular U and Z are independently distributed. The model additionally specifies a particular type of dependence[5] amongst U, X and Z using a simple joint Gaussian distribution for a transformation of U and $V = X - \beta_0 - \beta_1 Z$.

[5]This is of course essential because the model will be used to produce *particular* structures in the context of which the identifying power of variants of the model $\mathcal{D}$ will be studied.

The structural function of this model at $u = \tau$ is $h(x, \tau)$, obtained by setting $u = \tau$ in (10). In all the cases considered here each function $p_m(x)$ is a strictly monotonic decreasing function of scalar x and so $h(x, \tau)$ can be expressed as

$$\text{for } m \in \{1, \ldots, M\}: \qquad h(x, \tau) = y_m \text{ if } s_{m-1}(\tau) < x \leq s_m(\tau) \tag{12}$$

where the functions $s_m(x)$ are inverse functions such that for all $m \in \{1, \ldots, M-1\}$, $s_m(p_m(x)) = x$ for all x, and for all τ, $s_0(\tau) \equiv -\infty$, $s_M(\tau) \equiv \infty$.

Define the conditional variance: $\sigma^2_{w|v} \equiv 1 - \sigma^2_{wv}/\sigma_{vv}$. The distributions of X given Z and of Y given X and Z are as follows.

$$X|Z = z \sim N\left(\beta_0 + \beta_1 z, \sigma_{vv}\right) \tag{13}$$

$$\Pr[Y \leq y_m | X = x, Z = z] = \Phi\left(\frac{1}{\sigma_{w|v}}\left(\Phi^{-1}\left(p_m(x)\right) - \frac{\sigma_{wv}}{\sigma_{vv}}\left(x - \beta_0 - \beta_1 z\right)\right)\right) \tag{14}$$

The latent variable U is marginally distributed $Unif(0,1)$. It is distributed independently of X if and only if σ_{wv} is zero in which case Y is independent of Z given X and there is the following simplification.

$$\Pr[Y \leq y_m | X = x] = p_m(x)$$

Cases in which $\{p_m(x)\}_{m=1}^{M}$ are cumulative probabilities for covariate dependent ordered probit (including binary probit), Poisson and binomial (including binary logit) distributions are considered here.

In this triangular model (U, V) and Z are independently distributed. That restriction could be relaxed while retaining the Gaussian structure and the independence of U and Z and of V and Z by allowing σ_{wv} to vary with z in the conditional distribution (11).[6]

In the examples considered here σ_{wv} is constant for variations in z and the model is in the class of nonadditive error, *triangular* models considered in Chesher (2003) and Imbens and Newey (2003) for which point identification is possible with discrete or continuous Y without imposing the parametric restrictions of the model above, as long as X is continuous, as it is here. Inspecting (13) and (14) it is clear that with parametric restrictions a "control function" argument can be used to develop identification conditions.[7]

The examples set out below are built on particular structures that are admitted by this triangular model. However, in the analysis, the identifying power of particular cases of the *less restrictive* model $\mathcal{D}$, set out at the start of Section 3.1, is considered. The *only* independence restriction maintained in this model is that U and Z are independently distributed. The model whose identifying power is studied is *silent* about the nature of any dependencies between U and X and between X and Z.

---

[6] I am grateful to Lars Nesheim for this insight. Models with σ_{wv} constant or varying with z are all in the class admitted by the model $\mathcal{D}$ of Section 3.1.

[7] Smith and Blundell (1986) use this sort of argument in a parametric Gaussian simultaneous equation tobit model. See Newey, Powell and Vella (1999) for a control function approach in non-parametric additive error models.

The candidate structural functions considered here are restricted to be weakly monotonic and nondecreasing in x, $h^p(x,\tau)$, defined as follows.

$$\text{for } m \in \{1,\ldots,M\}: \qquad h^p(x,\tau) = y_m \text{ if } s^p_{m-1}(\tau) < x \leq s^p_m(\tau)$$

The structures used in the examples are all members of familiar parametric families. The candidate structural functions are generated by varying the values of these parameters which in the cases considered guarantees weak monotonicity of the cumulative probabilities $p_m(x)$ for variations in x.

For each value, z, of the instrument considered, the probabilities in the inequalities (6) and (7) for a particular candidate structural function $h^p(x,\tau)$ are calculated for the particular structure of interest (which determines particular distributions of X given Z and of Y given X and Z) as follows.[8]

$$\Pr[Y \leq h^p(X,\tau)|Z=z] = \sum_{m=1}^{M} \int_{s^p_{m-1}(\tau)}^{s^p_m(\tau)} \Phi\left(\frac{1}{\sigma_{w|v}}\left(\Phi^{-1}(p_m(x)) - \frac{\sigma_{wv}}{\sigma_{vv}}(x-\beta_0-\beta_1 z)\right)\right)\frac{1}{\sigma_{vv}^{1/2}}\phi\left(\frac{1}{\sigma_{vv}^{1/2}}(x-\beta_0-\beta_1 z)\right)dx$$

$$\Pr[Y < h^p(X,\tau)|Z=z] = \sum_{m=2}^{M} \int_{s^p_{m-1}(\tau)}^{s^p_m(\tau)} \Phi\left(\frac{1}{\sigma_{w|v}}\left(\Phi^{-1}(p_{m-1}(x)) - \frac{\sigma_{wv}}{\sigma_{vv}}(x-\beta_0-\beta_1 z)\right)\right)\frac{1}{\sigma_{vv}^{1/2}}\phi\left(\frac{1}{\sigma_{vv}^{1/2}}(x-\beta_0-\beta_1 z)\right)dx$$

In the integrands of the second expression, $p_{m-1}(x)$ appears because the calculation is of the probability that Y falls on points of support lying immediately below and *not on* the structural function. There is no integration over the lowest interval because the probability Y falls below its lowest point of support is zero.

The value of z is varied and it is determined whether

$$\Pr[Y \leq h^p(X,\tau)|Z=z] \geq \tau$$

and

$$\Pr[Y < h^p(X,\tau)|Z=z] < \tau$$

hold for all z in the support of Z. The candidate structural functions which, for the structure chosen do not violate these inequalities for any z in the support of Z constitute the set of structural functions identified by the model under consideration for the particular structure chosen.

[8] Numerical integration is done using the function `integrate` provided in R (Ihaka and Gentleman (1996)).

Table 1: Parameter values for the ordered probit model

Parameter	α_0	α_1	α_2	β_0	β_1	σ_{wv}	σ_{vv}
Value	0	1	1	0	2	0.6	1

4.2. Ordered probit.

Definitions. In the ordered probit version of the model of Section 4.1 the cumulative probabilities (with exogenous X) are:

$$p_m(x) = \Phi\left(\frac{1}{\alpha_2}(T_m - \alpha_0 - \alpha_1 x)\right), \; m \in \{1, \dots, M\} \tag{15}$$

where Φ is the standard normal distribution function, $\alpha_2 > 0$, and $\{T_m\}_{m=1}^{M}$ are constants with $T_M \equiv \infty$. These probabilities can arise by interval censoring of a latent Y^*

$$Y^* = \alpha_0 + \alpha_1 X + \alpha_2 U$$

with U independent of X and distributed $N(0,1)$ and for $m \in \{1, \dots, M\}$:

$$Y = y_m \text{ for } T_{m-1} < Y^* \leq T_m$$

with $T_0 \equiv -\infty$.

The functions $\{s_m(\tau)\}_{m=1}^{M-1}$ defining the structural function $h(x, \tau)$ are as follows in the case in which $\alpha_1 > 0$.

$$s_m(\tau) = \frac{1}{\alpha_1}\left(T_m - \alpha_0 - \alpha_2 \Phi^{-1}(\tau)\right), \; m \in \{1, \dots, M-1\} \tag{16}$$

Particular structures are obtained by choosing particular numerical values for the parameters:

$$\theta \equiv \{M, \alpha_0, \alpha_1, \alpha_2, \beta_{0,}\beta_1, \sigma_{wv}, \sigma_{vv}\}$$

and the identifying power of the following model $\mathcal{OP}$ is studied.

OP1. $Y = h(X, U)$ with, for $m \in \{1, \dots, M\}$:

$$h(x, u) = y_m \text{ for } p_{m-1}(x) < u \leq p_m(x)$$

with $p_m(x)$ defined as in (15), $a_2 > 0$ and $\{T_m\}_{m=0}^{M}$ known constants, $T_0 \equiv -\infty$, $T_M \equiv \infty$.

OP2. U distributed independently of Z, $Unif(0,1)$.

This model embodies the restrictions of the model $\mathcal{D}$ and some additional parametric restrictions.

Bounding inequalities. Figure 1 shows the bounding conditional probabilities calculated using the parameter values shown in Table 1 *and* using the structural function $h(x, \tau)$ *implied by those parameter values.* The four panes show the variation in the probabilities as the value of the instrument z varies. The value of τ is 0.5 in the left hand panes and 0.8 in the right hand panes. Because the structural function employed in the calculations is the structural function for the specified value of τ implied by the parameter values in Table 1 (which determine the probability distribution used in calculating the probabilities) the bounding probabilities always fall either side of the specified value of τ.

In the upper panes Y has $M = 10$ points of support and in the lower panes $M = 30$ (left) and $M = 100$ (right). The values of $T_1, \ldots, T_{M-1}$ are equally spaced in $[-10, 10]$. Increasing the number of points of support has a dramatic effect, bringing the bounds tightly onto the chosen value of τ.

