

Instrumental variable models for discrete outcomes

Andrew Chesher

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP30/08



An ESRC Research Centre

Instrumental Variable Models for Discrete Outcomes

ANDREW CHESHER*

CENTRE FOR MICRODATA METHODS AND PRACTICE AND UCL

Revised November 19th 2008

ABSTRACT. Single equation instrumental variable models for discrete outcomes are shown to be set not point identifying for the structural functions that deliver the values of the discrete outcome. Identified sets are derived for a general nonparametric model and sharp set identification is demonstrated. Point identification is typically not achieved by imposing parametric restrictions. The extent of an identified set varies with the strength and support of instruments and typically shrinks 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.

KEYWORDS: Partial identification, Nonparametric methods, Nonadditive models, Discrete distributions, Ordered choice, Endogeneity, Instrumental variables, Structural quantile functions, Incomplete models.

JEL CODES: C10, C14, C50, C51.

1. INTRODUCTION

This paper gives results on the identifying power of single equation instrumental variables (IV) models for a discrete outcome, Y , in which explanatory variables, X , may be endogenous. Outcomes can be binary, for example indicating the occurrence of an event; integer valued - for example recording counts of events; or ordered - for example giving a point on an attitudinal scale or obtained by interval censoring of an unobserved continuous outcome. Endogenous and other observed variables can be continuous or discrete.

The scalar discrete outcome Y is determined by a structural function thus:

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

and it is identification of the function h that is studied. Here X is a vector of possibly endogenous variables, U is a scalar continuously distributed unobservable random variable, normalised marginally uniformly distributed on the unit interval and h is restricted to be weakly monotonic, normalised non-decreasing in U . There are instrumental variables, Z , excluded from the structural function h , and U is distributed independently of Z for Z lying in a set Ω_Z . X may be endogenous in the sense that U and X may not be independently distributed. This is a single equation model in the sense that there is no specification of structural equations determining the value of X . In this respect the model is incomplete.

*Department of Economics, University College London, Gower Street, London WC1E 6BT, UK. Telephone: +442076795857. Email: andrew.chesher@ucl.ac.uk.

This single equation IV model is shown to *set identify* the structural function h . Identified sets are characterised and illustrated *via* examples. Identification is shown to be sharp in the sense that every structural function in an identified set is capable of delivering the probability distribution of observables used to calculate the identified set. Additional parametric restrictions do not typically secure point identification although they may reduce the extent of identified sets.

Underpinning the identification results is the following pair of inequalities which hold for all $\tau \in (0, 1)$ and all $z \in \Omega_Z$ under a probability measure generated by a structure admitted by the single equation IV model with structural function h .

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

It is shown that these inequalities sharply define the identified set for the probability measure used in these probability calculations in the sense that all functions h , and only functions h , that satisfy these inequalities for all $\tau \in (0, 1)$ and all $z \in \Omega_Z$ are elements of the observationally equivalent structures which generate that probability measure.

When the outcome Y is continuously distributed (in which case h is strictly monotonic in U) both probabilities in (1) are equal to τ and with additional completeness restrictions, the model point identifies the structural function as set out in Chernozhukov and Hansen (2005) where the function h is called a *structural quantile function*. This paper extends the analysis of structural quantile functions to cases in which outcomes are discrete.

Many applied researchers facing discrete outcome and endogenous explanatory variables using a control function approach. These are rooted in a more restrictive model which can be point identifying but the model's restrictions are not always applicable. There is a brief discussion in Section 4.

There are a few papers that take a single equation IV approach to endogeneity in parametric count data models, basing identification on moment conditions.¹ Mullahy (1997) and Windmeijer and Santos Silva (1997) consider a model in which the conditional expectation of a count variable given explanatory variables, $X = x$, and an unobserved scalar heterogeneity term, $V = v$, is *multiplicative*: $\exp(x\beta) \times v$, with X and V correlated and with V and instrumental variables Z having a degree of independent variation. This IV model can point identify β but the fine details of the functional form restrictions are influential in securing point identification and the method, based as it is on a multiplicative heterogeneity specification, does not apply when discrete variables have bounded support.

