Endogeneity and Discrete Outcomes

Andrew Chesher*
Centre for Microdata Methods and Practice, UCL & IFS

Revised July 26th 2007

Abstract. This paper studies models for discrete outcomes which permit explanatory variables to be endogenous. In these models there is a single nonadditive latent variate which is restricted to be locally independent of instruments. 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 single equation IV models which, when an 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.

Keywords: Partial identification, Nonparametric methods, Nonadditive models, Discrete distributions, Ordered probit, Poisson regression, Binomial regression, Endogeneity, Instrumental variables, Structural quantile functions.

1. Introduction

This paper studies instrumental variables models for discrete outcomes in which explanatory variables can be endogenous. Outcomes can be binary, for example indicating the occurrence or otherwise of an event; they can be integer valued - for example recording counts of events; they can be ordered - perhaps giving a position on an attitudinal scale or obtained by interval censoring of an unobserved continuous outcome. Endogenous and other observed variables can be continuous or discrete.

It is shown that these IV models, which can have point identifying power when outcomes are continuous, do not point identify when outcomes are discrete. However they do have partial identifying power and this can be useful, particularly in situations in which the conditions maintained in more restrictive point identifying models are untenable. The results of the paper are derived for models in which there are no parametric restrictions. Such restrictions can be imposed using the methods developed in the paper but typically they do not deliver point identification.

*I thank Victor Chernozhukov, Martin Cripps, Russell Davidson, Simon Lee, Arthur Lewbel, Charles Manski, Lars Nesheim, Adam Rosen and Richard Spady for stimulating comments and 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 acknowlegded. Since July 1st 2007 cemmap is sponsored by the UK Economic and Social Research Council whose support is acknowledged. 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, at a Festschift conference in honour of Tony Lancaster on April 15th 2007 and at a number of subsequent workshops and conferences. I am grateful to participants for comments. This paper is dedicated to Tony Lancaster.

In the models considered here a 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 X may not be independently distributed. Since Y is discrete and U is continuous h is a step function which, in the models studied in this paper, is restricted to be weakly monotonic (normalized non-decreasing) in its final argument, U.

This nonadditive specification is adopted because, when Y is discrete, it is difficult to produce economic examples in which the unobservable U appears additively in h and the elements of the problem admit structural interpretation. One consideration here is that in applications in economics the support of a discrete outcome Y is always independent of X and in an additive U model the support of U would then have to be dependent on X. Then U and X would have to be dependently distributed and would be hard to give U a structural interpretation. The paper proceeds with consideration of nonadditive structural functions.

There are instrumental variables excluded from h. They comprise a vector random variable Z, with the property that for some $\tau \in (0, 1)$ and all z in some set Ω_Z:

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

which is in the nature of a local (to τ and to Ω_Z) independence restriction. The important feature of this restriction is that the probability in (1) does not depend on z, the value, τ, on the right hand side of (1) being a normalization. In a more restrictive model, also studied here, U and Z are globally independent and the restriction (1) holds for all $\tau \in (0, 1)$ and all $z \in \Omega_Z$.

This paper considers identification of the function $h(x, \tau)$ under the restrictions embodied in this "single equation" instrumental variables model. Triangular system models with nonadditive structural functions can have point identifying power when the outcome Y is continuous or discrete but not when endogenous X is discrete. These more restrictive triangular models are nested within the single equation IV models studied here. They are not a panacea because their restrictions do not always flow from economic considerations; Tamer (2003) gives interesting examples. Their inability to deliver point identification when endogenous variables are discrete is a serious limitation. The single equation IV model is therefore worth studying.

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 for example Chernozhukov and Hansen (2005). Since a discrete outcome can be very close to continuous if it has many points of support it seems plausible that there is a *partial* identification result for the discrete outcome case. The contribution of this paper is the development of partial identification results for the discrete outcome model.

The key to analysis of the continuous outcome case is, as noted 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, non-trivial, conditions this leads to point identification of the function $h(\cdot, \tau)$.

When Y is discrete the restrictions of the model imply that $h(\cdot, \tau)$ simultaneously satisfies *inequalities*, as follows.

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

It is shown that this leads to set identification of the structural function $h(\cdot, \tau)$.[1]

To be specific, it is shown that structural functions h which do *not* satisfy these inequalities *cannot* be elements of structures which generate the probabilities used in calculating the inequalities. Given a particular distribution of Y and X given Z, say $F_{YX|Z}$, there can be many functions h which satisfy the inequalities and so a *set* of potential structural functions which are concordant with $F_{YX|Z}$. This can lead toinformative bounds on admissible structural functions when Y has many points of support or when instruments are strong in the sense of being accurate predictors of values taken by endogenous variables. It is shown that set identification achieved using the inequalities is *sharp* in the sense that for every structural function, h, that satisfies the inequalities there exists a distribution of U and X given Z, say $F_{UX|Z}$, such that $\{h, F_{UX|Z}\}$ generate the probability distribution $F_{YX|Z}$ used to calculate the probability inequalities. Estimates of $F_{YX|Z}$ naturally lead to estimates of identified sets.

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.

The results shed light on the impact of endogeneity in situations where outcomes are by their nature discrete, for example where they are binary or records of 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 a single equation IV 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 and it is not clear what structural interpretation can be given to the results obtained when they are applied.

A control function approach can deliver identification 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 triangular models with a discrete or continuous outcome and continuous endogenous arguments in the structural function. Chesher (2005) develops a set identifying triangular model for the case in which endogenous arguments are discrete. Abrevaya,

[1]Chernozhukov and Hansen (2001) call $h(x, \tau)$ a structural quantile function. They give these inequalities but do not consider their role in the partial identification of structural functions focussing instead on the continuous outcome case. Of course there are other inequalities but they are all implied by those stated here.

Hausman and Khan (2007) study testing for causal effects in a triangular model with a discrete outcome and a discrete endogenous variable and additional monotonicity restrictions. These control function approaches set the structural equation of interest in a triangular structural *system* of equations in which all latent variates (errors) and instruments are jointly independent.

This paper studies the discrete outcome case in a *single equation* setting. Hong and Tamer (2003) and Khan and Tamer (2006) study identification and inference for linear structural functions when there is endogeneity and a degree of discreteness induced by censoring but with some continuous variation in outcomes observable. The model of this paper places no parametric restrictions on structural functions and is similar to the single equation instrumental variable model used in Chernozhukov and Hansen (2005) and Chernozhukov, Imbens and Newey (2007) but with the difference that in this paper *purely discrete* outcomes are permitted. Chesher (2007b) compares and contrasts the control function and single equation IV approaches to identification of nonadditive structural functions.

Roehrig (1988), Benkard and Berry (2006) and Matzkin (2005) study nonparametric identification in non-additive error simultaneous equation models without a triangularity restriction but only for cases in which the outcomes are continuous. Tamer (2003) studies a two equation simultaneous system in which outcomes are binary, there is *no* triangularity restriction and the two latent variables that drive stochastic variation are distributed independently of instruments. A binary outcome, binary endogenous variable special case of the model studied here applies to the Tamer (2003) problem taken one equation at a time.

The results of the paper are informative about the effect of interval censoring on the identifying power of models. 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.

Lewbel, Linton and McFadden (2006) study identification and estimation of the distribution of a continuous unobserved variable when an observable binary outcome indicates whether the latent variable exceeds an observable and varying threshold. Manski and Tamer (2002) develop partial identification results for regression functions when there is interval censoring of outcomes (and explanatory variables). Neither model permits endogeneity. This paper differs from these in that it is concerned with identification of structural functions delivering values of the observed discrete variables rather than the structural functions delivering values of the unobserved, pre-censored, continuous latent variables, and this paper focuses on identification in the presence of endogeneity.

The paper is organised as follows. Section 2 briefly reviews identification in single equation IV models for continuous outcomes with nonadditive structural functions. The main results of the paper are given in Section 3 which is concerned with models for discrete outcomes and derives set identification results. Proofs of the main propositions are given in Annexes. Section 4 illustrates the nature of the set identification results in particular ordered probit, Poisson regression and binomial regression models, in all cases with endogeneity. Ordered and binary probit and binary logit models arise as special cases in these examples. Section 5 concludes.

2. Continuous outcomes

Consider the model $\mathcal{C}$ comprising the following two restrictions.

C1. $Y = h(X, U)$ with U continuously distributed and h strictly monotonic (normalized increasing) in its last argument. The function is normalised so that $U \in [0, 1]$.

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

These restrictions are at the core of the models for continuous outcomes Y considered by Chernozhukov and Hansen (2005) and Chernozhukov, Imbens and Newey (2007). The identifying power of the model $\mathcal{C}$ is now briefly reviewed.

Let $a(\tau, x, z)$ denote a conditional distribution function for U given X and Z

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

and let $\Omega_{X|Z}$ denote the support of X which may depend upon the value taken by Z. By virtue of Restriction C1 there is, for all $z \in \Omega_Z$:

$$\int_{\Omega_{X|Z}} a(\tau, x, z) dF_{X|Z}(x|z) = \tau \tag{3}$$

where $F_{X|Z}$ is the conditional distribution function of X given Z and integration is definite over the support of X.

Since h is strictly increasing in U, applying the function to both sides of the inequality on the right hand side of (2) gives

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

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

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

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

$$\Pr[Y \leq h(X, \tau) | Z = z] = \tau. \tag{6}$$

This argument fails at the first step (4) if h is not strictly increasing in U.

Without further restriction there are many functions satisfying (6). 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 (6) along with other maintained conditions limit $h(\cdot, \tau)$ to some class of functions. The results of this paper for the discrete outcome case are of precisely this nature.

In a model in which X is locally exogenous, that is in which for some τ and all x, $\Pr[U \leq \tau | X = x] = \tau$, there is

$$\Pr[Y \leq h(x, \tau) | X = x] = \tau \tag{7}$$

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 continuous and when Y is discrete.[2]

When Y is discrete and X is endogenous neither equation (6) nor equation (7) hold. The discrete outcome case is explored now.

3. Discrete outcomes

3.1. Probability inequalities. The model $\mathcal{C}$ is now amended to permit Y to be discrete. There is the following model: $\mathcal{D}$, comprising two restrictions.

D1. $Y = h(X, U)$ with U continuously distributed and h is weakly monotonic (normalized càglàd, non-decreasing) in its last argument. Its codomain is the ascending sequence $\{y_m\}_{m=1}^M$ which is independent of X. M may be unbounded. The function is normalised so that $U \in (0, 1)$.

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

Restriction D2, which is identical to Restriction C2 in the continuous outcome model, is a local-to-τ independence restriction. A more restrictive model $\mathcal{DG}$, which will also be considered, embodies a global, full independence restriction requiring Restriction D2 to hold for all $\tau \in (0, 1)$.[3]

An important implication of the weak monotonicity condition contained in Restriction D1 is that the function $h(x, u)$ can be characterized by functions $\{p_m(x)\}_{m=0}^M$ as follows:

$$\text{for } m \in \{1, \ldots, M\}: \qquad h(x, u) = y_m \text{ if } p_{m-1}(x) < u \leq p_m(x) \tag{8}$$

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

The equality (6) does not hold under the restrictions of model $\mathcal{D}$. This is because when Y is discrete, $h(X, U)$ is not strictly increasing in U, and the equality (5) fails to hold. It is shown in Annex 1 that in its place there are the following *inequalities* which hold for *all* $\tau \in (0, 1)$ and for all x and z.

$$\begin{aligned} \Pr[Y \leq h(X, \tau) | X = x, Z = z] &\geq a(\tau, x, z) \\ \Pr[Y < h(X, \tau) | X = x, Z = z] &< a(\tau, x, z) \end{aligned} \tag{9}$$

On taking expectations with respect to X given $Z = z$ on the left and right hand sides of these inequalities and using the independence restriction embodied in (3)

[2] In the discrete case h is normalized to be càglàd for variation in U.

[3] In the model $\mathcal{D}$ the content of Ω_Z can depend on τ.

which holds at the specific value τ in the model $\mathcal{D}$, and for all τ in the model $\mathcal{DG}$, the following inequalities (10) are obtained.

$$\text{For all } z \in \Omega_Z\text{:} \quad \begin{cases} \Pr[Y \leq h(X,\tau)|Z=z] \geq \tau \\ \\ \Pr[Y < h(X,\tau)|Z=z] < \tau \end{cases} \tag{10}$$

The inequalities exhaust the information about the structural function contained in the models.[4] They are the base upon which the identification analysis of the discrete outcome IV model is constructed.

3.2. Partial identification. Let $F_{YX|Z}$ and $F_{UX|Z}$ denote distribution functions of respectively (Y,X) and (U,X) conditional on Z defined as follows

$$F_{YX|Z}(y,x|z) \equiv \Pr[Y \leq y \cap X \leq x | Z=z] = \int_{s \leq x} F_{Y|XZ}(y|s,z) dF_{X|Z}(s|z)$$

$$F_{UX|Z}(u,x|z) \equiv \Pr[U \leq u \cap X \leq x | Z=z] = \int_{s \leq x} F_{U|XZ}(u|s,z) dF_{X|Z}(s|z)$$

where $F_{Y|XZ}$ and $F_{U|XZ}$ are distribution functions of respectively Y and U conditional on X and Z.[5]

Under the weak monotonicity condition embodied in the models $\mathcal{D}$ and $\mathcal{DG}$ each structure, $S^a \equiv \{h^a, F^a_{UX|Z}\}$, delivers a conditional distribution for (Y,X) given Z, $F^a_{YX|Z}$, as follows.

$$F^a_{YX|Z}(y_m, x|z) = F^a_{UX|Z}(p^a_m(x), x|z) \tag{11}$$

This holds for $m \in \{1,\ldots,M\}$ and the functions $\{p^a_m(x)\}^M_{m=0}$ characterise the structural function h^a as set out in (8).

Data are informative about distribution functions of observable variables Y and X conditional on $Z=z$ for each value of Z that can be observed. It is assumed henceforth that the set of values of Z that can be observed is the set Ω_Z which appears in Restriction D1. It may happen that distinct structures deliver indistinguishable distributions of observables for all $z \in \Omega_Z$, that is that there exists $S^a \neq S^{a'}$ such that $F^a_{YX|Z} = F^{a'}_{YX|Z}$ for all $z \in \Omega_Z$. Such structures are observationally equivalent.

The existence of observationally equivalent structures is plausible because of the possibility of offsetting variations in the functions $p^a_m(x)$ by altering the sensitivity of $F^a_{UX|Z}(u|x,z)$ to variations in x on the right hand side of (11) while leaving the left hand side unchanged. Crucially the independence restriction embodied in the IV models studied here places limits on the variations in the functions $p^a_m(x)$ that can be so compensated. The inequalities (10) are the key to understanding these limits. There is the following Theorem.