The effect of the value of z on the probabilities in (6) and (7) is small in Figure 1. This occurs because of the large range, $[-10, 10]$, over which the thresholds, T_m, are spaced relative to the range of variation used for z, $[-2, 2]$. These choices were made to facilitate demonstration of the impact of reduction in discretization on the bounding probabilities. It can be shown that in this endogenous ordered probit model

$$\lim_{z\to\infty} \Pr[Y \leq h(X, 0.5)|Z = z] = \lim_{z\to-\infty} \Pr[Y \leq h(X, 0.5)|Z = z] = 1$$

$$\lim_{z\to\infty} \Pr[Y < h(X, 0.5)|Z = z] = \lim_{z\to-\infty} \Pr[Y < h(X, 0.5)|Z = z] = 0$$

so in all these cases for values of z sufficiently large in magnitude the upper and lower bounding probabilities become close to respectively 1 and 0.

Candidate structural functions. Now candidate structural functions that differ from the structural function associated with the distribution of (Y, X) given Z employed in the probability calculations are considered. For each case the bounding probabilities (6) and (7) are calculated as functions of the instrument z. For this exercise the degree of discretization is chosen to be quite substantial with just 4 points of support and 3 finite thresholds $\{-2, 0, +2\}$.

In the first set of cases the structure generating the joint distribution of (Y, X) given Z has the parameter values shown in Table 1. The values of X at which the $\tau = 0.5$ structural function changes are shown in the first row of Table 2.[9]

Alternative candidate structural functions are obtained by using the expression (16) with $\alpha_0 = 0$, $\alpha_2 = 1$ and $\tau = 0.5$ and with α_1 varying in $\{1.0, 1.4, 1.8, 2.2\}$. The values of X at which these structural functions change (that is the values of $s_m^p(0.5)$) are shown in Table 2.

[9] To be clear, the $\tau = 0.5$ structural function is as follows.

$$\begin{array}{ll} h(x, 0.5) = y_1 & -\infty < x \leq -2 \\ h(x, 0.5) = y_2 & -2 < x \leq 0 \\ h(x, 0.5) = y_3 & 0 < x \leq +2 \\ h(x, 0.5) = y_4 & +2 < x \leq +\infty \end{array}$$

Table 2: Threshold values of putative structural functions in the ordered probit model

	x_1	x_2	x_3
$\alpha_1 = 1.0$	-2.00	0	2.00
$\alpha_1 = 1.4$	-1.43	0	1.43
$\alpha_1 = 1.8$	-1.11	0	1.11
$a_1 = 2.2$	-0.91	0	0.91

The results of the probability calculations are shown in Figure 3. When $a_1 = 1.0$ this candidate structural function *is* the structural function for the structure defined in Table 1. As a_1 moves away from 1.0 the upper and lower conditional probabilities move towards the line $\tau = 0.5$ but never cross it. None of these structural functions is identifiably distinct from the structural function, say $h_0(\cdot, 0.5)$, corresponding to the parameter values in Table 1 with $\tau = 0.5$. However, considering pane (d), the candidate structural function for this case ($a_1 = 2.2$) is identifiably distinct from $h_0(\cdot, \tau)$ for τ deviating by quite small amounts from 0.5.

Now consider the impact of increasing the strength of the instrument. Note that varying the value of β_1 has no effect in this regard unless the support of Z is held fixed over some finite interval because changing the value of β_1 just has the effect of redenominating the unit of measurement of Z. The strength of the instrument *is* increased by reducing σ_{vv} which reduces the noise in the relationship between X and Z. The results in Figure 4 are obtained with σ_{vv} reduced from its Table 1 value of 1.0 to 0.6, which corresponds to an increase in the R^2 for the linear regression of X on Z from 0.36 to 0.6. Now the structural function obtained with $\alpha_1 = 2.2$ can be ruled inadmissible as a $\tau = 0.5$ structural quantile function and there is no value of τ which for which this function is admissible as a structural function for the joint distribution of (Y, X) given Z employed in these calculations.

Finally consider the effect of increasing the number of points of support of Y. The results in Figure 5 are obtained using 7 points of support with finite thresholds at the following values:

$$\{-2.0, -1.2, -0.4, 0.0, +0.4, +1.2, +2.0\}$$

and with σ_{vv} reset to its Table 1 value of 1.0 as in Figure 3. With the finer discretization the structural functions obtained with $\alpha_1 = 1.8$ and $\alpha_1 = 2.2$ can be both ruled inadmissible as $\tau = 0.5$ structural quantile functions, indeed as structural functions for any value of τ.

Identification sets. Any particular structure, with structural function say h_0, implies a joint distribution of (Y, X) given Z, $F^0_{YX|Z}$. With knowledge of $F^0_{YX|Z}$ one can ask, what candidate structural functions, satisfying particular restrictions (e.g. monotonicity, or perhaps more specific parametric restrictions), are observationally indistinguishable from h_0 in the sense that the inequalities (6) and (7) are not violated for any value of the instrument z at some particular value of τ.

Consider the particular triangular ordered probit structure with the parameter values set out in Table 1 (except as stated below), the model $\mathcal{OP}$ and candidate structural functions $h^p(\cdot, 0.5)$ obtained by varying candidate values of α_0 and α_1 in (16).[10] Figure 6 shows combinations of values of α_0 and α_1 which generate structural functions for which the bounding probabilities (6) and (7) fall either side of $\tau = 0.5$ for all values of z are observationally indistinguishable from $\alpha_0 = 0$ and $\alpha_1 = 1$. Values within the outer region in Figure 6 are for the case in which $\sigma_{vv} = 1.0$ when the instrument Z is rather weak.[11] The inner region is the identification set when $\sigma_{vv} = 0.6$ so the instrument is stronger. Strengthening the instrument substantially reduces the extent of the identification set.

4.3. Poisson regression.

Definitions. In the Poisson version of the model of Section 4.1 the functions defining the structural function are:

$$p_m(x) = \exp(-\lambda(x)) \sum_{y=0}^{m-1} \frac{\lambda(x)^y}{y!} \qquad m \in \{1, 2, \dots\} \tag{17}$$

and the Poisson parameter $\lambda(x)$ is parameterized as follows.

$$\lambda(x) = \exp(\alpha_0 + \alpha_1 x) \tag{18}$$

The outcome Y is given by $Y = h(X, U)$ where

$$h(x, u) = m - 1 \text{ for } p_{m-1}(x) < u \leq p_m(x), \qquad m \in \{1, 2, \dots\} \tag{19}$$

with $p_0(x) \equiv 0$. When U and X are independently distributed and $U \sim Unif(0, 1)$ then Y has a conditional Poisson distribution given X with point probabilities:

$$\Pr[Y = y|X = x] = \exp(-\lambda(x)) \frac{\lambda(x)^y}{y!}, \qquad y \in \{0, 1, 2, \dots\}.$$

The rest of the model is as in Section 4.1 with $U = \Phi^{-1}(W)$ and $V = X - \beta_0 - \beta_1 Z$ jointly normally distributed independently of Z. Particular structures are obtained by choosing particular numerical values for the parameters:

$$\theta \equiv \{\alpha_0, \alpha_1, \beta_{0,}\beta_1, \sigma_{wv}, \sigma_{vv}\}$$

and the identifying power of the following model $\mathcal{P}$ is studied.

P1. $Y = h(X, U)$ with, the function h defined in (17), (18) and (19).

P2. U distributed independently of Z, $Unif(0, 1)$.

This model embodies the restrictions of the model $\mathcal{D}$ and some additional parametric restrictions.

Table 3: Parameter values for the Poisson model

Parameter	α_0	α_1	β_0	β_1	σ_{wv}	σ_{vv}	τ
Value	0	1	0	1.0	0.6	1	0.5

Bounding inequalities. Figure 7 shows the bounding conditional probabilities, (6) and (7), as functions of the value of the instrument z, calculated using the parameter values shown in Table 3 *and* using the structural function $h(x, \tau)$ *implied by those parameter values.*[12] The case with $\tau = 0.5$ is shown. The functions lie either side of $\tau = 0.5$, as they must, but as z becomes large they approach $\tau = 0.5$ very closely.

This happens because at large values of the instrument and at the parameter values considered here the endogenous X typically takes large values resulting in large values of the Poisson parameter $\lambda(x) = \exp(\alpha_0 + \alpha_1 x)$. When $\lambda(x)$ is large the variance of Y is large and probability mass is spread across many points of support. In this situation the impact of discreteness is very small. Conversely when z is small X tends to take small values and the Poisson parameter is close to zero. In that situation the distribution of Y is concentrated on just a few points of support close to zero and the impact of discreteness is very substantial.

Candidate structural functions. These effects are evident when candidate structural functions are considered. Figure 8 show the bounding probabilities, (6) and (7), calculated using the parameter values shown in Table 3 as functions of the value of the instrument: in pane (a) for $\alpha_1 = 1.0$ (reproducing Figure 7 but over a smaller range of values of z), and in panes (b) - (d) for $\alpha_1 \in \{0.95, 1.05, 1.10\}$. At large values of z even these small deviations from the value of α_1 (1.0) which determines the probabilities used in the calculation lead to violation of the inequality constraints.

The IV model has almost point identifying power when the dependence of X on the instrument and the support of the instrument are such that large values of the Poisson "parameter" are achieved but not otherwise. If a few small values of Y are all that is typically observed, perhaps because the instrument has limited support, then the IV model may have little identifying power.

4.4. Binomial regression.

Definitions. In the binomial version of the model of Section 4.1 the functions defining the structural function are:

$$p_m(x) = \sum_{y=0}^{m-1} \binom{N}{y} \gamma(x)^y \left(1 - \gamma(x)\right)^{N-y} \qquad m \in \{1, 2, \ldots, M\} \tag{20}$$

[10] The value of α_2 is irrelevant because $\tau = 0.5$ and $\Phi^{-1}(0.5) = 0$.