The paper is organised as follows. The main results of the paper are given in Section 2 which specifies an IV model for a discrete outcome and presents and discusses the set identification results. Section 3 presents two illustrative examples; one is very simple with a binary outcome and a binary endogenous variable and the other involves a parametric ordered-probit-type problem. Section 4 discusses alternatives to the set identifying single equation IV model and outlines some extensions including the case arising with panel data when there is a vector of discrete outcomes.

¹See the discussion in Section 11.3.2 of Cameron and Trivedi (1988).

2. IV MODELS AND THEIR IDENTIFYING POWER

This Section presents the main results of the paper. Section 2.1 defines a single equation instrumental variable model for a discrete outcome and develops the probability inequalities which play a key role in defining the identified set of structural functions. In Section 2.2 theorems are presented which characterise the set of structural functions identified by the IV model and state that identification thus achieved is sharp. Section 2.3 discusses the set identification result with brief comments on: the impact of support and strength of instruments and discreteness of outcome on the identified set, local independence restrictions, and estimation and inference.

2.1. Models. The following two restrictions define a model, \mathcal{D} , for a scalar discrete outcome.

- D1. $Y = h(X, U)$ where $U \in (0, 1)$ is continuously distributed and h is weakly monotonic (normalized càglàd, non-decreasing) in its last argument. X is a vector of explanatory variables. The codomain of h is the ascending sequence $\{y_m\}_{m=1}^M$ which is independent of X . M may be unbounded. The function h is normalised so that the marginal distribution of U is uniform.
- D2. There exists a vector Z such that $\Pr[U \leq \tau | Z = z] = \tau$ for all $\tau \in (0, 1)$ and all $z \in \Omega_Z$.

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

$$\text{for } m \in \{1, \dots, M\}: \quad h(x, u) = y_m \text{ if } p_{m-1}(x) < u \leq p_m(x) \quad (2)$$

with, for all x , $p_0(x) \equiv 0$ and $p_M(x) \equiv 1$. The structural function, h , is a non-decreasing step function, the value of Y incrementing as U passes through thresholds which depend on the value of the explanatory variables, X .

Restriction D2 requires that the conditional distribution of U given $Z = z$ be invariant with respect to z for variations within Ω_Z . If Z is a random variable and Ω_Z is its support then the model requires that U and Z be independently distributed. But Z is not required to be a random variable. For example values of Z might be chosen purposively, for example by an experimenter, and then Ω_Z is some set of values of Z that can be chosen.

Restriction D1 excludes the variables Z from the structural function h . These variables play the role of instrumental variables with the potential for contributing to the identifying power of the model if they are indeed “instrumental” in determining the value of the endogenous X . But the model \mathcal{D} places no restrictions on the way in which the variables X , possibly endogenous, are generated.

Data are informative about the conditional distribution function of (Y, X) given Z for $Z = z \in \Omega_Z$, denoted by $F_{YX|Z}(y, x|z)$. Let $F_{UX|Z}$ denote the joint distribution function of U and X given Z . Under the weak monotonicity condition embodied in the model \mathcal{D} an admissible structure $S^a \equiv \{h^a, F_{UX|Z}^a\}$ with structural function h^a

delivers a conditional distribution for (Y, X) given Z , $F_{YX|Z}^a$, as follows.

$$F_{YX|Z}^a(y_m, x|z) = F_{UX|Z}^a(p_m^a(x), x|z), \quad m \in \{1, \dots, M\} \quad (3)$$

Here the functions $\{p_m^a(x)\}_{m=0}^M$ are the threshold functions that characterize the structural function h^a as in (2) above.