Theorem 1. *Let* $\Pr_a$ *indicate probabilities calculated using a distribution function* $F^a_{YX|Z}$. *Consider a function* h *and a value* $\tau \in (0,1)$ *satisfying the conditions of model*

[4]The model $\mathcal{DG}$ implies other inequalities but they are all implied by (10).

[5]Inequalities hold element by element when applied to vectors.

$\mathcal{D}$. If for any $z \in \Omega_Z$ either of the inequalities

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

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

fails to hold then there exists no structure $S^* \equiv \{h^*, F^*_{UX|Z}\}$ satisfying the conditions of model $\mathcal{D}$ such that (i) $h^*(x,u)|_{u=\tau} = h(x,\tau)$ and (ii) $F^*_{YX|Z} = F^a_{YX|Z}$.

The proof, which is straightforward given the development up to this point, is given in Annex 2. The following Corollary is a direct consequence.

Corollary. *If for any $\tau \in (0,1)$ and any $z \in \Omega_Z$ either of the inequalities of Theorem 1 is violated by a function $h(x,u)$ then there is no structure in which $h(x,u)$ is a structural function that (i) satisfies the restrictions of model $\mathcal{DG}$ and (ii) is observationally equivalent to structures S^a that generate the distribution $F^a_{YX|Z}$.*

Theorem 2 concerns the sharpness of set identification induced by the inequalities of Theorem 1.

Theorem 2. *If $h(x,u)$ satisfies the restrictions of the model $\mathcal{DG}$ and does not lead to violation of the inequalities of Theorem 1 for any $z \in \Omega_Z$ and for any $\tau \in (0,1)$ then there exists a proper distribution function $F_{UX|Z}$ such that $S = \{h, F_{UX|Z}\}$ satisfies the restrictions of model $\mathcal{DG}$ and is observationally equivalent to structures S^a that generate the distribution $F^a_{YX|Z}$.*

A constructive proof of Theorem 2 for continuously varying X is given in Annex 3. A proof for discrete X can be developed along similar lines.

The essence of the result of Theorem 1 and its Corollary is that a putative structural function that violates either of the inequalities of Theorem 1 for any value of the instruments cannot be an element in a structure that generates the probabilities used in the calculation of the inequalities.

The essence of the result of Theorem 2 is that all structural functions that satisfy the inequalities of Theorem 1 for all values of the instruments $z \in \Omega_Z$ and of $\tau \in (0,1)$ are components of some admissible structure that generates the probabilities used to calculate the inequalities.

For any distribution $F^a_{YX|Z}$ there are typically many structural functions that satisfy the inequalities of Theorem 1. This is true even when structural functions are subject to parametric restrictions. Examples are given in Section 4. Thus the models $\mathcal{D}$ and $\mathcal{DG}$ *set identify* the structural function which underlies any distribution $F^a_{YX|Z}$ and there is set identification of parameter values when there are parametric restrictions.

Faced with a particular distribution $F^a_{YX|Z}$ and support of Z, Ω_Z, it may be possible to enumerate or otherwise characterize the set of structural functions that satisfy the inequalities of Theorem 1 for all $z \in \Omega_Z$, either for a particular value of τ under the restrictions of a model $\mathcal{D}$ or for all $\tau \in (0,1)$ under the global independence restriction of model $\mathcal{DG}$.

The nature and support of the instrumental variables is critical in determining the extent of the set of identified functions. In particular when there are values of Z

for which values taken by X can be predicted with high accuracy then there can be close to point identification of $h(X,U)$ at those values of X. When h is subject to parametric restrictions knowledge of h at just a few values of X and τ may secure point identification of parameter values. Unfortunately in the commonly encountered case in microeconometrics in which instruments are relatively weak, not having strong predictive power, there is only partial identification of the structural function.

The degree of discreteness in the distribution of Y affects the extent of partial identification. Examples are given in Section 4. The difference between the two probabilities in the inequalities of Theorem 1,$\Delta_\tau(z)$, is the conditional probability of the event: (Y,X) realisations lie *on* the structural function:

$$\Delta_\tau(z) \equiv \Pr[Y = h(X,\tau)|Z = z]$$

which is an event of measure zero when Y is continuously distributed. 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. In some circumstances variation in instruments can affect the degree of discreteness in the distribution of Y. The Poisson example studied in Section 4 is a case in point.

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.[6]

3.3. Estimation and inference. The focus of this paper is identification, specifically the feasibility of using data to gain information about structural functions generating discrete outcomes when there may be endogeneity and there are instrumental variable restrictions. Here are a few observations on estimation and inference, topics which are left for future research.

It is clear that, given an estimate $\hat{F}_{YX|Z}$ and any additional restrictions to be satisfied by the structural function h the set of functions that satisfy these restrictions and do not violate the "estimated" inequalities for any value of the instruments can be enumerated. Sampling variation in the estimate $\hat{F}_{YX|Z}$ will be transmitted to the set estimate thus obtained. Informative inference is likely to require further restrictions on the structural function. In the absence of parametric restrictions it will be desirable have restrictions that induce some smoothness in the X related variation in $h(X,U)$.

Another approach to set estimation exploits the fact that the inequalities of Theorem 1 can be expressed as conditional moment inequalities involving binary indicator variables. Under the model $\mathcal{D}$ there is, for some specified τ and all $z \in \Omega_Z$:[7]

$$E_{YX|Z}[1[Y \leq h(X,\tau)]|Z = z] - \tau \geq 0$$

$$E_{YX|Z}[1[Y < h(X,\tau)]|Z = z] - \tau < 0$$

[6] The completeness conditions advanced, for example, in Chernozhukov Imbens and Newey (2007) have the effect of achieving point identification in cases in which bounds coincide, but there is the difficulty for practitioners that economics does not speak loudly, perhaps not at all, about the plausibility of such conditions.

[7] $1[C]$ is 1 if C is true and 0 otherwise.

and under the model $\mathcal{DG}$ these inequalities hold for all $\tau \in (0, 1)$.

For any positive vector valued function $w(z)$ there are the unconditional moment inequalities:

$$E_{YXZ}[(1[Y \leq h(X, \tau)] - \tau)\, w(Z)] \geq 0$$
$$E_{XYZ}[(1[Y < h(X, \tau)] - \tau)\, w(Z)] < 0.$$

Under the model $\mathcal{D}$ where only one value of τ is involved these can serve as the basis for inference as in for example Andrews, Berry and Jia (2004), Moon and Schorfheide (2006), Pakes, Porter, Ho and Ishii (2006) and Rosen (2006).

Under the model $\mathcal{DG}$ there is a *continuum* of moment inequalities on which there seem to be few research results at this time although inference with point identification induced by a continuum of moment *equalities* is quite well understood, see for example Carrasco and Florens (2000).

4. Illustrations and elucidation

This Section illustrates the set identification results, showing in particular cases how the set identifying probability inequalities and identified sets vary with the nature and strength of instruments and the degree of discreteness in outcomes. The illustrations use particular parametric structures commonly used in microeconometric practice. Specifically ordered probit, Poisson and binomial structures are considered in a setting in which there is potentially endogenous variation in continuous explanatory variables.[8]

These particular parametric examples are useful because the degree of discreteness in outcomes can be tuned by altering characteristics of these structures. For example the ordered probit structure can span the range from binary to almost continuous outcomes by choice of number and spacing of thresholds. This range is spanned in a different way in the binomial structure by varying the "number of trials" parameter. In the Poisson case the effective degree of discreteness increases with the "mean parameter" introducing the possibility that variation in an instrument can affect the degree of discreteness in an outcome.

The parametric examples also bring home an important message of the paper, namely that even with parametric restrictions on structural functions the single equation IV model does not in general deliver point identification. This difficulty arises because of the incompleteness of the IV model which is silent about the genesis of endogenous variables and the role played by instrumental variables in that process.

Of course to develop the examples employed in this Section it is necessary to employ a complete data generating structure. The examples are generated as special cases of a structure in which the structural function of interest is augmented by an equation relating the endogenous variable to an instrument and a latent variable which can be correlated with the latent variable in the structural function. Transformed latent variables have a joint Gaussian distribution. The structures obey the restrictions of the model $\mathcal{DG}$ because the structural function is nondecreasing in a scalar continuously distributed latent variate which is distributed independently of

[8] Discrete endogenous variables are considered in Chesher (2007c).

an instrumental variable. *Only* the restrictions of the single equation IV model $\mathcal{DG}$ are used in the subsequent identification analysis.

4.1. A generic structure for discrete outcomes with endogeneity. The general form of the structure used in the illustrative examples is as follows.

The structural function is $h(X,U)$ with h nondecreasing in U which is normalised $Unif(0,1)$ so that there is the characterisation:

$$\text{for } m \in \{1,\ldots,M\}: \qquad Y = h(X,U) = y_m \text{ if } p_{m-1}(X) < U \leq p_m(X) \tag{12}$$

for some ascending sequence $\{y_m\}_{m=1}^M$, with $p_0(X) \equiv 0$ and $p_M(X) \equiv 1$.

If U and X are independently distributed then $\Pr[Y \leq y_m|X = x] = p_m(x)$. Particular examples are obtained by choosing particular functions $\{p_m(X)\}_{m=1}^M$. In the illustrations that follow these functions are specified as cumulative probabilities for classical covariate dependent ordered probit (including binary probit), Poisson and binomial (including binary logit) distributions with *endogeneity introduced* by allowing U and X to be jointly dependent in a manner now described.

Let Φ and Φ^{-1} denote respectively the standard normal distribution and quantile function. Values of X are generated by the auxiliary equation

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

and, with $W \equiv \Phi^{-1}(U)$, there is a joint Gaussian distribution of the latent variates conditional on $Z = z$.

$$\left[\begin{array}{c} W \\ V \end{array}\right] |Z = z \sim N\left(\left[\begin{array}{c} 0 \\ 0 \end{array}\right], \left[\begin{array}{cc} 1 & \sigma_{wv} \\ \sigma_{wv} & \sigma_{vv} \end{array}\right]\right) \tag{13}$$

It follows that X given $Z = z$ is distributed $N(\beta_0 + \beta_1 z, \sigma_{vv})$ and the conditional distribution function of Y given $X = x$ and $Z = z$ is

$$F_{Y|XZ}(y_m|x,z) = \Phi\left(\frac{1}{\sigma_{w|v}}\left(\Phi^{-1}(p_m(x)) - \frac{\sigma_{wv}}{\sigma_{vv}}(x - \beta_0 - \beta_1 z)\right)\right) \tag{14}$$

where $\sigma^2_{w|v} \equiv 1 - \sigma^2_{wv}/\sigma_{vv}$ is the conditional variance of W given V.

The structural equations for Y and X have a triangular form and (U,V) and Z are jointly independently distributed. A model that embodies those restrictions can have point identifying power when outcomes are discrete as shown in Chesher (2003). Those restrictions are *not* imposed by the single equation IV model now considered.

4.2. Ordered probit structures. In the ordered probit illustration the structural function is characterised by

$$p_m(x) = \Phi\left(\frac{1}{\alpha_2}(T_m - \alpha_0 - \alpha_1 x)\right), \; m \in \{0,\ldots,M\} \tag{15}$$

where Φ is the standard normal distribution function, $\alpha_2 > 0$, and $\{T_m\}_{m=0}^M$ are constants with $T_0 = -\infty$ and $T_M \equiv \infty$. 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 $m \in \{1,\ldots,M\}$: $Y = y_m$ for $T_{m-1} < Y^* \leq T_m$. In this case the values of the thresholds will usually be known but in other ordered response settings they may not be known.

In the particular case explored here there is throughout $\alpha_0 = \alpha_1 = 1 = \alpha_2 = 1$, $\sigma_{wv} = 0.6$, $\beta_0 = 0$ and $\beta_1 = 1$. In the first two cases studied $\sigma_{vv} = 1$ so the instrument is relatively weak, the squared correlation between X and Z being 0.36. In the first case studied Y takes 4 values with $\{T_m\}_{m=1}^{M-1} = \{-0.7, 0.0, +0.7\}$.

Now consider the admissibility of structural functions

$$p_m(x) = \Phi\left(\frac{1}{a_2}(T_m - a_0 - a_1 x)\right),\ m \in \{1,\ldots,M\} \tag{16}$$

with $a_0 = \alpha_0 = 1$ and $a_2 = \alpha_2 = 1$ and with a_1 taking values in

$$\{0.6, 0.7, 0.8, 0.9, 1.0, 1.2, 1.4, 1.6, 1.8\}.$$

The value of a_1 is allowed to vary around the value (1.0) in the structural function actually generating the probability distributions in the calculations that follow. The values of a_0 and a_2 and the thresholds are held fixed at the values in the structural function that generates the probability distributions. Variations in z in the interval $[-3, 3]$ are considered.[9]

Figure 1 shows the upper and lower bounding probabilities of Theorem 1 calculated for $a_1 \in \{1.0, 1.2, 1.4, 1.6, 1.8\}$ and the structural function obtained when $\tau = 0.5$.[10] In Figure 2 the value of τ is 0.75. In Figure 3 the value of τ is returned to 0.5 and the instrument is strengthened by reducing σ_{vv} to 0.48. This also has the effect of increasing the strength of endogeneity in the sense that the squared correlation between W and V rises from 0.36 to 0.75. In Figure 4 the value of σ_{vv} is returned to 1 and the discreteness of the outcome is reduced with $M = 11$ with

$$\{T_m\}_{m=1}^{M-1} = \{-1.25, -0.85, -0.5, -0.25, 0.0, 0.25, 0.5, 0.85, 1.25\}.$$

Table 1 summarises these settings of the parameters of the structure that generates the probabilities used in the calculations. In Figures 5-8 the parameter settings are as in Figures 1-4 but with a_1 varying in $\{0.6, 0.7, 0.8, 0.9, 1.0\}$.

First consider Figures 1 and 5. Here and in all the Figures the upper and lower bounding probabilities for $a_1 = 1$ lie respectively above and below the line making the value of τ under consideration (0.5 except in Figures 2 and 6 where $\tau = 0.75$). This is as it must be because the probabilities are calculated using $\alpha_1 = 1$. In Figures 1 and

[9] Computations were done in the R environment (Ihaka and Gentleman (1996)). The `integrate` function was used in computing the bounding probabilities and identifed sets were determined by solving an optimisation problem using recursive calls to the `optimise` function to calculate solutions.

[10] When $a_1 > 0$ the structural function $h(x,\tau)$ with parameters set equal to a_0, a_1 and a_2 can be expressed as:

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

where

$$s_m(\tau) = \frac{1}{a_1}\left(T_m - a_0 - a_2\Phi^{-1}(\tau)\right),\ m \in \{1,\ldots,M\}$$

with inequalities reversed when $a_1 < 0$.