[11] The slight irregularities along the edges of the identification sets are artefacts of the plotting procedure which utilizes values of a logical variable on a 100×100 grid.

[12] Numerical integration is done using the integrate function in R (Ihaka and Gentleman (1996)). Values of X at which structural functions break are calculated using the uniroot and ppois functions in R. Summations are truncated when the proportionate change in the total on adding an additional segment is smaller than 10^{-6}.

Table 4: Parameter values for the binomial model

Parameter	α_0	α_1	β_0	β_1	σ_{wv}	σ_{vv}	τ
Value	0	1	0	1.0	0.6	1	0.5

with $M = N + 1$ and the binomial parameter $\gamma(x)$ is parameterized as follows.

$$\gamma(x) \equiv \frac{\exp(\alpha_0 + \alpha_1 x)}{1 + \exp(\alpha_0 + \alpha_1 x)} \tag{21}$$

The outcome Y is given by $Y = h(X, U)$ where

$$h(x, u) = m - 1 \text{ for } p_{m-1}(x) < u \leq p_m(x), \qquad m \in \{1, 2, \ldots, M\} \tag{22}$$

with $p_0(x) \equiv 0$. When U and X are independently distributed and $U \sim Unif(0, 1)$ then Y has a conditional binomial distribution given X with point probabilities:

$$\Pr[Y = y|X = x] = \binom{N}{y} \gamma(x)^y \left(1 - \gamma(x)\right)^{N-y}, \qquad y \in \{0, 1, 2, \ldots, N\}.$$

When $N = 1$ this is a standard parametric binary logit model.

The rest of the model is as in Section 4.1 with $U = \Phi^{-1}(W)$ and $V = X - \beta_0 - \beta_1 Z$ jointly normally distributed independently of Z. Particular structures are obtained by choosing particular numerical values for the parameters:

$$\theta \equiv \{N, \alpha_0, \alpha_1, \beta_{0,}\beta_1, \sigma_{wv}, \sigma_{vv}\}$$

and the identifying power of the following model $\mathcal{B}$ is studied.

B1. $Y = h(X, U)$ with, the function h defined in (20), (21) and (22).

B2. U distributed independently of Z, $Unif(0, 1)$.

This model embodies the restrictions of the model $\mathcal{D}$ and some additional parametric restrictions.

Bounding inequalities. Figure 9 shows the bounding conditional probabilities, (6) and (7), as functions of the value of the instrument z, calculated using the parameter values shown in Table 4 *and* using the structural function $h(x, \tau)$ *implied by those parameter values.*[13] The case with $\tau = 0.5$ is shown. The four panes show the bounding probabilities for $N \in \{1, 3, 6, 10\}$. The functions lie either side of $\tau = 0.5$, as they must and as the number of points of support increases the bounds come to sit tighter on $\tau = 0.5$.

[13] Calculations are done using the integrate, uniroot and pbinom functions of R 2.3.1.

Candidate structural functions. In Figure 10 the parameter values in Table 4 still define the probability distribution used to calculate the bounding conditional probabilities but the structural function is what is obtained in this binomial model with $\alpha_0 = 0.5$ rather than the value 0.0 in Table 4.[14] The effect is to shift both bounding probabilities upwards and for $N = 6$ and $N = 10$ the $\tau = 0.5$ line is crossed indicating that in these cases this candidate structural function is identifiably distinct from the Table 4 $\tau = 0.5$ structural function.

A similar exercise is reported in Figure 11, this time using a candidate structural function with $\alpha_0 = 0$ but $\alpha_1 = 0.5$ rather than the Table 4 value of 1.0. Again for N sufficiently large the candidate $\tau = 0.5$ structural function is identifiably distinct from that defined by the parameter values in Table 4.

The binary case. When $N = 1$ the outcome Y is binary and when X is exogenous ($\sigma_{wv} = 0$) it has a binary logit distribution conditional on X.

It is evident from Figures 9, 10 and 11 that at the parameter values considered so far the bounding probabilities are far apart in this binary case so that there is a large class of candidate structural functions admitted by the model. The key to reducing the size of the identified set lies in strengthening the instrument by reducing the value of σ_{vv}.

Figure 12 shows the effect of varying σ_{vv} in the binary case. In every pane except pane (d) the bounding probabilities are for a $\tau = 0.5$ structural function calculated using the structural function implied by the parameter values in Table 4 except for the value of σ_{vv} which is varied and the value of σ_{wv} which is adjusted to keep the correlation between W and V constant.

Pane (a) reproduces the case with $\sigma_{vv} = 1$ and $\sigma_{wv} = 0.6$ already displayed in the same pane in Figure 9 but with the difference that the instrument is adjusted over the narrower range $[-2, 2]$. Pane (b) shows the situation with $\sigma_{vv} = 0.1$. The bounds come more tightly onto $\tau = 0.5$ and much more so in pane (c) in which $\sigma_{vv} = 0.01$. The upper bounding probability is minimised at a different value of z than the lower bounding probability is maximised. This has to occur because in this binary case the probability that Y lies on the $\tau = 0.5$ structural function conditional on $Z = z$ (which is the distance between the bounding probability functions at $Z = z$) cannot fall below 0.5.

In pane (d) the bounding probabilities are calculated for the case $\sigma_{vv} = 0.1$ using

[14] Here are the expressions for the bounding probabilities for a candidate $\tau = 0.5$ stuctural function

$$p(x) \equiv \frac{\exp(a_0 + a_1 x)}{1 + \exp(a_0 + a_1 x)}$$

when $N = 1$ and parameter values are as shown in Table 4.

$$\Pr[Y \leq h^p(X,\tau)|Z = z] = \int_{-\infty}^{-a_0/a_1} \Phi\left(\frac{1}{0.8}\left(\Phi^{-1}(\frac{1}{1+\exp(x)}) - 0.36\,(x - 2z)\right)\right)\phi\left((x-2z)\right)dx + \int_{-a_0/a_1}^{\infty}\phi\left((x-2z)\right)dx$$

$$\Pr[Y < h^p(X,\tau)|Z = z] = \int_{-a_0/a_1}^{\infty} \Phi\left(\frac{1}{0.8}\left(\Phi^{-1}(\frac{1}{1+\exp(x)}) - .36\,(x - 2z)\right)\right)\phi\left((x-2z)\right)dx$$

Note that when a_0 is equal to 0 these expressions are not sensitive to variations in a_1. This feature is peculiar to the $\tau = 0.5$ case with $N = 1$.

the structural function implied by a binary logit model with $\alpha_0 = 1$ and $\alpha_1 = 1$ rather than the values respectively 0 and 1 that define the probabilities used in the calculation. With the strong instrument implied by $\sigma_{vv} = 0.1$ this structural function can be ruled out as inadmissible. The dashed lines in pane (d) show the bounding probabilities for this case when $\sigma_{vv} = 1$. With the weaker instrument the structural function for the case with $\alpha_0 = 1$ and $\alpha_1 = 1$ cannot be ruled inadmissible given the information conveyed in the distribution of Y and X given Z generated with $\alpha_0 = 0$ and $\alpha_1 = 1$.

Identification sets. Consider the particular binomial structure with the parameter values set out in Table 4, and possible observationally indistinguishable alternative structural functions $h^p(\cdot, 0.5)$ under the model $\mathcal{B}$ obtained by varying candidate values of α_0 and α_1. Figure 13 shows combinations of values of candidate values of α_0 and α_1 which are observationally indistinguishable from $\alpha_0 = 0$ and $\alpha_1 = 1$.[15] Values within the outer region are observationally indistinguishable when $N = 4$. Values within the inner regions are observationally indistinguishable when $N = 6$ and $N = 12$. Reductions in discreteness substantially reduce the extent of the identification set.

5. Concluding remarks

1. The simple examples of the previous Section are useful for exploring the implications of the results of this paper. However they have the disadvantage that because their genesis is in simple parametric models they distract from the fact that the results of the paper apply in the absence of parametric restrictions.

2. Several simplifying assumptions could be relaxed. For example extension to many endogenous variables and introduction of exogenous covariates in the structural function is straightforward.

3. When Y is discrete the model $Y = h(X, U)$ with h non-decreasing in U and for some τ and all z $\Pr[U \leq \tau | Z = z] = \tau$, may only set identify the function $h(\cdot, \tau)$. The following pair of inequalities which simultaneously hold for all z define the class of structural functions that are concordant with a joint distribution of (Y, X) given Z.

$$\begin{aligned} \Pr[Y &\leq h(X, \tau) | Z = z] \geq \tau \\ \Pr[Y &< h(X, \tau) | Z = z] < \tau \end{aligned}$$

   Given such a joint distribution and information about the support of Z it may be feasible to enumerate or otherwise characterize this class of functions. An estimate of the joint distribution of (Y, X) given Z can be used to produce an analogue estimator of this class of functions. Informative empirical analysis will require further restrictions, for example "smoothness" restrictions limiting the number of turning points in the structural function as X varies. In

[15] The irregularities along the edges of the identification sets are artefacts of the plotting procedure which utilizes values of a logical variable on a 100×100 grid.

the examples studied here a monotonicity restriction has been imposed. Further analysis of these characterization problems and associated estimation and inference questions is left for future research.

4. The calculations presented here suggest that when discreteness of outcomes is a significant feature the identifying power of IV models can be low unless instruments are very strong, that is very highly correlated with the endogenous variable. The marginal value of the additional restrictions embodied in triangular, causal chain, models is very high in this circumstance but whether those restrictions are plausible is a matter for case by case consideration in the economic or other context of the application.