Distinct structures admitted by the model \mathcal{D} can deliver identical distributions of Y and X given Z for all $z \in \Omega_Z$. Such structures are *observationally equivalent* and the model is set, not point, identifying because within a set of admissible observationally equivalent structures there can be more than one distinct structural function. This can happen because on the right hand side of (3) certain variations in the functions $p_m^a(x)$ can be offset by altering the sensitivity of $F_{UX|Z}^a(u, x|z)$ to variations in x so that the left hand side of (3) is left unchanged.

Crucially the independence restriction D2 places limits on the variations in the functions $p_m^a(x)$ that can be so compensated and results in the model having *nontrivial set identifying power*. The key to defining the sets of structural functions identified by the model lies in a pair of probability inequalities which are the subject of the following Theorem.

Theorem 1. *Let $S_a \equiv \{h^a, F_{UX|Z}^a\}$ be a structure admitted by the model \mathcal{D} delivering a distribution function for (Y, X) given Z , $F_{YX|Z}^a$, and let \Pr_a indicate probabilities calculated using this distribution. The following inequalities hold.*

$$\text{For all } z \in \Omega_Z \text{ and } \tau \in (0, 1): \begin{cases} \Pr_a[Y \leq h^a(X, \tau)|Z = z] \geq \tau \\ \Pr_a[Y < h^a(X, \tau)|Z = z] < \tau \end{cases} \quad (4)$$

The proof is straightforward given the following Lemma which is proved in the Annex.

Lemma 1. *Let $Y = h(X, U)$ with U , X and h satisfying Restriction D1. The following inequalities hold for all $\tau \in (0, 1)$ and for all x and z .*

$$\begin{aligned} \Pr[Y \leq h(X, \tau)|X = x, Z = z] &\geq F_{U|XZ}(\tau|x, z) \\ \Pr[Y < h(X, \tau)|X = x, Z = z] &< F_{U|XZ}(\tau|x, z) \end{aligned} \quad (5)$$

Proof of Theorem 1. *Applying Lemma 1 to each structure S_a considered in Theorem 1 gives the following inequalities which hold for all $\tau \in (0, 1)$ and for all x and z .*

$$\begin{aligned} \Pr_a[Y \leq h^a(X, \tau)|X = x, Z = z] &\geq F_{U|XZ}^a(\tau|x, z) \\ \Pr_a[Y < h^a(X, \tau)|X = x, Z = z] &< F_{U|XZ}^a(\tau|x, z) \end{aligned}$$

Let $F_{X|Z}^a$ be the distribution function of X given Z associated with $F_{YX|Z}^a$. Using this distribution to take expectations over X given $Z = z$ on the left hand sides of these inequalities delivers the left hand sides of the inequalities (4). Taking expectations similarly on the right hand sides yields the distribution function of U given $Z = z$

associated with $F_{U|Z}^a(\tau|z)$ which is equal to τ for all $z \in \Omega_Z$ and $\tau \in (0, 1)$ under the conditions of model \mathcal{D} .

2.2. Identification. Consider the model \mathcal{D} , a structure S_a admitted by it, and the set \tilde{S}_a of all structures admitted by \mathcal{D} and observationally equivalent to S_a . Let $\tilde{\mathcal{H}}_a$ be the set of structural functions which are components of structures contained in \tilde{S}_a . Let $F_{YX|Z}^a$ be the joint distribution function of (Y, X) given Z delivered by the observationally equivalent structures in the set \tilde{S}_a .

The model \mathcal{D} set identifies the structural function generating $F_{YX|Z}^a$ - it must be one of the structural functions in the set $\tilde{\mathcal{H}}_a$. The inequalities (4) characterize this set as follows: the set of structural functions $\tilde{\mathcal{H}}_a$ is precisely the set of functions which satisfy the inequalities (4) when probabilities are calculated using distribution $F_{YX|Z}^a$. No structural function lying outside this set can be a component of an admissible structure observationally equivalent to S_a . Every structural function lying in this set is a component of an admissible structure observationally equivalent to S_a . There is the following Theorem.