Table 1: Parameter values for the structures generating probability distrubtions in the ordered probit examples

Figures	α_0	α_1	α_2	β_0	β_1	σ_{wv}	σ_{vv}	τ	M
1 & 5	1	1	1	0	1	0.6	1.00	0.50	4
2 & 6	1	1	1	0	1	0.6	1.00	0.75	4
3 & 7	1	1	1	0	1	0.6	0.48	0.50	4
4 & 8	1	1	1	0	1	0.6	1.00	0.50	10

5, as a_1 moves above (below) 1, the upper (lower) bounding probabilities approach the marked $\tau = 0.5$ line but they do *not* cross it. None of these values of a_1 is identifiably distinct from the value $\alpha_1 = 1$ for the $\tau = 0.5$ structural function. In Figures 2 and 6 in which $\tau = 0.75$ the lines for $a_1 \in \{0.6, 1.6, 1.8\}$ do cross the $\tau = 0.75$ line. The implication of Theorem 1 is that these are not values in the identified set of values of α_1 at these parameter settings.

As the value of τ varies different values of a_1 become admissible. There is a set of admissible values for each τ and the set of admissible values under the global independence model $\mathcal{DG}$ is the intersection of these sets. For the parameter α_1 on which we are focussing here the identified set at the settings in Figures 1, 2, 5 and 6 is the interval $(0.63, 1.23)$.

Here and in the other Figures the upper (lower) bounding probabilities approach 1 (0) as the absolute value of the instrument becomes large in magnitude. This happens because as z moves to very low or high values the distribution of X given $Z = z$ is relocated to very low or high values with the result that, with high probability Y falls on an extreme point of its support. All structural functions considered here come to coincide and deliver an extreme point of support at sufficiently extreme values of X. As z moves to extreme values, extreme values of X become highly likely and the probability that Y lies on or below any structural function approaches 1 and the probability it lies strictly below any structural function approaches 0.

In Figures 3 and 7 the value of τ is returned to 0.50 and the instrument is strengthened by reducing the variance of V to 0.48. Now the lines for $a_1 \in \{0.6, 1.6, 1.8\}$ do cross the $\tau = 0.50$ line and the result of Theorem 1 is that these are not values in the identified set of values of α_1 at these parameter settings. Strengthening the instrument reduces the extent of the identified set. Under the global independence condition of model $\mathcal{DG}$ the identified set for the parameter α_1 is the interval $(0.73, 1.16)$ with the stronger instrument.

Changing the value of the coefficient β_1 in the auxiliary equation has no effect on identified sets as long as the support of Z changes concomitantly - changing β_1 has the same effect as changing the scale on which Z varies. The support of the instrument has a very significant effect on the content of identified sets. Obviously the crossings of the τ lines noted above will only be helpful in practice if values of z in the vicinity of those crossings can be observed. In the particular parametric case studied here there is additionally a phenomenon whereby if z varies unboundedly then, under model $\mathcal{DG}$, there can be point identification achieved by crossings at extreme values of τ This would not be a useful route to point identification in practice and relies on parametric restrictions that have the effect of "linking" the behaviour of structural functions at extreme values of the endogenous variables to their behaviour at more moderate values.

In Figures 4 and 8 the instrument is returned to its original strength and the degree of discreteness in the outcome is reduced, Y now having 10 points of support. Now the lines for $a_1 \in \{0.6, 0.7, 1.4, 1.6, 1.8\}$ all cross the $\tau = 0.50$ line and the implication of Theorem 1 is that these are not values in the identified set of values of α_1 at these parameter settings. Reducing discreteness reduces the extent of the identified set. Under the global independence condition of model $\mathcal{DG}$ the identified set for the parameter α_1 is the interval $(0.79, 1.10)$ with this less discrete outcome.

In practice one would calculate joint identified sets (or estimates of them) for multiple parameters. Here for ease of exposition the focus has been on just one parameter. With discrete endogenous variables it is feasible to take a completely nonparametric approach and set identify the full structural function as shown in Chesher (2007b).

4.3. Poisson structures. In the Poisson example there are threshold functions:

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

with $\lambda(x)$ parameterized as follows.

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

This specification is commonly found in applied work. If U and X are independently distributed then

$$P[Y = y_m | X = x] = \frac{\lambda(x)^{y_m}}{y_m!} \exp(-\lambda(x))$$

and with $y_m = m - 1$, Y has a Poisson distribution conditional on X. The model $\mathcal{D}$ permits X to be endogenous and requires U and Z to be independently distributed. The identifying power of this model in a particular case is now investigated.

In the illustration the structure generating probability distributions used in the calculations has $\alpha_0 = -0.5$ and $\alpha_1 = 0.5$. X is endogenous with $\sigma_{wv} = 0.6$, $Var(X|Z) = \sigma_{vv} = 1$ and $\beta_0 = 0$, $\beta_1 = 1$.

Figures 9 and 10 show the variation with the value of the instrument in the bounding probabilities of Theorem 1, calculated for $\tau = 0.5$ structural functions using $\{p_m(x)\}_{m=1}^{\infty}$ as defined above with $\lambda(x) = \exp(a_0 + a_1 x)$, with $a_0 = -0.5$ and a_1 varying in $\{0.1, 0.2, \dots, 0.8, 0.9\}$. At low values of the instrument these variations in a_1 about the probability distribution generating value $\alpha_1 = 0.5$ are not identifiably distinguishable using the model $\mathcal{D}$. However at high values of the instrument quite small departures from $\alpha_1 = 0.5$ can be identified. Indeed, if large enough values lie in the support of Z then virtually all departures can be detected and there is point identification.

This "identification at infinity" phenomenon arises because Z has a positive effect on X which in turn has a positive effect on the threshold functions. The result is that as Z increases the discrete outcome tends to take larger and larger values and to become more dispersed, the discreteness in the outcome virtually disappears, and a position where there is point identification is approached. This would not occur with some other specifications of $\lambda(x)$ or if the impact of Z on X is limited. In practice one encounters rather small values of the counts in the data to which Poisson models are brought and point identification *via* this device using a single equation IV model is infeasible.

4.4. Binomial structures. In the binomial structure there are threshold functions:

$$p_m(x) = \sum_{y=0}^{m-1} \binom{N}{y} \gamma(x)^y \left(1 - \gamma(x)\right)^{N-y} \qquad m \in \{1, 2, \ldots, M\} \tag{19}$$

with $M = N + 1$ and in this illustration $\gamma(x)$ is parameterized as follows.

$$\gamma(x) \equiv \frac{\exp(\alpha_0 + \alpha_1 x)}{1 + \exp(\alpha_0 + \alpha_1 x)} \tag{20}$$

When U and X are independently distributed Y has a binomial distribution conditional on X with point probabilities as follows.

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

When $N = 1$ this is a parametric binary logit model. The model $\mathcal{D}$ permits X to be endogenous requiring U and Z to be independent. The identifying power of this model in a particular case is now investigated.

In the illustration the structure generating probability distributions used in the calculations has $\alpha_0 = 0.75$ and $\alpha_1 = 1.0$. X is endogenous with $\sigma_{wv} = 0.6$, $Var(X|Z) = \sigma_{vv} = 1$ and $\beta_0 = 0$, $\beta_1 = 1$.

Figures 11-16 show the bounding probabilities of Theorem 1 calculated for $\tau = 0.5$ structural functions using $\{p_m(x)\}_{m=1}^{M}$ as defined above with

$$\gamma(x) = \frac{\exp(a_0 + a_1 x)}{1 + \exp(a_0 + a_1 x)}$$

$a_0 = 0.75$ and a_1 varying as follows.

$$a_1 \in \{0.6, 0.7, 0.8, 0.9, 1.0, 1.2, 1.4, 1.6, 1.8\}$$

The first two Figures (11 and 12) in the sequence show bounding probabilities for the case $N = 1$ in which Y is generated by a binary logit structure with endogenous X. In each case the calculation is done using the probability distribution generated by the structure which has $\alpha_1 = 1$. In no case does the bounding probability cross the $\tau = 0.5$ line and there is no close approach at all. Very wide variations in a_1 are accommodated in the identified set when Y is binary. The single equation IV model is quite uninformative in this binary case, a case studied in more detail in Chesher (2007c).

In Figures 13 and 14 $N = 3$ and none of the bounding probability functions cross the $\tau = 0.5$ line although there is very close approach for values of a_1 far from 1. In Figures 15 and 16 $N = 6$ and now structural functions with a_1 outside the interval $[0.7, 1.4]$ can be ruled inadmissible when the probability distribution generating structure has $\alpha_1 = 1$. As N is increased further the bounding probabilities become increasingly informative and close to point identification is achieved by the single equation IV model.

5. Concluding remarks

Single equation IV models for discrete outcomes are not point identifying. However they can have partial identifying power and econometric analysis using these models can be informative. They have the advantages that they are less restrictive than commonly used triangular system models which lead to control function approaches to estimation and that they can be employed when endogenous explanatory variables are discrete. They offer the possibility of a single equation attack on estimation in the sorts of incomplete models studied in Tamer (2003). The minimalist, but perhaps palatable, restrictions imposed in the single equation IV model lead to partial identification of deep structural objects which may be of greater interest than point identification of the various averages of structural features which have featured largely in the recent nonparametric identification literature.

Some of the restrictions embodied in the model and examples can be relaxed. For example it is easy to generalise to the case in which exogenous variables appear in the structural function. In the examples there is just one instrumental variable and parametric restrictions are considered. The results of the paper do apply when there are many instruments and it is interesting to consider the over-partial-identification that may then result. This, and tests for the validity of partially identifying instrumental variable restrictions are the subject of current research. A fully nonparametric analysis of a model for binary outcomes with discrete endogenous variables is contained in Chesher (2007c).[11]

Some restrictions cannot be relaxed without a major consequent change in the nature of the identification that can be achieved. The monotonicity restriction and the requirement that there is just one scalar unobservable variable in the structural function are examples. In both cases one would lose the ability to identify pointwise features of the structural function although various average structural features may be identifiable. Even if one feels it essential to permit multiple sources of heterogeneity in econometric models for discrete outcomes the results of this paper are of interest as they show what one loses as one steps away away from the classical Cowles-type model restriction of one-error-per-outcome.

It is clear from the examples studied here that when discreteness is a significant aspect of an outcome the identifying power of a single equation IV model can be low unless instruments are very strong. 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. Finally, thinking about survey

[11] Some results were given at the 60th Birthday Conference in Honour of Peter Robinson on May 25th 2007.

design, it is clear that the impact of discreteness on the identifying power of models is an element that should be considered when deciding whether to illicit banded responses and when deciding what type of banding to employ. Measurement matters.

References

Abrevaya, Jason, Hausman, Jerry A., and Shakeeb Kahn (2007): "Testing for Causal Effects in a Generalized Regression Model with Endogenous Regressors," unpublished manuscript.

Andrews, Donald W.K., Berry, Stephen, and Panle Jia (2004): "Confidence Regions for Parameters in Discrete Games with Multiple Equilibria, with an Application to Discount Chain Store Locations," Manuscript, Yale University, Department of Economics.

Benkard, C. Lanier and Stephen 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.

Carrasco, Marine, and Jean Pierre Florens (2000): "Generalization of GMM to a continuum of moment conditions," *Econometric Theory,* 16, 797–834.

Chernozhukov, Victor And Christian Hansen (2001): "An IV Model of Quantile Treatment Effects," MIT Department of Economics Working Paper No. 02-06.

Chernozhukov, Victor And Christian Hansen (2005): "An IV Model of Quantile Treatment Effects," *Econometrica*, 73, 245-261.

Chernozhukov, Victor, Imbens, Guido W., and Whitney K. Newey (2007): "Instrumental Variable Estimation of Nonseparable Models," *Journal of Econometrics*, 139, 4-14.

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.

Chesher, Andrew D.,(2007a): "Instrumental Values," *Journal of Econometrics*, 139, 15 – 34.

Chesher, Andrew D.,(2007b): "Identification of nonadditive structural functions," 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.

Chesher, Andrew D.,(2007c): "The Partial Identifying Power of Instrumental Variable Models for Binary Data," mimeo, in preparation.

Hong, Han, and Elie Tamer (2003): "Inference in Censored Models with Endogenous Regressors," *Econometrica*, 71, 905-932.

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, cemmap, London, December 12th - 13th 2003.

Kahn, Shakeeb and Elie Tamer (2006): "Quantile Minimum Distance Estimation of Randomly Censored Regression Models," unpublished manuscript.

Lewbel, Arthur, Linton, Oliver and Daniel L. McFadden (2006): "Estimating Features of a Distribution from Binomial Data," STICERD Discussion Paper EM/2006/507.

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," unpublished manuscript.

Moon, Hyungsik Roger and Frank Schorfheide (2006): "Boosting Your Instruments: Estimation with Overidentifying Inequality Moment Conditions," IEPR Working Paper No. 06.56.

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.

Pakes, Ariel., Porter, J., Ho, K., and J. Ishii, (2006): "Moment Inequalities and Their Application," mimeo, Harvard University.

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

Rosen, Adam (2006): "Confidence Sets for Partially Identified Parameters that Satisfy a Finite Number of Moment Inequalities," CeMMAP Working Paper, CWP25/06.

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

Tamer, Elie (2003): "Incomplete Simultaneous Discrete Response Model with Multiple Equilibria," *Review of Economic Studies*, 70, 147-165.

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.

Annex 1

Probability inequalities

It is shown that under the restrictions of model $\mathcal{D}$ there are the following inequalities which hold for all x, z and τ.

$$\begin{aligned}
&\Pr[Y \leq h(X,\tau)|X=x, Z=z] \geq a(\tau, x, z) \\
&\Pr[Y < h(X,\tau)|X=x, Z=z] < a(\tau, x, z)
\end{aligned} \tag{A1.1}$$

Consider the *first* inequality. Directly from the definition of Y:

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

and because $h(x,\tau) = y_m$ if and only if $\tau \in (p_{m-1}(x), p_m(x)]$ there is the following.

$$\Pr[h(X,U) \leq h(X,\tau)|X=x, Z=z] = \\ \sum_{m=1}^{M} 1[h(x,\tau)=y_m] \Pr[h(X,U) \leq h(X,p_m(x))|X=x, Z=z]$$

$1[\cdot]$ is the indicator function, equal to 1 if its argument is true and 0 otherwise. Since $h(x,u)$ is nondecreasing with points of increase only at $u \in \{p_1(x), \dots, p_M(x)\}$

$$\begin{aligned}
\Pr[h(X,U) \leq h(X,p_m(x))|X=x, Z=z] &= \Pr[U \leq p_m(x)|X=x, Z=z] \\
&\equiv a(p_m(x), x, z)
\end{aligned}$$

and so

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

and since $a(t,x,z)$ is a strictly increasing function of t and $h(x,\tau) = y_m$ if and only if $\tau \in (p_{m-1}(x), p_m(x)]$ there is on substituting $h(X,U) = Y$

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

which is the required inequality.

Consider the *second* inequality in (A1.1). Directly from the definition of Y:

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

and because $h(x,\tau) = y_m$ if and only if $\tau \in (p_{m-1}(x), p_m(x)]$ there is the following.