5. In some situations investigators have control over the amount of discreteness present in observed outcomes. For example survey designers can choose to illicit banded responses and choose the particular banding to employ. The results of this paper reveal the impact of these choices on the ability to identify interesting structural features.

6. This paper has studied the modelling of, and impact of, endogeneity in microeconometric models with discrete outcomes. On the occasion of the Festschrift Conference for Tony Lancaster it is apposite to remark that Lancaster (1985) studied[16] these issues in the context of labour market outcomes, namely unemployment durations and achieved and reservation wages. That paper was unusual for its time in its focus on the extent to which parametric models[17] used in this field possess nonparametric identifying power for particular interesting structural elements.

REFERENCES

BENKARD, C. LANIER AND STEVEN BERRY (2006): "On the Nonparametric Identification of Nonlinear Simultaneous Models: Comment on B. Brown (1983) and Roehrig (1988)," *Econometrica*, 74, 1429-1440.

CAMERON, A. COLIN AND PRAVIN K. TRIVEDI (1998): *Regression Analysis of Count Data.* An Econometric Society Monograph, Cambridge University Press: Cambridge.

CHERNOZHUKOV, VICTOR AND CHRISTIAN HANSEN (2005): "An IV Model of Quantile Treatment Effects," *Econometrica*, 73, 245-261.

CHERNOZHUKOV, VICTOR, GUIDO W. IMBENS AND WHITNEY K. NEWEY (2007): "Instrumental Variable Identification and Estimation of Nonseparable Models via Quantile Conditions," forthcoming in *Journal of Econometrics.*

CHESHER, ANDREW D., (2003): "Identification in nonseparable models," *Econometrica*, 71, 1405-1441.

CHESHER, ANDREW D.,(2005): "Nonparametric identification under discrete variation," *Econometrica*, 73, 1525-1550.

[16] There is a partial analysis in Lancaster and Chesher (1984).

[17] Such as that developed in the paper, and like that employed in the pioneering study by Kiefer and Neumann (1981).

Chesher, Andrew D.,(2007): "Identification of nonadditive structural functions," forthcoming as Chapter 1, Volume 3, *Advances in Economics and Econometrics, Theory and Applications: 9th World Congress of the Econometric Society*, edited by Richard Blundell, Torsten Persson and Whitney Newey. Cambridge University Press: Cambridge.

Ihaka, Ross, and Robert Gentleman (1996): "R: A language for data analysis and graphics," *Journal of Computational and Graphical Statistics*, 5, 299-314.

Imbens, Guido W., and Whitney K. Newey (2003): "Identification and estimation of triangular simultaneous equations models without additivity," Manuscript, paper presented at the 14th EC^2 Meeting, c*e*mmap, London, December 12th - 13th 2003.

Keifer, Nicholas M., and George R. Neumann (1981): "Individual effects in a non-linear model: Explicit treatment of heterogeneity in the empirical job search model," *Econometrica*, 49, 965-979.

Lancaster, Tony (1985): "Simultaneous equations models in applied search theory," *Journal of Econometrics*, 28, 113-126.

Lancaster, Tony and Andrew D. Chesher (1984): "Simultaneous equations with endogenous hazards," in *Studies in Labour Market Dynamics*, edited by: George R. Neumann and Niels Westergard-Nielsen. Springer Verlag: Heidelberg.

Manski, Charles, F., and Elie Tamer (2002): "Inference on regressions with interval data on a regressor or outcome," *Econometrica*, 70, 519-546.

Matzkin, Rosa L. (2003): "Nonparametric Estimation of Nonadditive Random Functions," *Econometrica*, 71, 1339 - 1376.

Matzkin, Rosa L. (2005): "Nonparametric Simultaneous Equations," manuscript.

Mullahy, John (1997): "Instrumental variable estimation of count data models: applications to models of cigarette smoking behavior," *Review of Economics and Statistics*, 79, 586-593.

Roehrig, Charles S. (1988): "Conditions for identification in nonparametric and parametric models", *Econometrica* 56, 433-447.

Newey, Whitney K., James L. Powell, and F. Vella (1999): "Nonparametric Estimation of Triangular Simultaneous Equations Models," *Econometrica* 67, 565-603.

Rothenberg, Thomas J. (1971): "Identification in parametric models," *Econometrica*, 39, 577-591.

Smith, Richard, J., and Richard W. Blundell (1986): "An exogeneity test for a simultaneous equation tobit model with an application to labor supply," *Econometrica*, 54, 679-685.

Windmeijer, Frank A.G., and João M.C. Santos Silva: (1997): "Endogeneity in count data models: an application to demand for health care," *Journal of Applied Econometrics*, 12, 281-294.

Figure 1: Endogenous ordered probit structure. Conditional probabilities given $Z = z$ that Y falls at or below the τ structural function (upper line) and strictly below the τ structural function (lower line) for $\tau \in \{0.5, 0.8\}$.

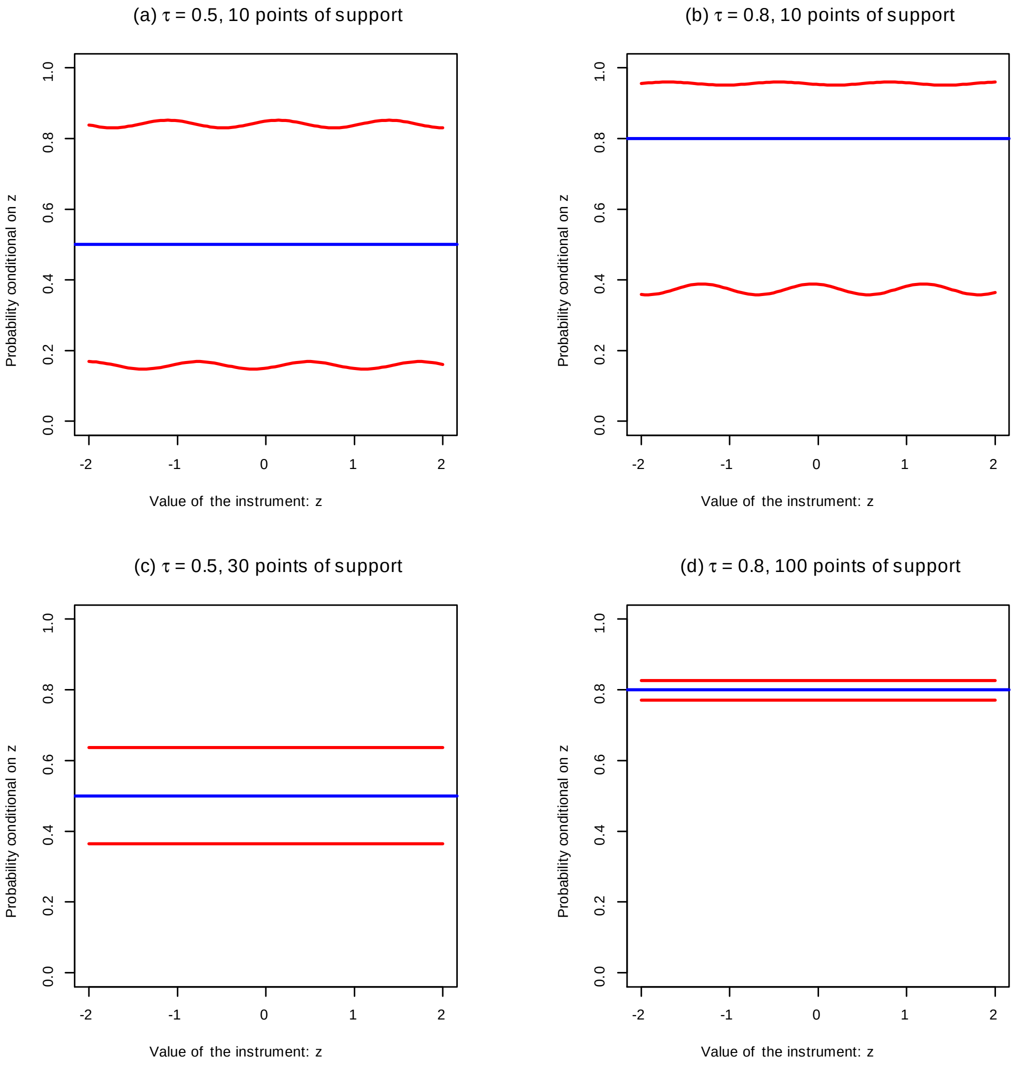

Figure 2: Endogeonous ordered probit structure. Conditional probabilities given $Z = z$ that Y falls at or below (upper line) or strictly below (lower line) a candidate $\tau = 0.5$ structural function obtained by moving the value of the slope parameter α_1 away from 1.0.