Theorem 2. *Let S_a be a structure admitted by the model \mathcal{D} and delivering the distribution function $F_{YX|Z}^a$. Let $S_* \equiv \{h_*, F_{UX|Z}^*\}$ be any observationally equivalent structure admitted by the model \mathcal{D} . Let \Pr_a indicate probabilities calculated using the distribution function $F_{YX|Z}^a$. The following inequalities are satisfied.*

$$\text{For all } z \in \Omega_Z \text{ and } \tau \in (0, 1): \quad \begin{cases} \Pr_a[Y \leq h^*(X, \tau)|Z = z] \geq \tau \\ \Pr_a[Y < h^*(X, \tau)|Z = z] < \tau \end{cases} \quad (6)$$

Proof of Theorem 2. *Let \Pr_* indicates probabilities calculated using $F_{YX|Z}^*$. Because the structure S_* is admitted by model \mathcal{D} , Theorem 1 implies that for all $z \in \Omega_Z$ and $\tau \in (0, 1)$:*

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

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

Since S_a and S_ are observationally equivalent, $F_{YX|Z}^* = F_{YX|Z}^a$ and the inequalities (6) follow on substituting “ \Pr_a ” for “ \Pr_* ”.*

There is the following Corollary whose proof, which is elementary, is omitted.

Corollary. *Let \Pr_a indicate probabilities calculated using a distribution function $F_{YX|Z}^a$ generated by a structure S_a admitted by the model \mathcal{D} . If the inequalities (6) are violated for any $(z, \tau) \in \Omega_Z \times (0, 1)$ then $h^* \notin \tilde{\mathcal{H}}_a$.*

The consequence of these results is that for any probability measure $F_{YX|Z}^a$ generated by an admissible structure the set of functions that satisfy the inequalities (6) contains all members of the set of structural functions $\tilde{\mathcal{H}}_a$ identified by the model \mathcal{D} . In fact the sets are identical, a sharpness result which follows from the following Theorem.

Theorem 3 *If $h^*(x, u)$ satisfies the restrictions of the model \mathcal{D} and the inequalities (6) then there exists a proper distribution function $F_{UX|Z}^*$ such that $S^* = \{h^*, F_{UX|Z}^*\}$ satisfies the restrictions of model \mathcal{D} and is observationally equivalent to structures S^a that generate the distribution $F_{YX|Z}^a$.*

A proof of Theorem 3 is given in the Annex. The proof is constructive. For a given distribution $F_{YX|Z}^a$ and each value of $z \in \Omega_Z$ and each structural function h^* satisfying the inequalities (6) a proper distribution function $F_{UX|Z}^*$ is constructed which respects the independence condition of Restriction D2 and has the property that at the chosen value of z the pair $\{h^*, F_{UX|Z}^*\}$ deliver the distribution function $F_{YX|Z}^a$ at that value of z .

2.3. Discussion.

2.3.1. Intersection bounds and sharpness. For a distribution, $F_{YX|Z}^a$, of Y and X given Z and a set, Ω_Z , of values of the instrumental variables the set of structural functions identified by the single equation IV model \mathcal{D} is the set of structural functions h^* that satisfy the inequalities (6) for all $z \in \Omega_Z$ and all $\tau \in (0, 1)$.

Let $\tilde{\mathcal{H}}_a(z)$ be the set of structural functions satisfying the inequalities (6) for all $\tau \in (0, 1)$ at a value $z \in \Omega_Z$. Because for each τ the inequalities must hold for all $z \in \Omega_Z$, the identified set, $\tilde{\mathcal{H}}_a$, associated with structures S^a that generate the distribution $F_{YX|Z}^a$ is the intersection of the sets $\tilde{\mathcal{H}}_a(z)$ for $z \in \Omega_Z$ which can be expressed as follows.

$$\tilde{\mathcal{H}}_a = \left\{ h^* : \text{for all } \tau \in (0, 1) \left(\begin{array}{l} \min_{z \in \Omega_Z} \Pr_a[Y \leq h^*(X, \tau) | Z = z] \geq \tau \\ \max_{z \in \Omega_Z} \Pr_a[Y < h^*(X, \tau) | Z = z] < \tau \end{array} \right) \right\} \quad (7)$$