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

Since $h(x, u)$ is nondecreasing with points of increase only at $u \in \{p_1(x), \ldots, p_M(x)\}$

$$\begin{aligned} \Pr[h(X,U) \leq h(X, p_{m-1}(x))|X = x, Z = z] &= \Pr[U \leq p_{m-1}(x)|X = x, Z = z] \\ &\equiv a(p_{m-1}(x), x, z) \end{aligned}$$

and so

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

and because $a(t, x, z)$ is a strictly increasing function of t and $h(x, \tau) = y_m$ if and only if $\tau \in (p_{m-1}(x), p_m(x)]$ there is on substituting $h(X, U) = Y$,

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

which is the required inequality.

Annex 2

Proof of Theorem 1

Consider structures $S^* \equiv \{h^*, F^*_{UX|Z}\}$ satisfying the conditions of model $\mathcal{D}$ and such that $h^*(x, u)|_{u=\tau} = h(x, \tau)$. The inequalities in the statement of Theorem 1 must hold for all $z \in \Omega_Z$ with $\Pr^a$ replaced by $\Pr^*$. Therefore if for some $z \in \Omega_Z$ either inequality fails to hold then $F^*_{YX|Z} \neq F^a_{YX|Z}$.

Annex 3

Proof of Theorem 2

The proof proceeds by considering a structural function $h(x, u)$, that:

1. is weakly monotonic nondecreasing for variations in u,
2. is characterised by functions $\{p_m(x)\}_{m=0}^{M}$,
3. satisfies the inequalities of Theorem 1 when probabilities are calculated using a conditional distribution $F_{YX|Z}$.

A proper conditional distribution $F_{UX|Z}$ is constructed which respects the independence restriction (U and Z are independent) and has the property that the distribution function generated by $\{h, F_{UX|Z}\}$ is identical to $F_{YX|Z}$ used to calculate the probabilities in Theorem 1.[12]

[12] This method of construction builds on a method proposed for the discrete endogenous variable, binary outcome case by Martin Cripps.

Most of the proof is concerned with constructing a distribution for U conditional on *both* X *and* Z, $F_{U|XZ}$. This is combined with $F_{X|Z}$, the (identified) distribution of X conditional on Z implied by $F_{YX|Z}$, in order to obtain the required distribution of (U, X) conditional on Z.

The construction of $F_{UX|Z}$ is done for a representative value, z, of Z. The argument of the proof can be repeated for any z such that the inequalities of Theorem 1 are satisfied.

Unless otherwise stated below $m \in \{1, \ldots, M\}$ where M is the number of points of support of Y and M may be unbounded. It is helpful to introduce some abbreviated notation.

Define conditional probabilities as follows.

$$\alpha_m(x) \equiv \Pr[Y = y_m | X = x, Z = z] = F_{Y|XZ}(y_m|x, z)$$

$$\alpha_m \equiv \Pr[Y = y_m | Z = z] = \int\limits_{x \in \Omega_{X|Z}} \alpha_m(x) dF_{X|Z}(x|z)$$

$$\bar{\alpha}_m(x) \equiv \sum_{j=1}^{m} \alpha_j(x) \qquad\qquad \bar{\alpha}_m \equiv \sum_{j=1}^{m} \alpha_j$$

Define $\alpha_0(x) \equiv \alpha_0 \equiv 0$, and $\sum_{m=1}^{0} \alpha_m \equiv 0$ and note that $\bar{\alpha}_M = 1$ and for all x, $\bar{\alpha}_M(x) = 1$. Both $\{\alpha_m(x)\}_{m=0}^{M}$ and $\{\alpha_m\}_{m=0}^{M}$ depend on z but, to avoid clutter, dependence on z is not made explicit at many points in the notation in this Annex.

Define functions:

$$u_m(v) = \begin{cases} 0 & , \quad 0 < v \leq \sum_{j=1}^{m-1} \alpha_j \\ v - \sum_{j=1}^{m-1} \alpha_j & , \quad \sum_{j=1}^{m-1} \alpha_j < v \leq \sum_{j=1}^{m} \alpha_j \\ \alpha_m & , \quad \sum_{j=1}^{m} \alpha_j < v \leq 1 \end{cases}$$

which have the property $\sum_{m=1}^{M} u_m(v) = v$.

Define sets as follows. Let $\Omega_{X|Z}$ denote the support of X conditional on Z. Let ϕ denote the empty set.

For $m \in \{1, \ldots, M\}$:

$$X_m(s) \equiv \{x : p_m(x) = s\}$$

for $m \in \{1, \ldots, M-1\}$:

$$X_m[s] \equiv \{x : p_m(x) \leq s\}$$

and, for the case $m = M$:

$$X_M[s] \equiv \{x : p_{M-1}(x) \leq s\}.$$

Define

$$s_m(v) \equiv \min_s \left\{ s : \int_{x \in X_m[s]} \alpha_m(x) dF_{X|Z}(x|z) \geq u_m(v) \right\}$$

and define functions $\beta_m(v, x)$ as follows.

$$\beta_m(v, x) \equiv \begin{cases} \alpha_m(x) & , \quad x \in X_m[s_m(v)] \\ 0 & , \quad x \notin X_m[s_m(v)] \end{cases}$$

$$\beta(v,x) \equiv \sum_{m=1}^{M} \beta_m(v,x)$$

For a structural function $h(x,u)$ characterised by $\{p_m(x)\}_{m=1}^{M}$ the distribution function $F_{U|XZ}$ is defined as

$$F_{U|XZ}(u|x,z) \equiv \beta(u,x)$$

where z is the value of Z upon which there is conditioning at various points in the definition of $\beta(v,x)$. The distribution function $F_{UX|Z}$ is then obtained as

$$F_{UX|Z}(u|x,z) = \int\limits_{s\leq x} F_{U|XZ}(u|s,z)dF_{X|Z}(s|z).$$

It is now shown that $F_{U|XZ}$ is a *proper distribution function* exhibiting the *independence property* $U \perp\!\!\!\perp Z$, that is:

$$\int_{\Omega_{X|Z}} F_{U|XZ}(u|x,z)dF_{X|Z}(x|z) \equiv F_{U|Z}(u|z) = u$$

for all u.

It is required to show that (1) $\beta(0,x) = 0$ for all x, (2) $\beta(1,x) = 1$ for all x, (3) for each x, $\beta(v,x) \geq \beta(v',x)$ for all $v > v'$ and (4) independence, specifically: $E_{X|Z}[\beta(v,X)|z] = v$ for all z and recall that $\beta(v,x)$ depends on z although this is not made explicit in the notation.

1. $\beta(0,x) = 0$ for all x. For each m, $u_m(0) = 0$, so $s_m(0) = 0$. Therefore for each m, $X_m[s_m(0)] = \phi$, so for each m and all x, $\beta_m(0,x) = 0$ and so for all x, $\beta(0,x) = 0$.

2. $\beta(1,x) = 1$ for all x. For each m, $u_m(1) = \alpha_m$, so $s_m(1) = 1$. Therefore for each m, $X_m[s_m(1)] = \Omega_{X|Z}$, so for each m and all x, $\beta_m(1,x) = \alpha_m(x)$ and so, on summing across m, for all x, $\beta(1,x) = 1$.

3. $\beta(v,x) \geq \beta(v',x)$ for all $v > v'$. The functions $u_m(v)$ are nondecreasing, therefore the functions $s_m(v)$ are nondecreasing. Therefore for all $v > v'$, $X_m[s_m(v)] \subseteq X_m[s_m(v')]$, and so each function $\beta_m(v,x)$ is nondecreasing from which the result follows.

4. Independence. For each m and all v:

$$\int\limits_{x\in\Omega_{X|Z}} \beta_m(v,x)dF_{X|Z}(x|z) = \int\limits_{x\in X_m[s_m(v)]} \alpha_m(x)dF_{X|Z}(x|z) = u_m(v)$$

   and so

$$\int\limits_{x\in\Omega_{X|Z}} \beta(v,x)dF_{X|Z}(x|z) = \int\limits_{x\in\Omega_{X|Z}} \sum_{m=1}^{M} \beta_m(v,x)dF_{X|Z}(x|z) = \sum_{m=1}^{M} u_m(v) = v.$$

It is now shown that $F_{UX|Z}$, defined above, has an *observational equivalence* property. Specifically it is shown that, when $h(x,u)$ satisfies the inequalities of Theorem 1, the structure $\{h, F_{UX|Z}\}$, employing $F_{UX|Z}$ defined above, generates $F_{YX|Z}$ which defines the values of the probabilities $\{\alpha_m(x)\}_{m=1}^M$ and $\{\alpha_m\}_{m=1}^M$ that are employed in its construction.

Expressed in terms of the functions $\{p_m(x)\}_{m=0}^M$ the inequalities take the following form.

$$\sum_{m=1}^{M} \int\limits_{p_{m-1}(x)<u\leq p_m(x)} \bar{\alpha}_{m-1}(x)dF_{X|Z}(x|z) < u \leq \sum_{m=1}^{M} \int\limits_{p_{m-1}(x)<u\leq p_m(x)} \bar{\alpha}_m(x)dF_{X|Z}(x|z)$$

This involves the cumulative probabilities $\bar{\alpha}_m(x)$. It is convenient to write the inequalities in terms of the point probabilities $\{\alpha_m\}_{m=1}^M$, as follows.

$$\sum_{m=1}^{M-1} \int\limits_{p_m(x)<u} \alpha_m(x)dF_{X|Z}(x|z) < u \leq \alpha_1 + \sum_{m=2}^{M} \int\limits_{p_{m-1}(x)<u} \alpha_m(x)dF_{X|Z}(x|z) \qquad \text{(A3.1)}$$

It is now shown that, under this condition h together with $F_{U|XZ}$ defined above (that is $\beta(u,x)$) generates $F_{Y|XZ}$. That happens if and only if the following conditions hold: for each m and all x:

$$\beta(p_m(x), x) = \bar{\alpha}_m(x)$$

which is true if, for all i and j and all x the following conditions hold.

$$\beta_j(p_i(x), x) = \begin{cases} \alpha_j(x) & , \quad j \leq i \\ 0 & , \quad j > i \end{cases} \qquad \text{(A3.2)}$$

It is now shown that when the constraints (A3.1) are satisfied this condition is satisfied.

First consider the case $j = i$. Every term on the left hand side of (A3.1) is nonnegative, so when the constraints are satisfied, for each i and all u

$$\int\limits_{p_i(x)<u} \alpha_i(x)dF_{X|Z}(x|z) < u$$

that is for some $\delta(u) > 0$

$$\int\limits_{p_i(x)<u} \alpha_i(x)dF_{X|Z}(x|z) = u + \delta(u)$$

equivalently

$$\int\limits_{p_i(x)<u-\delta(u)} \alpha_i(x)dF_{X|Z}(x|z) = u$$

and so for all v, $s_i(v) > v$. It follows that for each i, $X_i[s_i(v)]$ contains $X_i[v]$ of which $X_i(v) = \{x : p_i(x) = v\}$ is a subset. Therefore, if the constraints (A3.1) hold then $\beta_i(p_i(x), x) = \alpha_i(x)$.

Now consider the case $i > j$. Because each $\beta_j(v, x)$ is nondecreasing in v, and because for all $i > j$, $p_i(x) \geq p_j(x)$, $\beta_j(p_i(x), x) \geq \beta_j(p_j(x), x)$. But the maximum value that $\beta_j(v, x)$ can take is $\alpha_j(x)$. It follows that for all $i \geq j$, $\beta_j(p_i(x), x) = \alpha_j(x)$.

Now consider the case $i = j - 1$. Consider some j and $u < p_j(x)$ and the right hand side of the constraints (A3.1). All contributions from terms in the summation with $i > j$ are zero. All contributions with $i < j$ are bounded by α_i and so there is the following inequality.

$$\int\limits_{p_{j-1}(x)<u} \alpha_j(x) dF_{X|Z}(x|z) \geq u - \sum_{i=1}^{j-1} \alpha_i$$

Consider the content of the set $X_j[s_j(v)]$. For $v < \sum_{i=1}^{j-1} \alpha_i$, $u_j(v)$ is zero and the set is empty. For $p_j(x) > v \geq \sum_{i=1}^{j-1} \alpha_i$, the inequality above requires that $s_j(v) \leq v$. In particular, for $i < j$, $s_j(p_i(x)) < p_i(x)$. It follows that $X_j[s_j(v)]$ has no intersection with $X_i(v)$ and so $\beta_j(p_i(x), x) = 0$ for $i < j$.

Finally consider the case $i < j - 1$. Because each $\beta_j(v, x)$ is nondecreasing in v, and because for all $i < j - 1$, $p_i(x) \leq p_{j-1}(x)$, $\beta_j(p_i(x), x) \leq \beta_j(p_{j-1}(x), x) = 0$ and so for all $i < j - 1$, $\beta_j(p_i(x), x) = 0$. This concludes the demonstration that (A3.2) holds.

It has been shown that $F_{U|XZ}(u|x, z) = \beta(u, x)$, constructed as above, is a proper distribution function respecting the independence restriction, $U \perp\!\!\!\perp Z$, delivering, with the structural function h, the conditional distribution function $F_{Y|XZ}$. It follows that $F_{U|XZ}$ defined as above, brought together with $F_{X|Z}$ to produce $F_{UX|Z}$, combines with h to deliver $F_{YX|Z}$.

The inequalities of Theorem 1 are crucial in endowing $\{h, F_{UX|Z}\}$ with the observational equivalence property. It has been shown that for each h that satisfies those inequalities there exists at least one distribution $F_{UX|Z}$ such that $\{h, F_{UX|Z}\}$ generates the distribution $F_{YX|Z}$ used to calculate the probability inequalities of Theorem 1.

Figure 1: Base case: $\tau = 0.50$. Upper and lower bounding probabilities for ordered probit structural functions with $\tau = 0.75$, $a_0 = a_2 = 1$, and a_2 varying as shown. Parameter values in the structure generating the probability distributions used in the calculations are $\alpha_0 = \alpha_1 = \alpha_2 = 1$, $\beta_0 = 0$, $\beta_1 = 1$, $\sigma_{wv} = 0.6$, $\sigma_{vv} = 1$, $M = 4$ points of support.