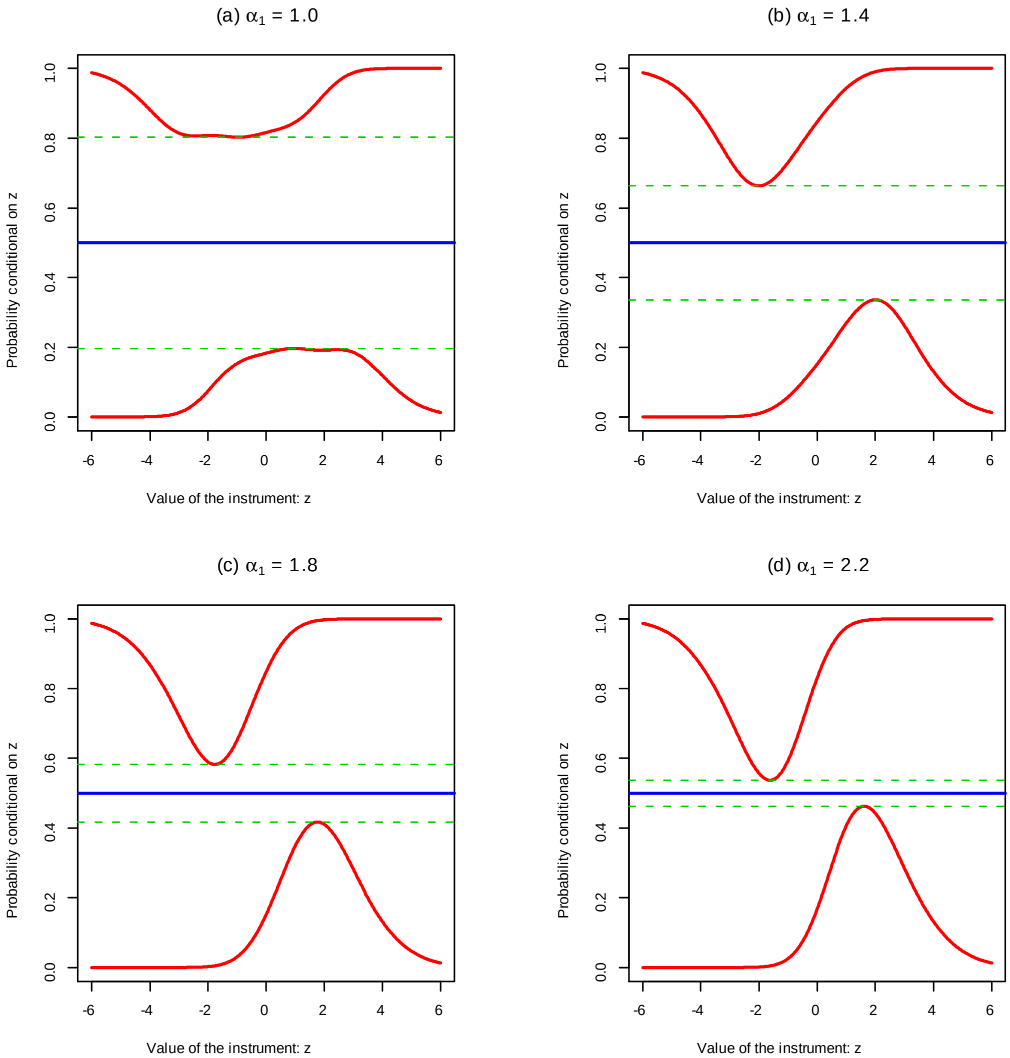

Figure 3:

Figure 4: Endogeonous ordered probit structure with stronger instrument ($\sigma_{vv} = 0.6$) Conditional probabilities given $Z = z$ that Y falls at or below (upper line) or strictly below (lower line) a candidate $\tau = 0.5$ structural function obtained by moving the value of the slope parameter α_1 away from 1.0.

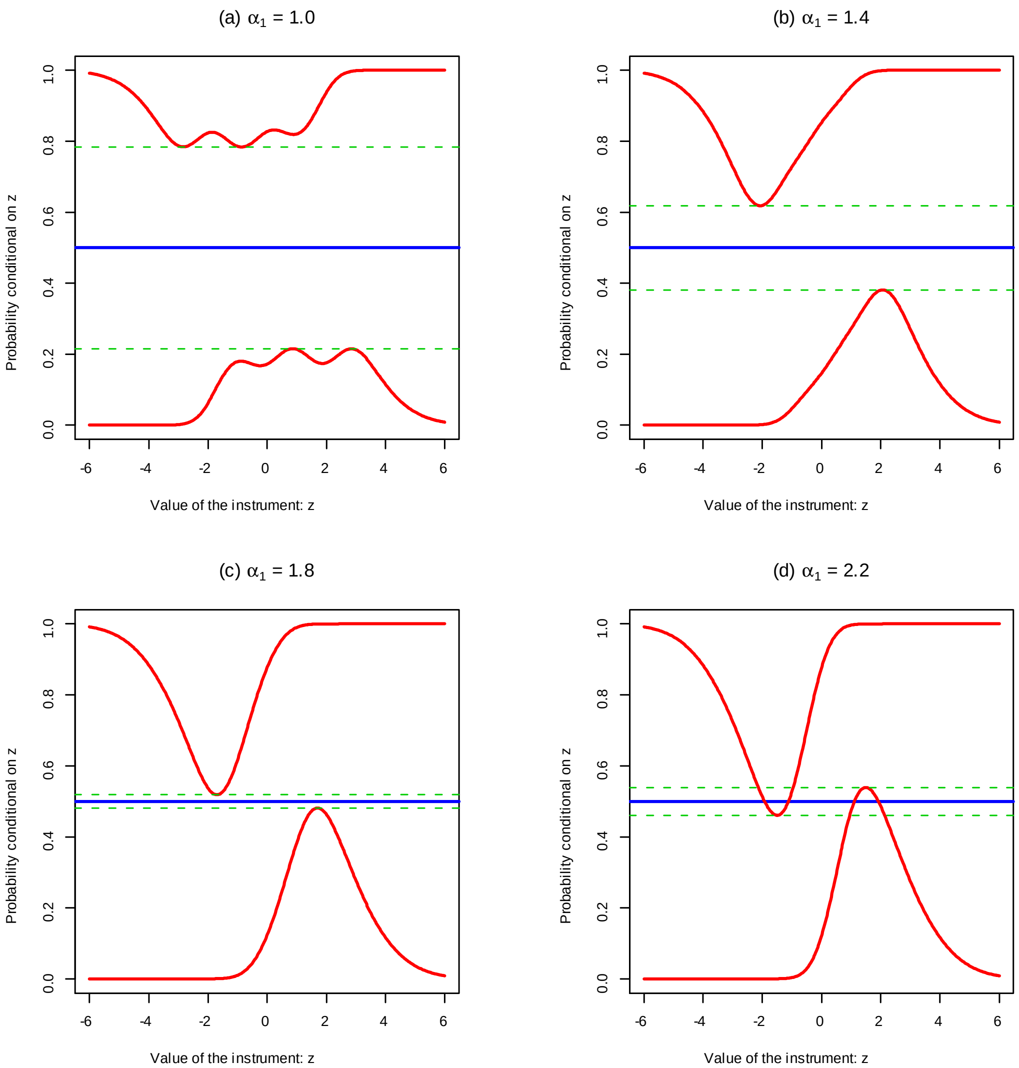

Figure 5: Endogeonous ordered probit structure with reduced discretization (7 points of support) Conditional probabilities given $Z = z$ that Y falls at or below (upper line) or stictly below (lower line) a candidate $\tau = 0.5$ structural function obtained by moving the value of the slope parameter α_1 away from 1.0.

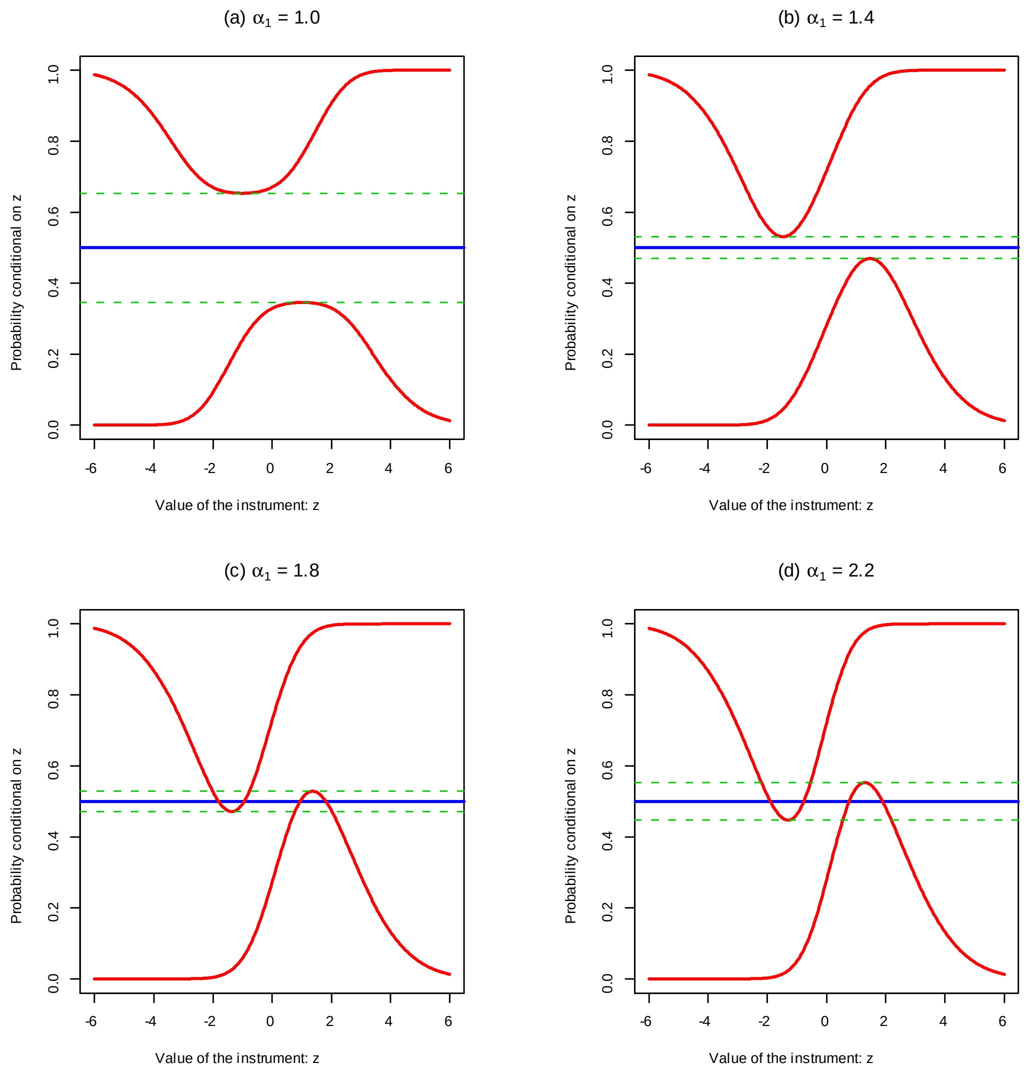

Figure 6: Sets of candidate values of α_0 and α_1 in a $\tau = 0.5$ structural function which are observationally indistinguishable from the values, $\alpha_0 = 0$ and $\alpha_1 = 1$ (filled circle), for an ordered probit structure as defined in Table 1. σ_{vv} is the variance of X given Z. Higher values corresponding to a weaker instrument.