Theorem 3 shows that for each z in Ω_Z and for each h^* in a set $\tilde{\mathcal{H}}_a(z)$ there exists an admissible distribution $F_{UX|Z}^*(u, x|z)$ such that h^* together with that distribution delivers the conditional distribution $F_{YX|Z}^a(y, x|z)$. It follows that for every h^* in the intersection $\tilde{\mathcal{H}}_a$ there is an admissible distribution $F_{UX|Z}^*$ which together with h^* delivers the conditional distribution $F_{YX|Z}^a$ for all values of Z in Ω_Z .

In particular cases the set $\tilde{\mathcal{H}}_a$ can be enumerated or otherwise characterised. Two examples are given in Section 3, one of them involving parametric restrictions. Parametric restrictions on structural functions reduce the identified set to a subset of $\tilde{\mathcal{H}}_a$ but do not in general produce a point identifying model.

2.3.2. Strength and support of instruments. It is clear from (7) that the support of the instrumental variables, Ω_Z , is critical in determining the extent of an identified set. The strength of the instruments is also critical.

When instrumental variables are good predictors of some particular value of the endogenous variables, say x^* , the identified sets for the values of threshold crossing functions at $X = x^*$ will tend to be small in extent. In the extreme case of perfect prediction there can be point identification.

For example, suppose X is discrete with K points of support, x_1, \dots, x_K , and suppose that for some value z^* of Z , $P[X = x_{k^*} | Z = z^*] = 1$. Then the values of all the threshold functions at $X = x_{k^*}$ are point identified and, for $m \in \{1, \dots, M\}$:²

$$p_m(x_{k^*}) = P[Y \leq m | Z = z^*]. \quad (8)$$

2.3.3. Discreteness of outcomes. The degree of discreteness in the distribution of Y affects the extent of the identified set. One of the examples in Section 3 demonstrates this in the context of parametric ordered-probit-type models where increasing the number of classes substantially reduces the size of the identified set of parameter values.

The difference between the two probabilities in the inequalities (4) which underpin the identification result is the conditional probability of the event: (Y, X) realisations lie *on* the structural function. This is an event of measure zero when Y is continuously distributed. As the support of Y grows more dense then as the distribution of Y comes to be continuous the maximal probability mass (conditional on X and Z) on any point of support of Y will pass to zero and the upper and lower bounds will come to coincide.

Even when the bounds coincide there can remain more than one observationally equivalent structural function admitted by the model. In the absence of parametric restrictions this is always the case when the support of Z is less rich than the support of X . The continuous outcome case is studied in Chernozhukov and Hansen (2005) and Chernozhukov, Imbens and Newey (2007) where completeness conditions are provided under which there is point identification of a structural function.

2.3.4. Local independence. It is possible to proceed under weaker independence restrictions, for example: $P[U \leq \tau | Z = z] = \tau$ for $\tau \in \tau_L$, some restricted set of values of τ , and $z \in \Omega_Z$. It is straightforward to show that, with this amendment to the model, Theorems 1 and 2 hold for $\tau \in \tau_L$ from which set identification of $h(\cdot, \tau)$ for $\tau \in \tau_L$ can be developed.

2.3.5. Estimation and inference. The identified set can be estimated by calculating (7) using an estimate of the distribution of $F_{Y|X|Z}$. Chernozhukov, Lee and Rosen (2008) give results on inference in the presence of intersection bounds.