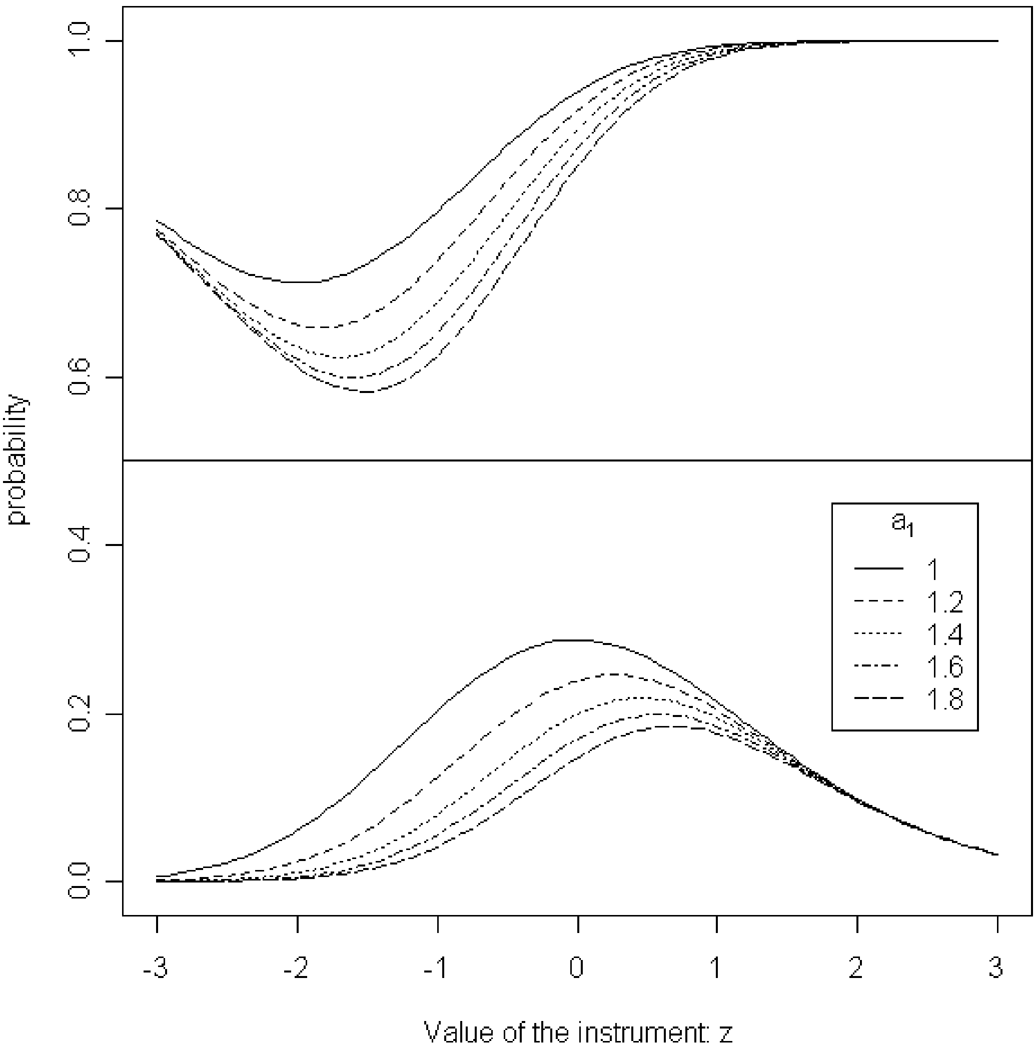

Figure 2: Base case: $\tau = 0.75$. Upper and lower bounding probabilities for ordered probit structural functions with $\tau = 0.75$, $a_0 = a_2 = 1$, and a_2 varying as shown. Parameter values in the structure generating the probability distributions used in the calculations are $\alpha_0 = \alpha_1 = \alpha_2 = 1$, $\beta_0 = 0$, $\beta_1 = 1$, $\sigma_{wv} = 0.6$, $\sigma_{vv} = 1$, $M = 4$ points of support.

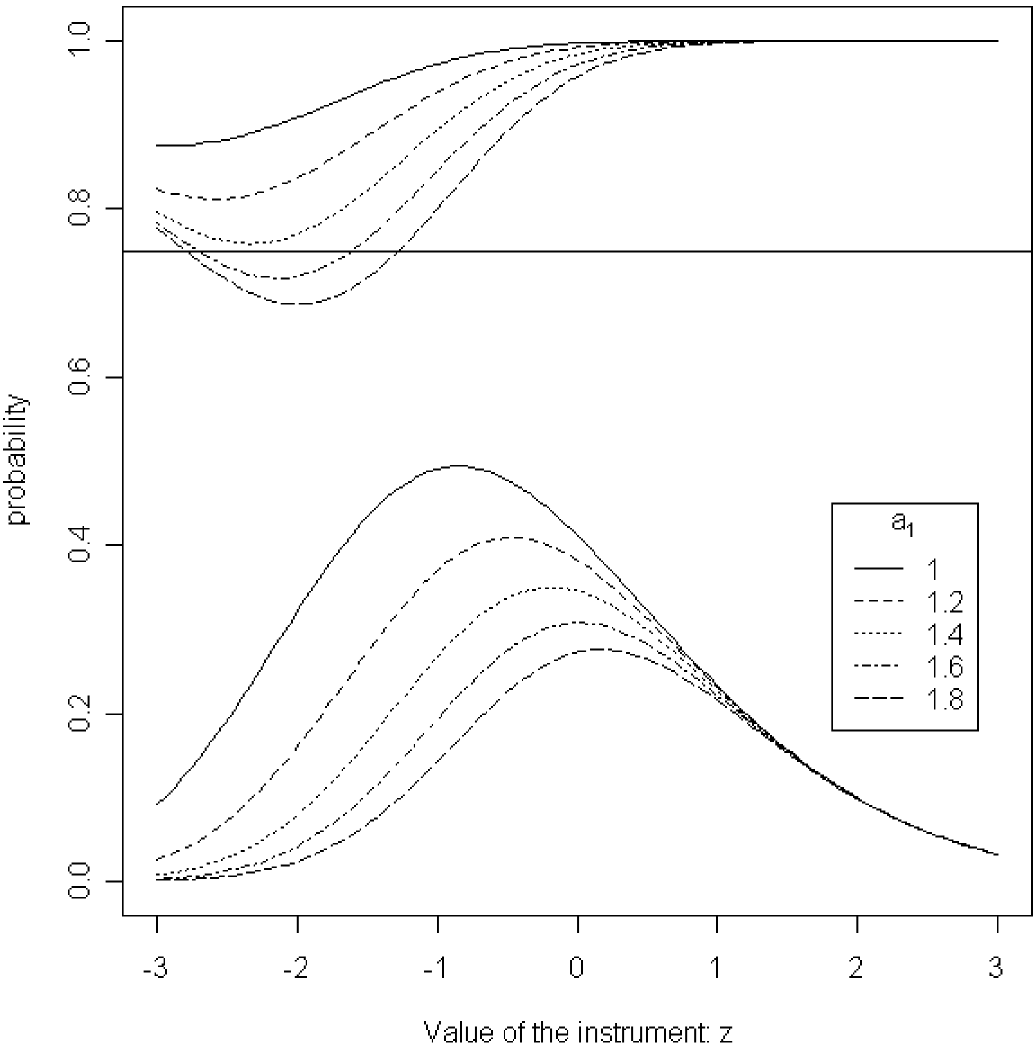

Figure 3: Strong Instrument. Upper and lower bounding probabilities for ordered probit structural functions with $\tau = 0.50$, $a_0 = a_2 = 1$, and a_2 varying as shown. Parameter values in the structure generating the probability distributions used in the calculations are $\alpha_0 = \alpha_1 = \alpha_2 = 1$, $\beta_0 = 0$, $\beta_1 = 1$, $\sigma_{wv} = 0.6$, $\sigma_{vv} = 0.48$, $M = 4$ points of support.

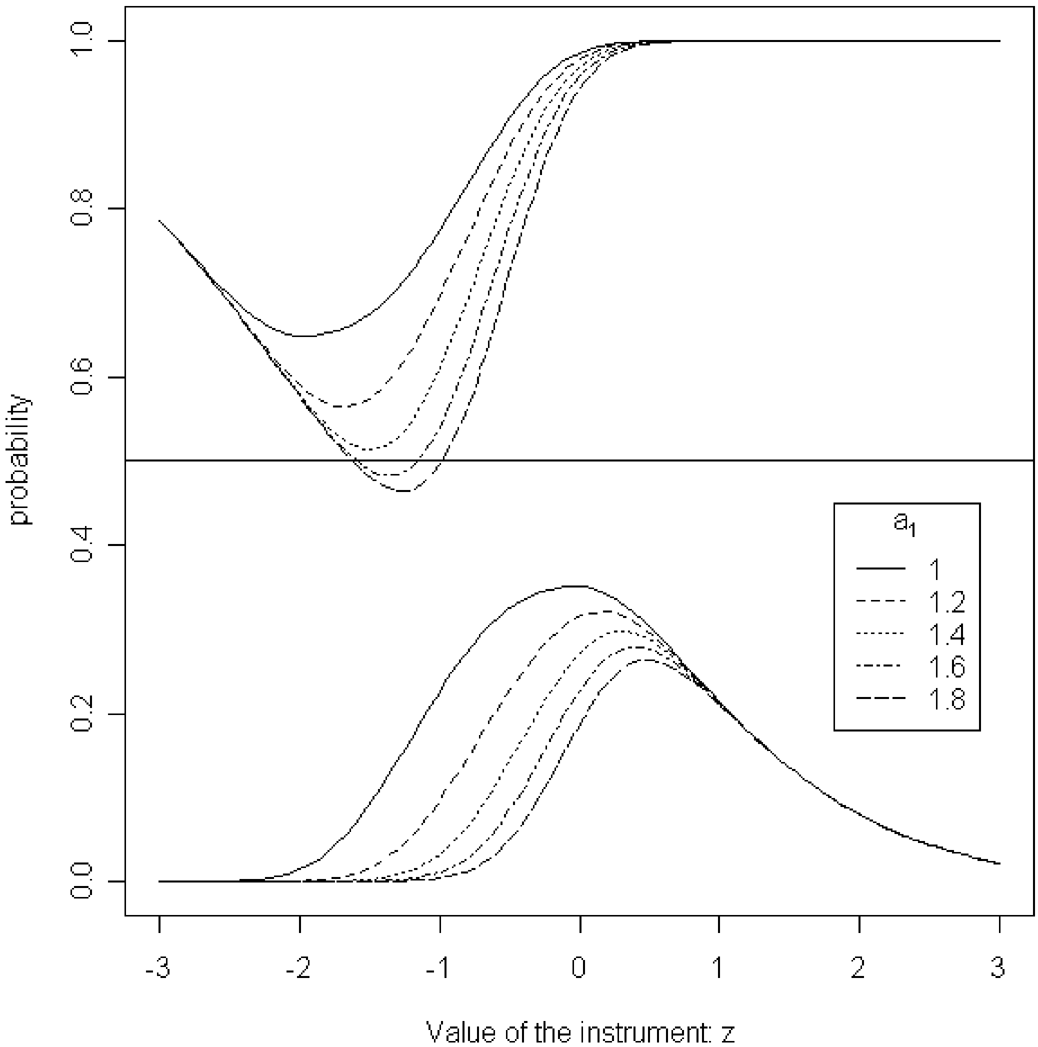

Figure 4: Less discrete $M = 10$. Upper and lower bounding probabilities for ordered probit structural functions with $\tau = 0.50$, $a_0 = a_2 = 1$, and a_2 varying as shown. Parameter values in the structure generating the probability distributions used in the calculations are $\alpha_0 = \alpha_1 = \alpha_2 = 1$, $\beta_0 = 0$, $\beta_1 = 1$, $\sigma_{wv} = 0.6$, $\sigma_{vv} = 1$, $M = 10$ points of support.

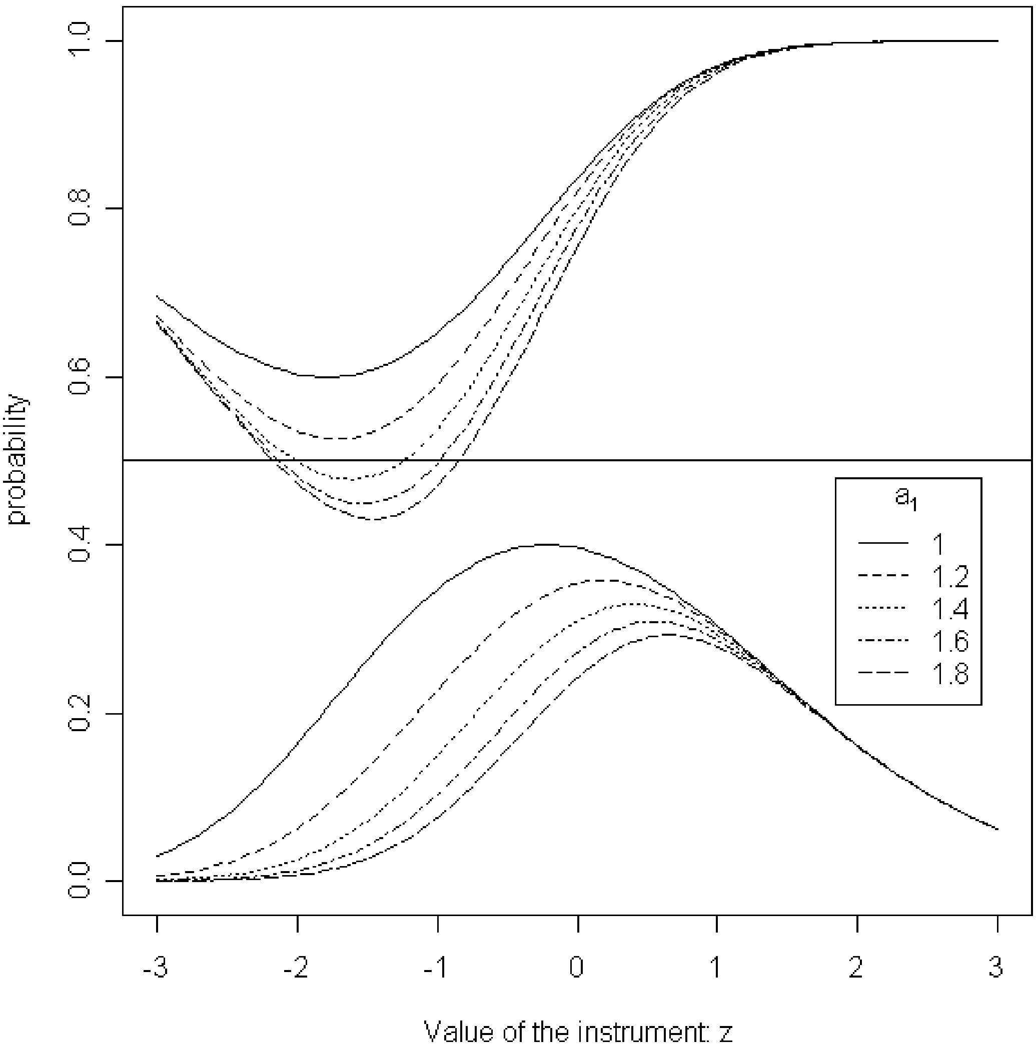

Figure 5: Base case: $\tau = 0.50$. Upper and lower bounding probabilities for ordered probit structural functions with $\tau = 0.75$, $a_0 = a_2 = 1$, and a_2 varying as shown. Parameter values in the structure generating the probability distributions used in the calculations are $\alpha_0 = \alpha_1 = \alpha_2 = 1$, $\beta_0 = 0$, $\beta_1 = 1$, $\sigma_{wv} = 0.6$, $\sigma_{vv} = 1$, $M = 4$ points of support.