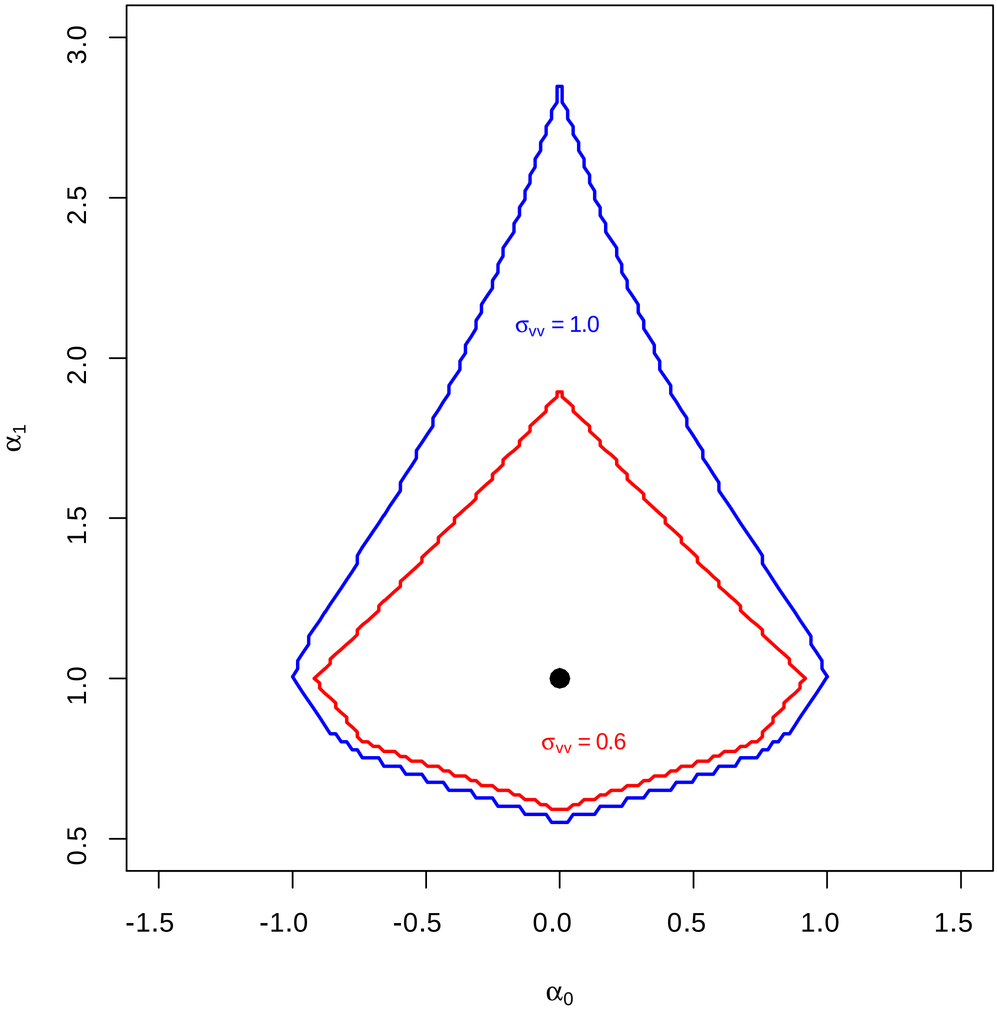

Figure 7: Bounding probabilities for the $\tau = 0.5$ structural function of the Poisson model with parameter values shown in Table 3.

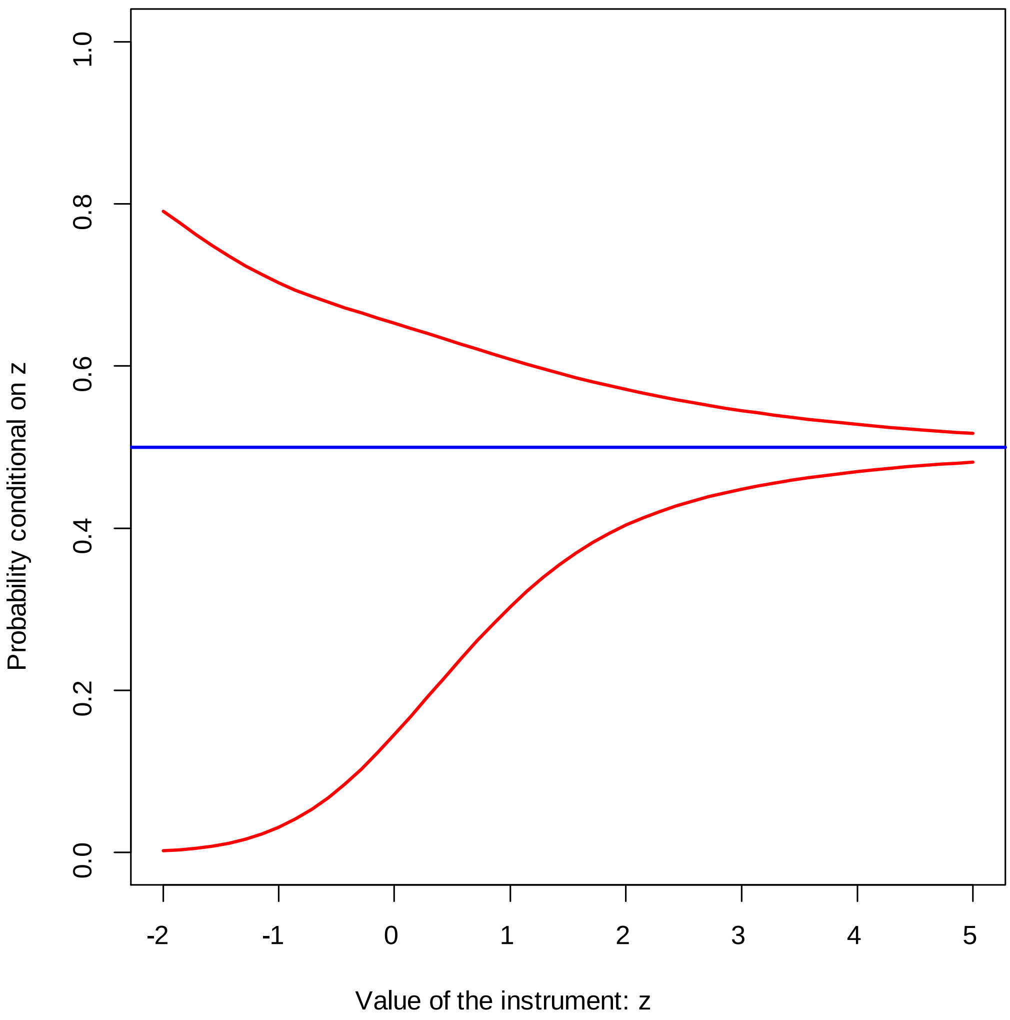

Figure 8: Bounding probabilities for a $\tau = 0.5$ structural function in the Poisson model as the value of α_1 is moved away from the value $\alpha_1 = 1$ in the stucture generating the distributions used in the calculations.