The probability inequalities of Theorem 1 can be expressed as conditional moment inequalities involving binary indicators. Inference can be based on marginal moment

²This is so because

$$P[Y \leq m | Z = z^*] = \sum_{k=1}^K P[U \leq p_m(x_k) | x_k, z^*] P[X = x_k | z^*] = P[U \leq p_m(x_{k^*}) | x_{k^*}, z^*],$$

the second equality following because of perfect prediction at z^* . Because of the independence restriction and the uniform marginal distribution normalisation embodied in Restriction D2, for any value p :

$$p = P[U \leq p | z^*] = \sum_{k=1}^K P[U \leq p | x_k, z^*] P[X = x_k | z^*] = P[U \leq p | x_{k^*}, z^*]$$

which delivers the result (8) on substituting $p = p_m(x_{k^*})$.

inequalities obtained from these by taking expectations over the distribution of the instrumental variables after multiplication by a positive valued weight function.

Consider the case in which Z is a random variable and let E_{YXZ}^a indicate expectation with respect to a joint distribution F_{YXZ}^a . Under the model \mathcal{D} there are the following marginal moment inequalities holding for all $\tau \in (0, 1)$, for any bounded positive vector valued function $w(z)$ and for all functions h in the identified set associated with F_{YXZ}^a .

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

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

Taking this to data and evaluating over a grid of values of τ leads to set estimation and inference which falls in the class of problems addressed by Andrews, Berry and Jia (2004), Moon and Schorfheide (2006), Pakes, Porter, Ho and Ishii (2006) and Rosen (2006).³

3. ILLUSTRATIONS AND ELUCIDATION

This Section illustrates the set identification results with two examples.⁴ The first has a binary outcome and a discrete endogenous variable which for simplicity in this illustration is specified as binary. It is shown how the probability inequalities of Theorem 2 deliver inequalities on the values taken by the threshold crossing function which determine the binary outcome. In this simple case it is easy to develop admissible distributions for unobservables which taken with each member of the identified set deliver the probabilities used to construct the set. The resulting constructive demonstration of sharpness of identification serves as a template for such constructions in the general discrete endogenous variable problem and is somewhat more transparent than the construction in the proof of Theorem 2 which deals with continuous vector valued endogenous variables.

The second example employs a parametric ordered-probit-type model such as might be used when analysing interval censored data or data on ordered choices. This example demonstrates that parametric restrictions alone are not sufficient to deliver point identification. By varying the number of “choices” the impact on set identification of the degree of discreteness of an outcome is clearly revealed. In both examples one can clearly see the effect of instrument strength on the extent of identified sets.

3.1. Binary outcomes and binary endogenous variables. In the first example there is a threshold crossing model for a binary outcome Y with binary explanatory variable X , which may be endogenous. An unobserved scalar random variable

³Full exploitation of the restrictions of the model yields a *continuum* of moment inequalities on which there are 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).

⁴In Chesher (2007) there are additional examples covering cases in which the outcome has a binomial or a Poisson structural function.

U is continuously distributed, normalised Uniform on $(0, 1)$ and restricted to be distributed independently of instrumental variables Z .⁵ The model is as follows.

$$Y = h(X, U) \equiv \begin{cases} 0 & , \quad 0 < U \leq p(X) \\ 1 & , \quad p(X) < U \leq 1 \end{cases}, \quad U \perp\!\!\!\perp Z \in \Omega_Z, \quad U \sim Unif(0, 1)$$

The distribution of X is restricted to have support independent of U and Z with 2 distinct points of support: $\{x_1, x_2\}$.

The values taken by $p(X)$ are denoted by $\theta_1 \equiv p(x_1)$ and $\theta_2 \equiv p(x_2)$. These are the structural features whose identifiability is of interest. Here is a shorthand notation for the conditional probabilities about which data are informative..

$$\begin{aligned} \alpha_1(z) &\equiv P[Y = 0|x_1, z] & \alpha_2(z) &\equiv P[Y = 0|x_2, z] \\ \beta_1(z) &\equiv P[X = x_1|z] & \beta_2(z) &\equiv P[X = x_2|z] \end{aligned}$$

Here expressions like $P[Y = 0|x_1, z]$ are shorthand for $P[Y = 0|X = x_1, Z = z]$.