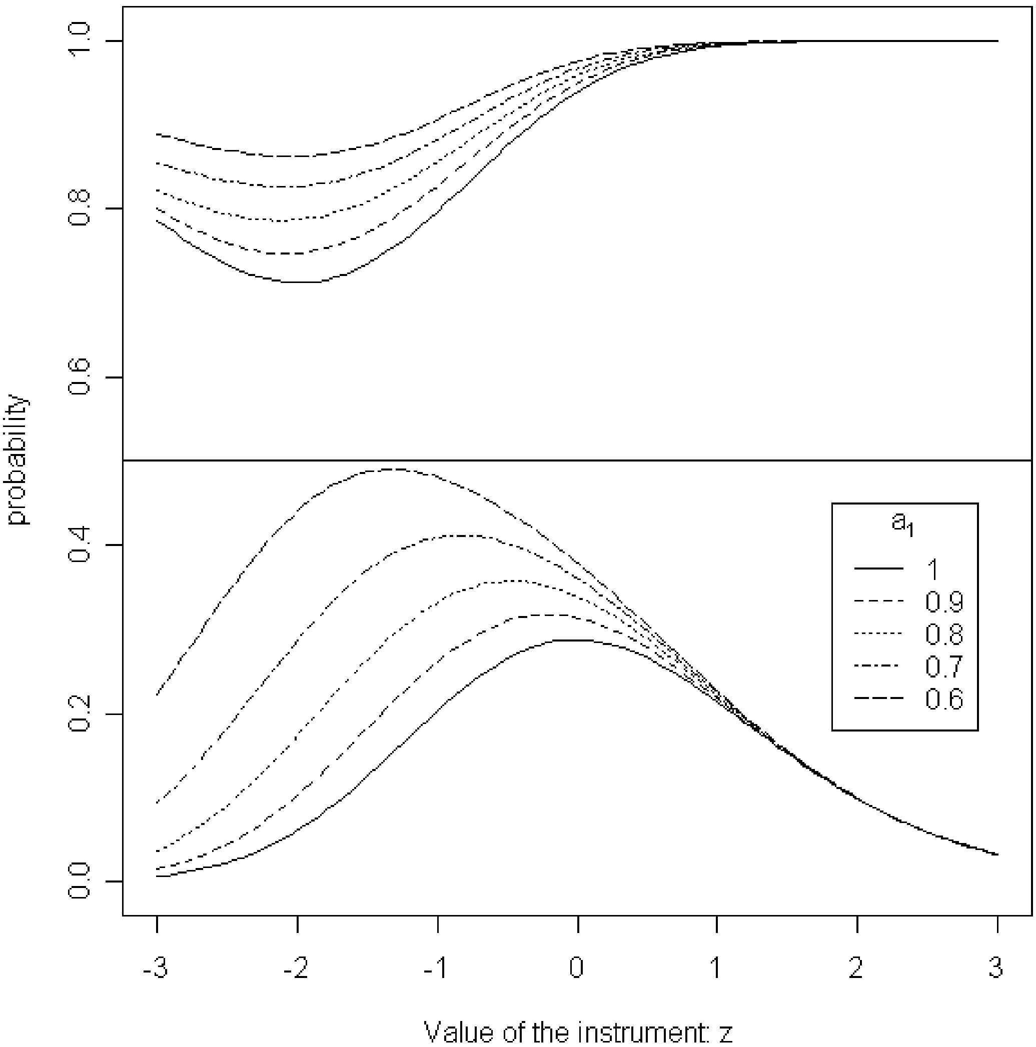

Figure 6: Base case: $\tau = 0.75$. Upper and lower bounding probabilities for ordered probit structural functions with $\tau = 0.75$, $a_0 = a_2 = 1$, and a_2 varying as shown. Parameter values in the structure generating the probability distributions used in the calculations are $\alpha_0 = \alpha_1 = \alpha_2 = 1$, $\beta_0 = 0$, $\beta_1 = 1$, $\sigma_{wv} = 0.6$, $\sigma_{vv} = 1$, $M = 4$ points of support.

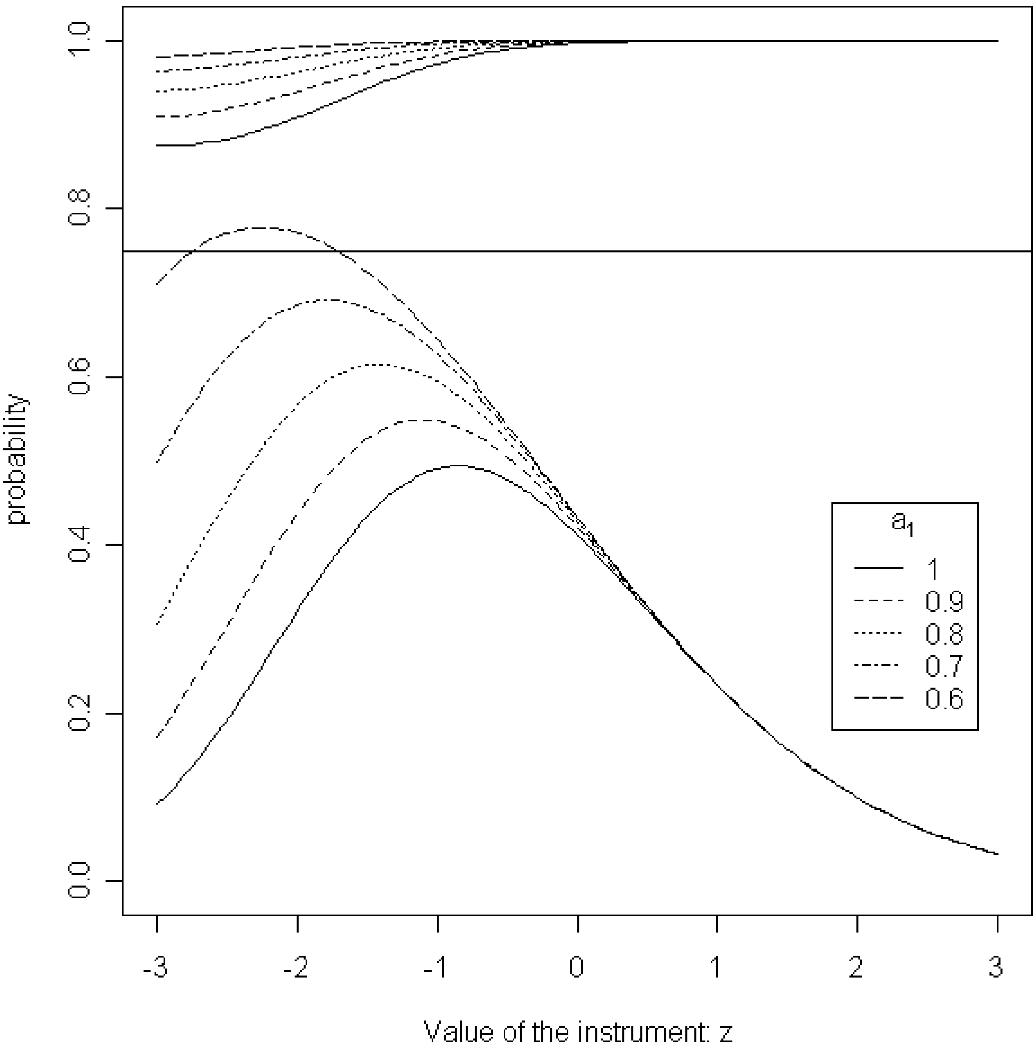

Figure 7: Strong Instrument. Upper and lower bounding probabilities for ordered probit structural functions with $\tau = 0.50$, $a_0 = a_2 = 1$, and a_2 varying as shown. Parameter values in the structure generating the probability distributions used in the calculations are $\alpha_0 = \alpha_1 = \alpha_2 = 1$, $\beta_0 = 0$, $\beta_1 = 1$, $\sigma_{wv} = 0.6$, $\sigma_{vv} = 0.48$, $M = 4$ points of support.

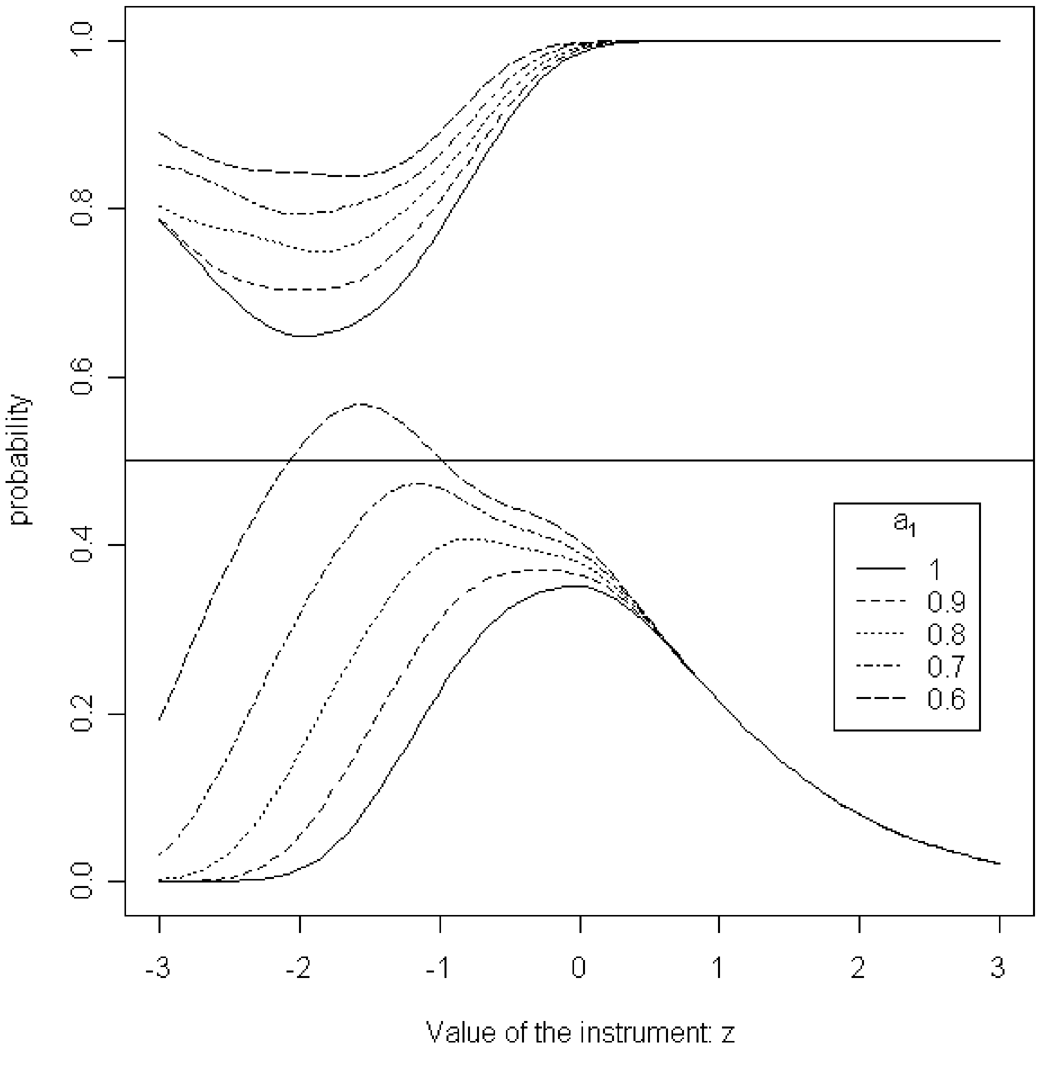

Figure 8: Less discrete $M = 10$. Upper and lower bounding probabilities for ordered probit structural functions with $\tau = 0.50$, $a_0 = a_2 = 1$, and a_2 varying as shown. Parameter values in the structure generating the probability distributions used in the calculations are $\alpha_0 = \alpha_1 = \alpha_2 = 1$, $\beta_0 = 0$, $\beta_1 = 1$, $\sigma_{wv} = 0.6$, $\sigma_{vv} = 1$, $M = 10$ points of support.

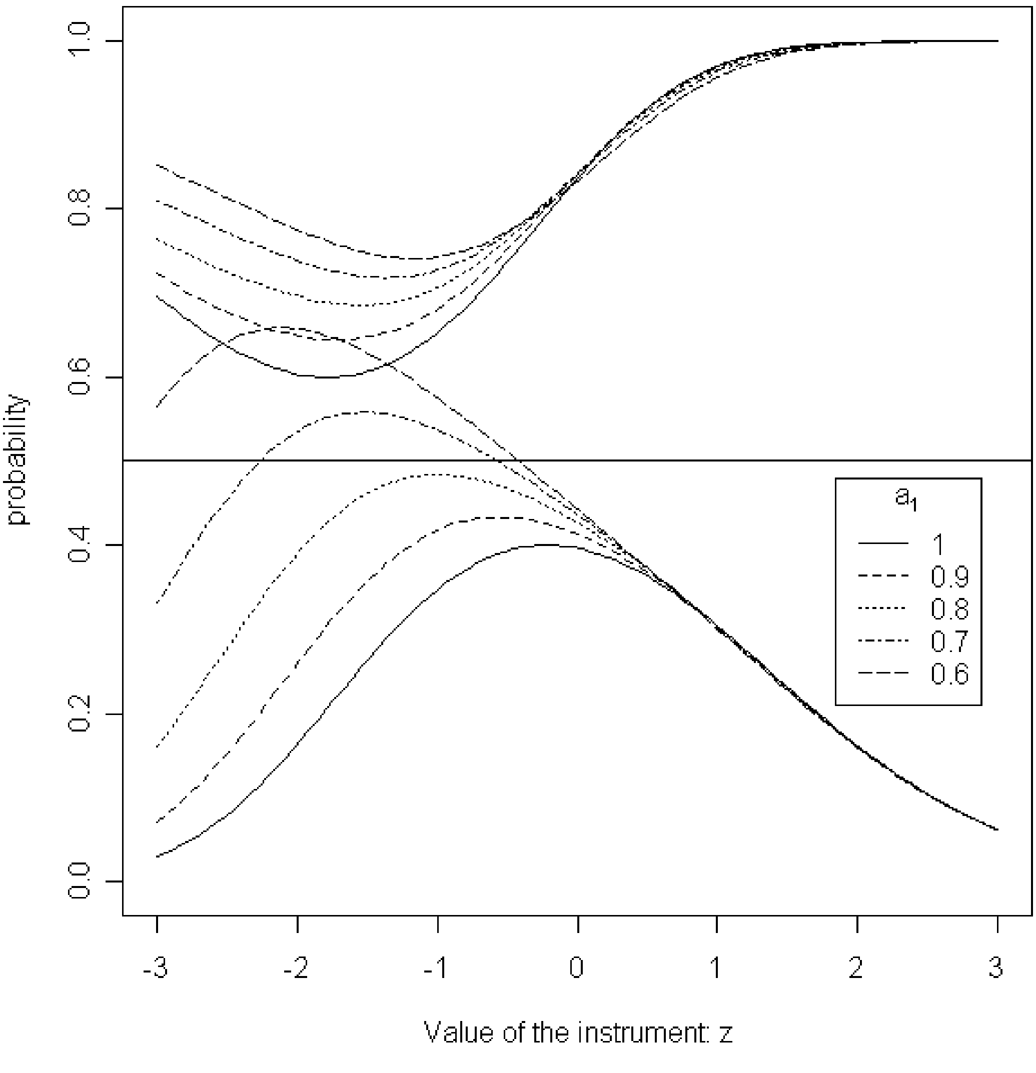

Figure 9: Upper and lower bounding probabilities for Poisson structural functions with $\tau = 0.50$, $a_0 = -0.5$, and a_1 varying as shown. Parameter values in the structure generating the probability distributions used in the calculations are $\alpha_0 = -0.5$, $\alpha_1 = 0.5$, $\beta_0 = 0$, $\beta_1 = 1$, $\sigma_{wv} = 0.6$, $\sigma_{vv} = 1$.