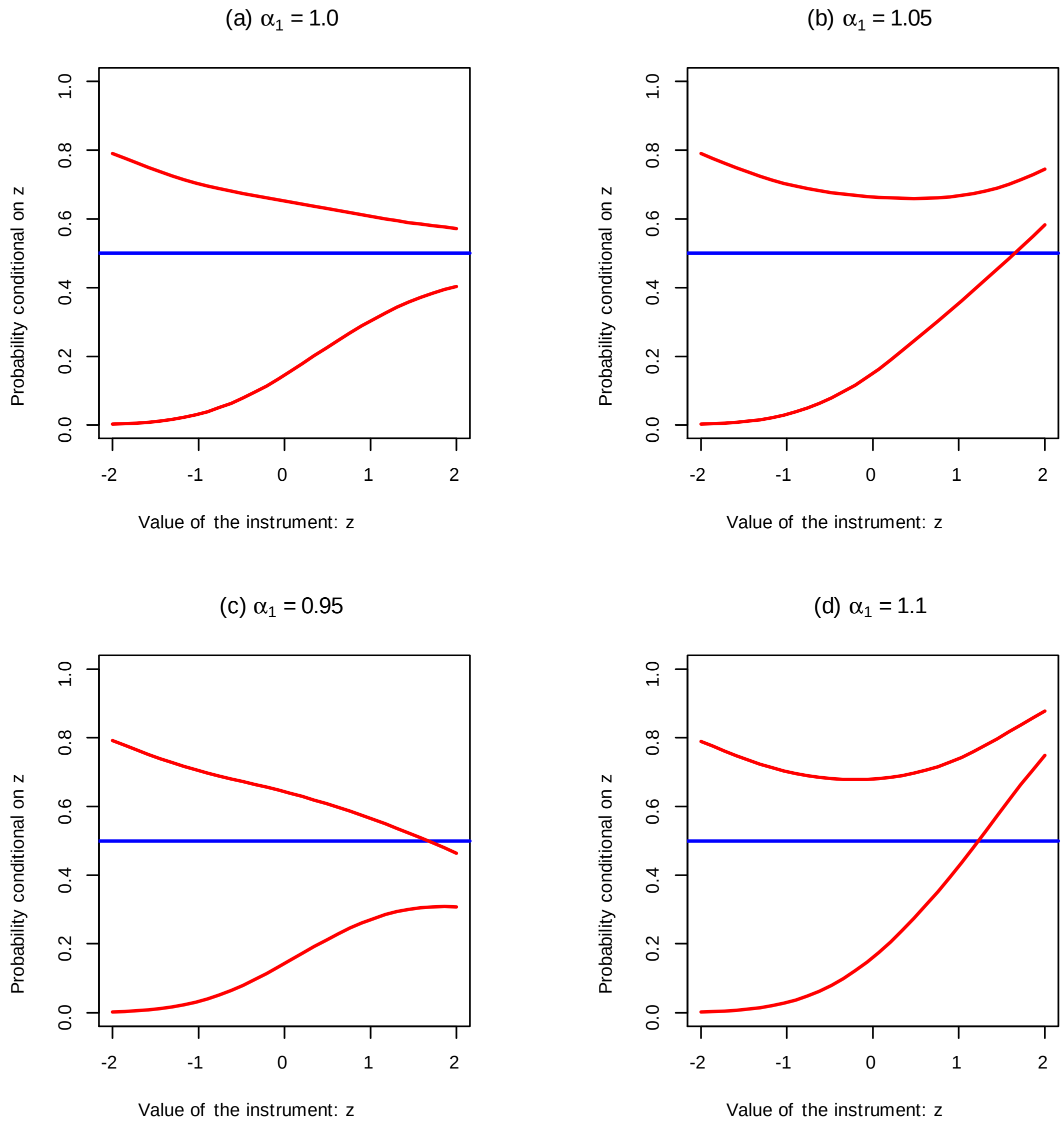

Figure 9: Bounding probabilities for the $\tau = 0.5$ structural function of the binomial model with parameter values shown in Table 4. N is the number of binomial trials and $Y \in \{0, 1, 2, \ldots, N\}$.

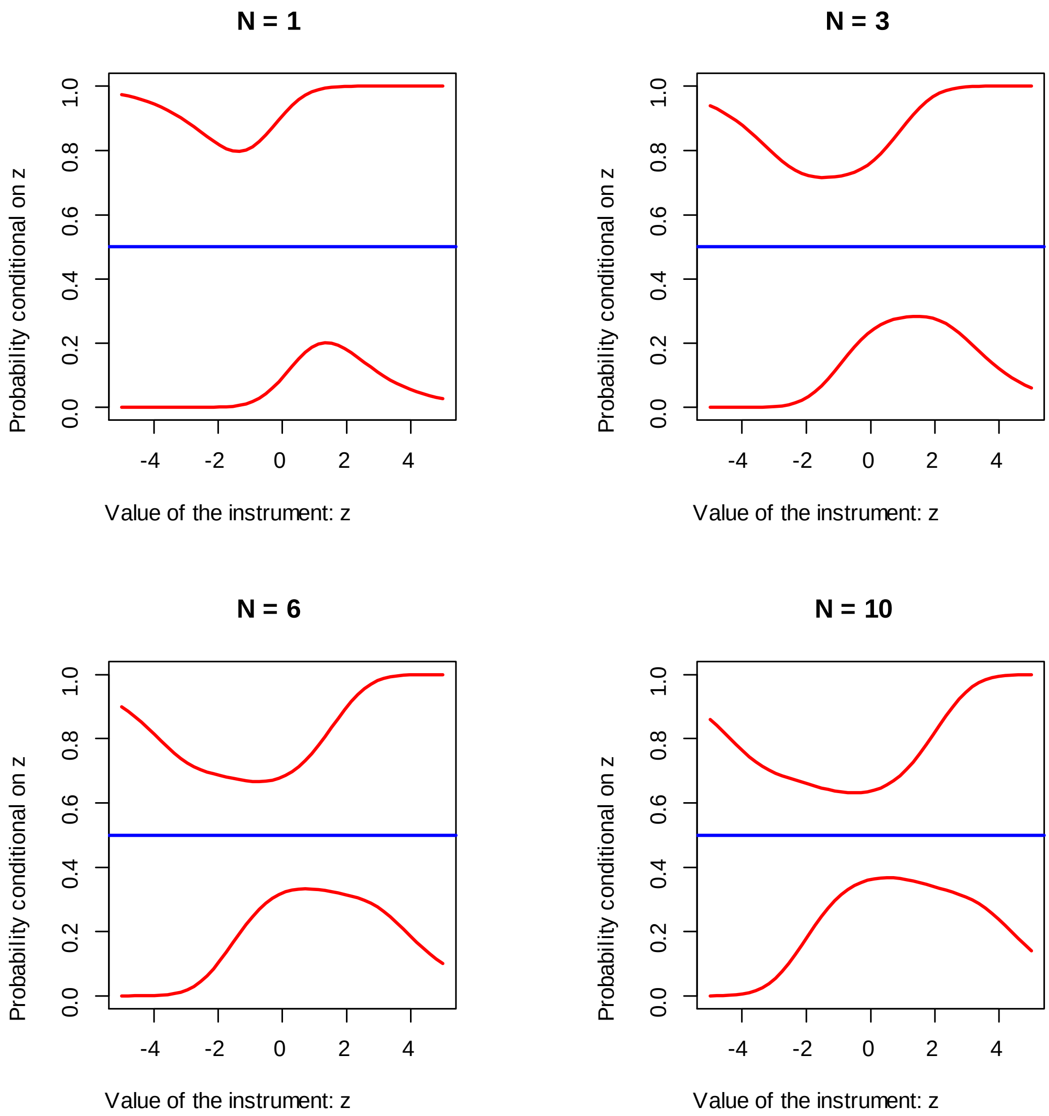

Figure 10: Bounding probabilities calculated for a candidate $\tau = 0.5$ structural function of the binomial model with $\alpha_0 = 0.5$ and $\alpha_1 = 1.0$ when probability distributions are defined by the parameter values shown in Table 4. N is the number of binomial trials and $Y \in \{0, 1, 2, \ldots, N\}$.

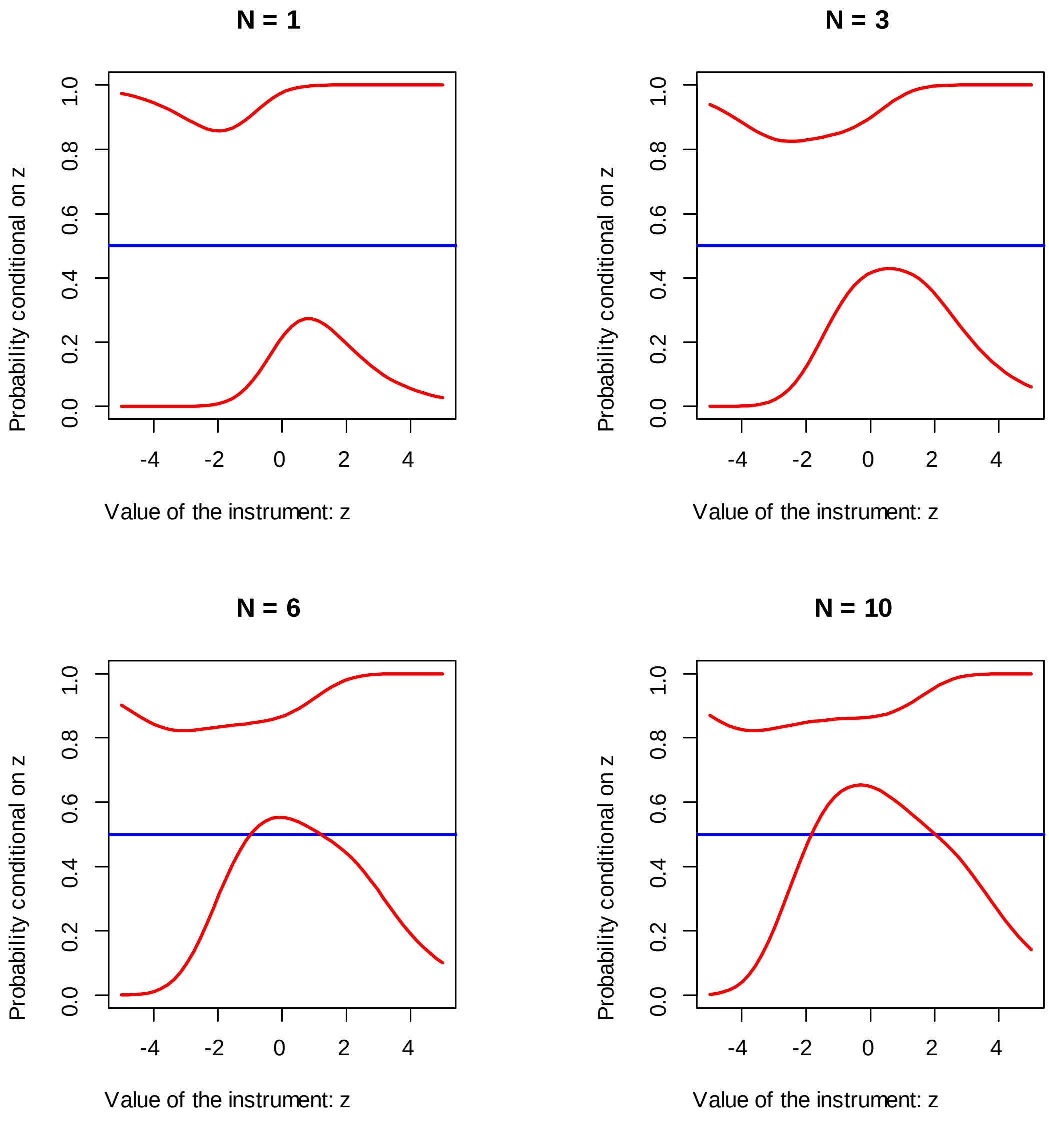

Figure 11: Bounding probabilities calculated for a candidate $\tau = 0.5$ structural function of the binomial model with $\alpha_0 = 0$ and $\alpha_1 = 0.5$ when probability distributions are defined by the parameter values shown in Table 4. N is the number of binomial trials and $Y \in \{0, 1, 2, \ldots, N\}$.