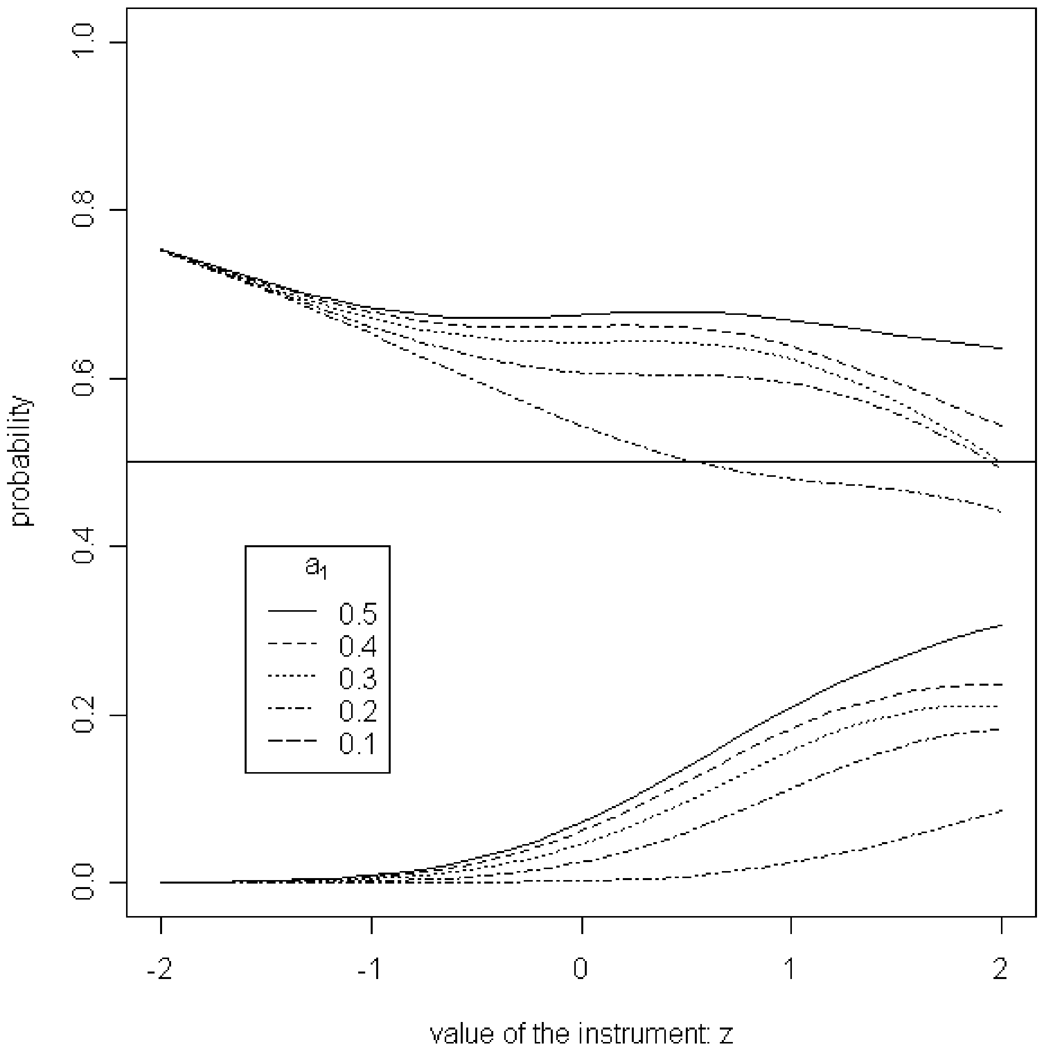

Figure 10: Upper and lower bounding probabilities for Poisson structural functions with $\tau = 0.50$, $a_0 = -0.5$, and a_1 varying as shown. Parameter values in the structure generating the probability distributions used in the calculations are $\alpha_0 = -0.5$, $\alpha_1 = 0.5$, $\beta_0 = 0$, $\beta_1 = 1$, $\sigma_{wv} = 0.6$, $\sigma_{vv} = 1$.

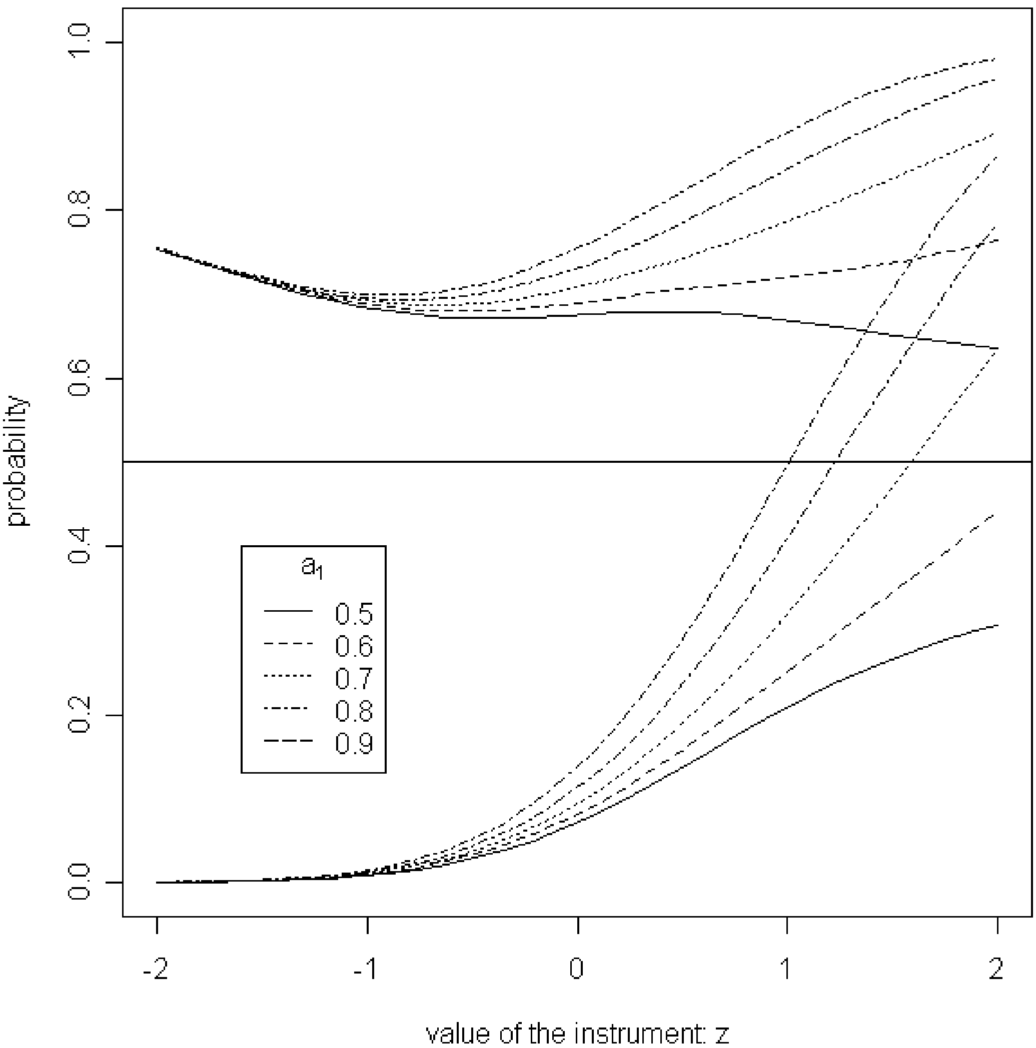

Figure 11: $N = 1$: binary logit. Upper and lower bounding probabilities for binomial structural functions with $\tau = 0.50$, $a_0 = 0.75$, and a_1 varying as shown. Parameter values in the structure generating the probability distributions used in the calculations are $\alpha_0 = 0.75$, $\alpha_1 = 1.0$, $\beta_0 = 0$, $\beta_1 = 1$, $\sigma_{wv} = 0.6$, $\sigma_{vv} = 1$.

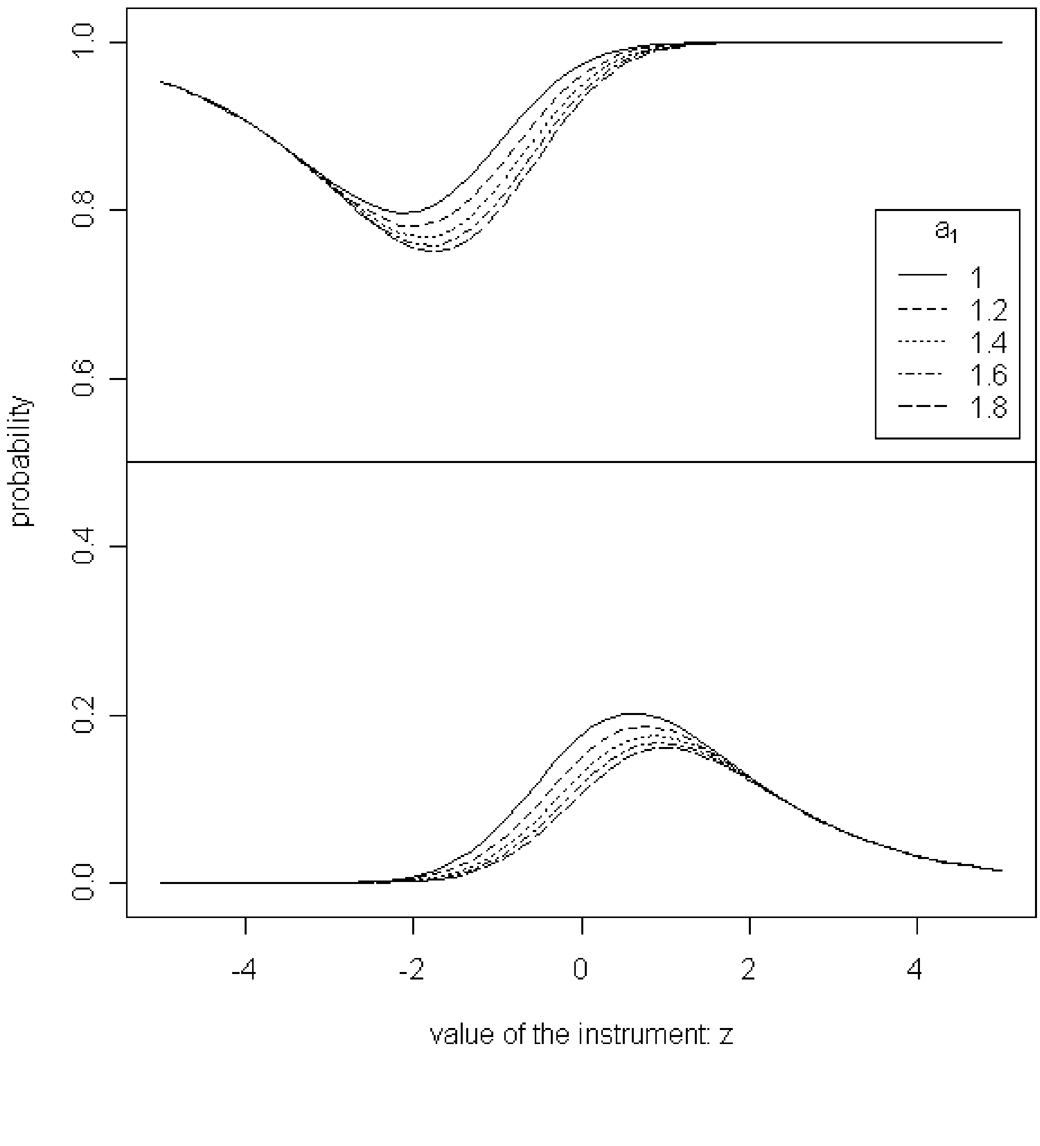

Figure 12: $N = 1$: binary logit. Upper and lower bounding probabilities for binomial structural functions with $\tau = 0.50$, $a_0 = 0.75$, and a_1 varying as shown. Parameter values in the structure generating the probability distributions used in the calculations are $\alpha_0 = 0.75$, $\alpha_1 = 1.0$, $\beta_0 = 0$, $\beta_1 = 1$, $\sigma_{wv} = 0.6$, $\sigma_{vv} = 1$.

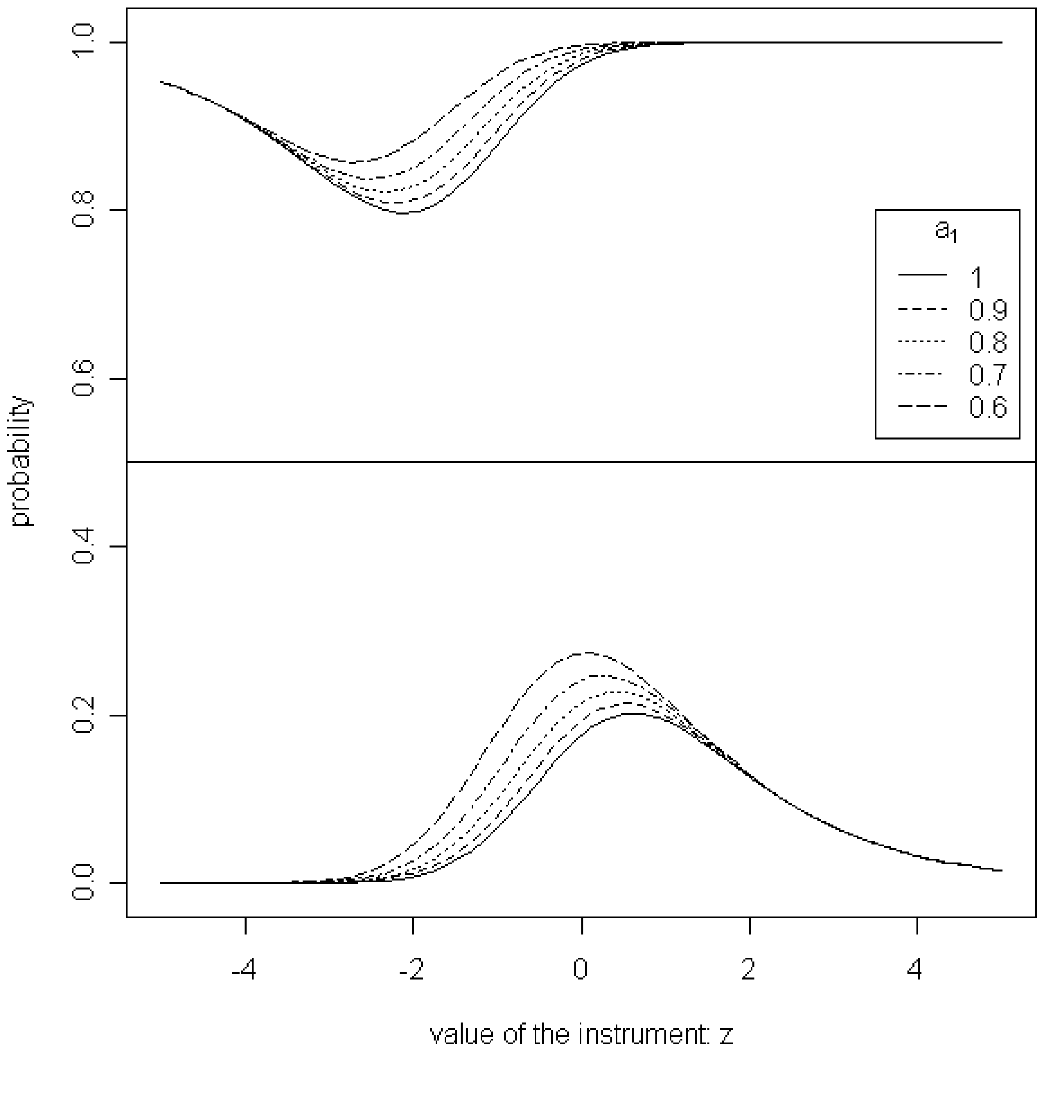

Figure 13: $N = 3$. Upper and lower bounding probabilities for binomial structural functions with $\tau = 0.50$, $a_0 = 0.75$, and a_1 varying as shown. Parameter values in the structure generating the probability distributions used in the calculations are $\alpha_0 = 0.75$, $\alpha_1 = 1.0$, $\beta_0 = 0$, $\beta_1 = 1$, $\sigma_{wv} = 0.6$, $\sigma_{vv} = 1$.

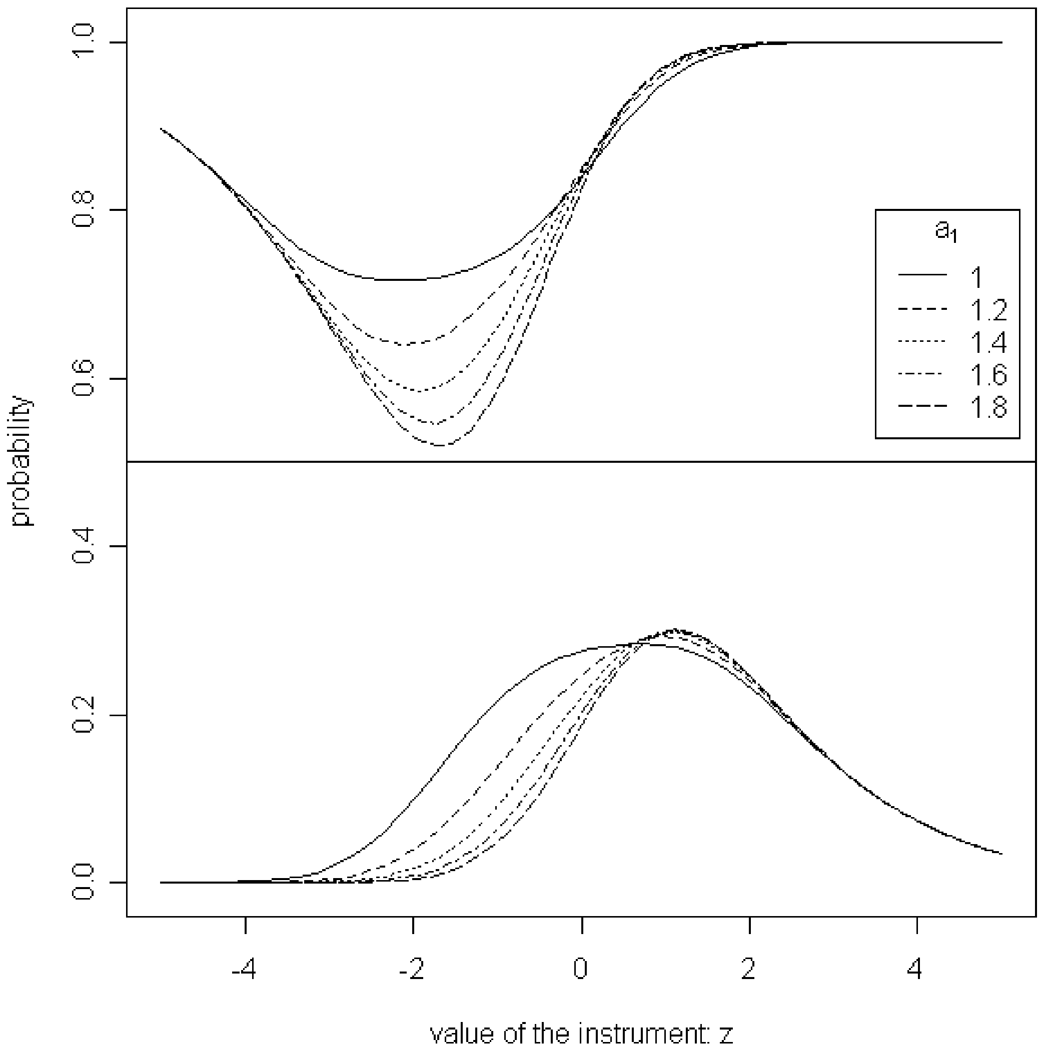

Figure 14: $N = 3$. Upper and lower bounding probabilities for binomial structural functions with $\tau = 0.50$, $a_0 = 0.75$, and a_1 varying as shown. Parameter values in the structure generating the probability distributions used in the calculations are $\alpha_0 = 0.75$, $\alpha_1 = 1.0$, $\beta_0 = 0$, $\beta_1 = 1$, $\sigma_{wv} = 0.6$, $\sigma_{vv} = 1$.

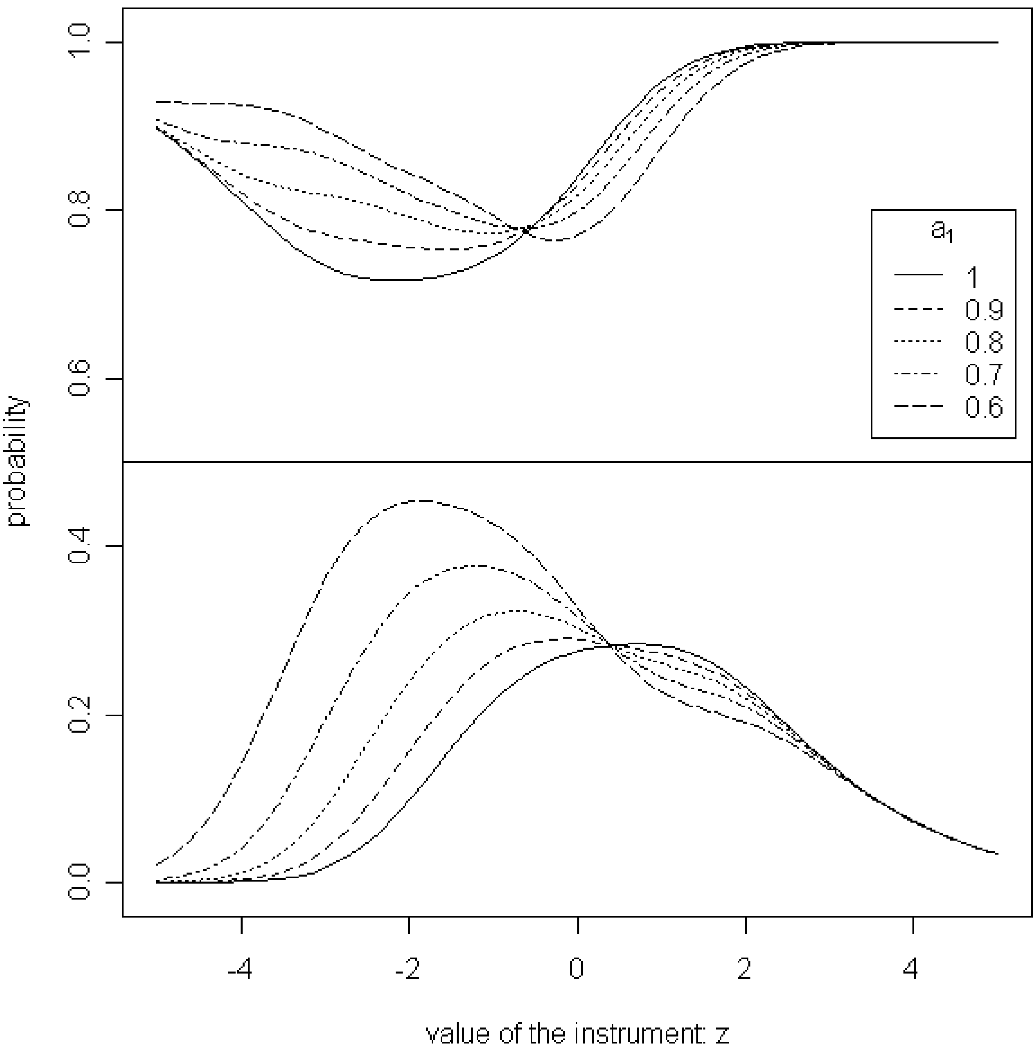

Figure 15: $N = 6$. Upper and lower bounding probabilities for binomial structural functions with $\tau = 0.50$, $a_0 = 0.75$, and a_1 varying as shown. Parameter values in the structure generating the probability distributions used in the calculations are $\alpha_0 = 0.75$, $\alpha_1 = 1.0$, $\beta_0 = 0$, $\beta_1 = 1$, $\sigma_{wv} = 0.6$, $\sigma_{vv} = 1$.

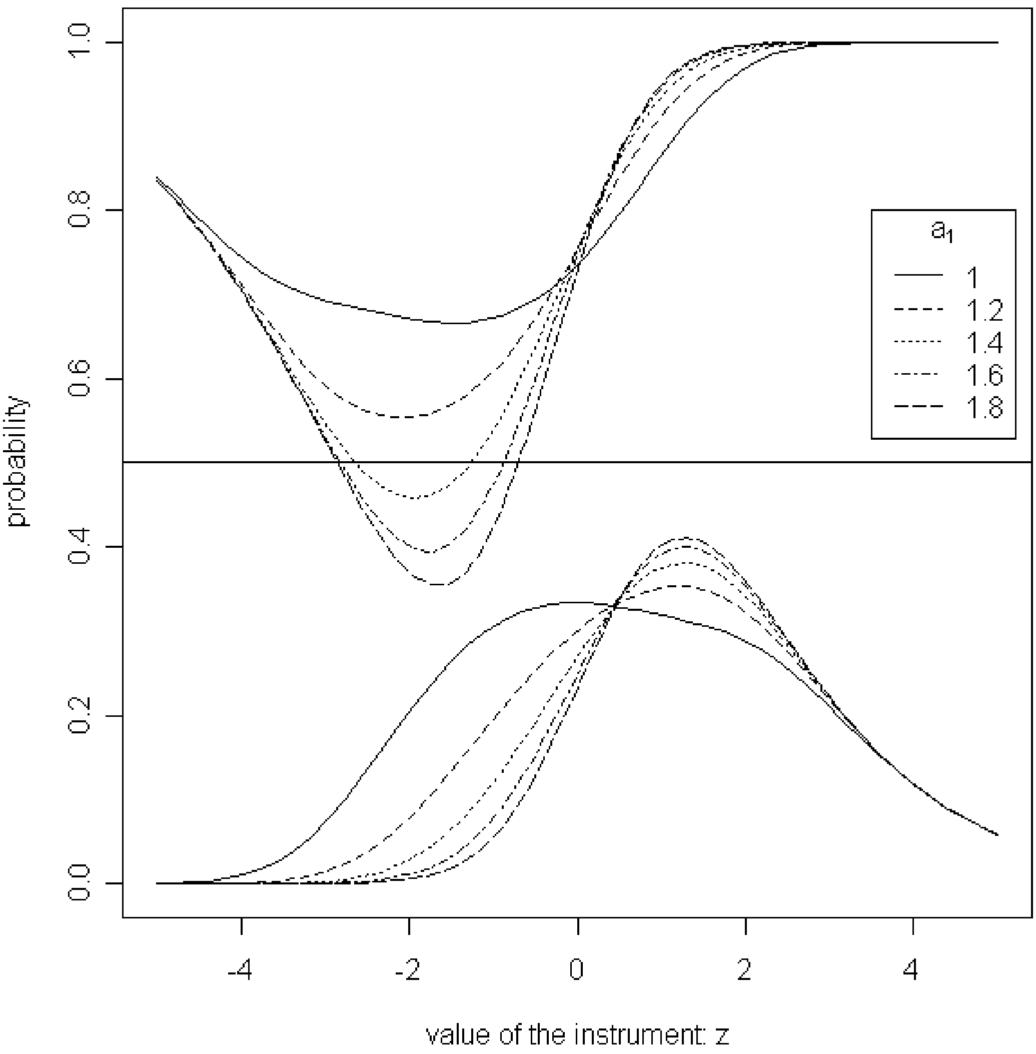

Figure 16: $N = 6$. Upper and lower bounding probabilities for binomial structural functions with $\tau = 0.50$, $a_0 = 0.75$, and a_1 varying as shown. Parameter values in the structure generating the probability distributions used in the calculations are $\alpha_0 = 0.75$, $\alpha_1 = 1.0$, $\beta_0 = 0$, $\beta_1 = 1$, $\sigma_{wv} = 0.6$, $\sigma_{vv} = 1$.