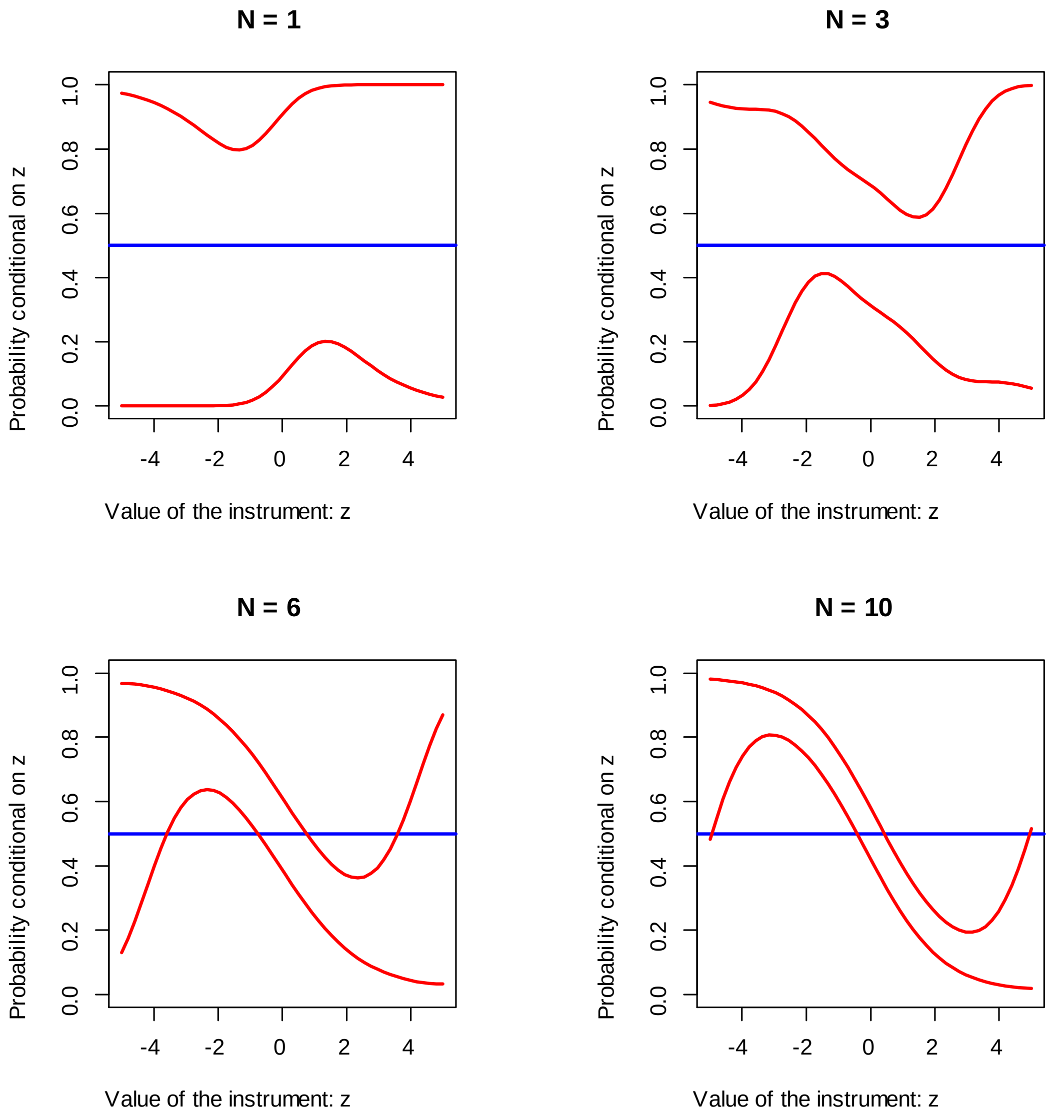

Figure 12: Bounding probabilities for the $\tau = 0.5$ structural function of the binomial model with $N = 1$ and parameter values shown in Table 4 except for σ_{vv} and σ_{wv} which vary to demonstrate the effect of increasing the strength of the instrument. In panes (a) - (c) the structural function used to construct the bounding functions is the one implied by those parameter values. In pane (d) the structural function is the $\tau = 0.5$ structural function for a structure with $\alpha_0 = 1$ rather than $\alpha_0 = 0$. The dashed lines in pane (d) are for the case $\sigma_{vv} = 1$ and the solid lines are for the case $\sigma_{vv} = 0.1$.

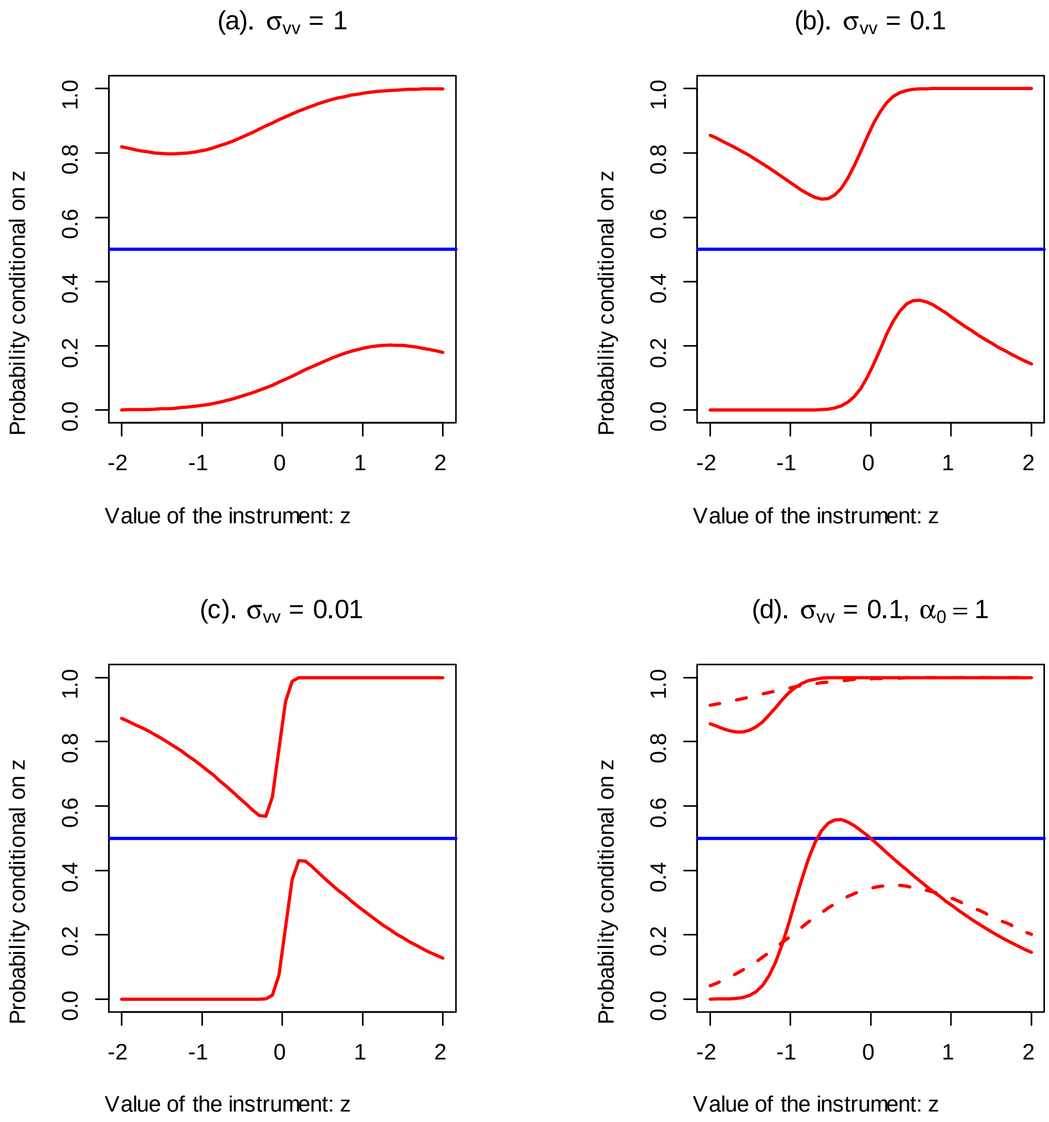

Figure 13: Sets of candidate values of α_0 and α_1 in a $\tau = 0.5$ structural function which are observationally indistinguishable from the values, $\alpha_0 = 0$ and $\alpha_1 = 1$ (filled circle), for a binomial structure as defined in Table 4. N is the number of binomial trials.