The set of values of $\Theta \equiv \{\theta_1, \theta_2\}$ identified by the model for a particular distribution of Y and X given $Z = z \in \Omega_Z$ is now obtained by applying the results given earlier. There is a set associated with each value of z in Ω_Z and the identified set for variations in z over Ω_Z is the intersection of the sets obtained at each value of z . The sharpness of the identified set is demonstrated by a constructive argument.

The identified set. First, expressions are developed for the probabilities that appear in the inequalities (4) which define the identified set. With these in hand it is straightforward to characterise the identified set. The ordering of θ_1 and θ_2 is important and in general is not restricted *a priori*.

First consider the case in which $\theta_1 \leq \theta_2$. Consider the event $\{Y < h(X, \tau)\}$. This occurs if and only if $h(X, \tau) = 1$ and $Y = 0$, and since $h(X, \tau) = 1$ if and only if $p(X) < \tau$ there is the following expression.

$$P[Y < h(X, \tau)|z] = P[(Y = 0) \cap (p(X) < \tau)|z] \tag{9}$$

So far as the inequality $p(X) < \tau$ is concerned there are three possibilities: $\tau \leq \theta_1$, $\theta_1 < \tau \leq \theta_2$ and $\theta_2 < \tau$. In the first case $p(X) < \tau$ cannot occur and the probability (9) is zero. In the second case $p(X) < \tau$ only if $X = x_1$ and the probability (9) is therefore

$$P[Y = 0 \cap (X = x_1)|z] = \alpha_1(z)\beta_1(z).$$

In the third case $p(X) < \tau$ whatever value X takes and the probability (9) is therefore

$$P[Y = 0|z] = \alpha_1(z)\beta_1(z) + \alpha_2(z)\beta_2(z).$$

⁵Here and in the second example, the independence restriction is the conditional independence required by Restriction D2, allowing the possibility that Z is not a random variable. The notation $U \perp\!\!\!\perp Z$ is shorthand for this conditional independence.

The situation is as follows.

$$P[Y < h(X, \tau)|z] = \begin{cases} 0 & , \quad 0 \leq \tau \leq \theta_1 \\ \alpha_1(z)\beta_1(z) & , \quad \theta_1 < \tau \leq \theta_2 \\ \alpha_1(z)\beta_1(z) + \alpha_2(z)\beta_2(z) & , \quad \theta_2 < \tau \leq 1 \end{cases}$$

The inequality $P[Y < h(X, \tau)|Z = z] < \tau$ restricts the identified set because in each row above the value of the probability must not exceed any value of τ in the interval to which it relates and in particular must not exceed the *minimum* value of τ in that interval. The result is the following pair of inequalities.

$$\alpha_1(z)\beta_1(z) \leq \theta_1 \quad \alpha_1(z)\beta_1(z) + \alpha_2(z)\beta_2(z) \leq \theta_2 \quad (10)$$

Now consider the event $\{Y \leq h(X, \tau)\}$. This occurs if and only if $h(X, \tau) = 1$ when any value of Y is admissible or $h(X, \tau) = 0$ and $Y = 0$. There is the following expression.

$$P[Y \leq h(X, \tau)|z] = P[(Y = 0) \cap (\tau \leq p(X))|z] + P[p(X) < \tau|z]$$

As before there are three possibilities to consider: $\tau \leq \theta_1$, $\theta_1 < \tau \leq \theta_2$ and $\theta_2 < \tau$. In the first case $\tau \leq p(X)$ occurs whatever the value of X and

$$P[Y \leq h(X, \tau)|z] = \alpha_1(z)\beta_1(z) + \alpha_2(z)\beta_2(z)$$

in the second case $\tau \leq p(X)$ when $X = x_2$ and $p(X) < \tau$ when $X = x_1$, so

$$P[Y \leq h(X, \tau)|z] = \beta_1(z) + \alpha_2(z)\beta_2(z)$$

while in the third case $p(X) < \tau$ whatever the value taken by X so

$$P[Y \leq h(X, \tau)|z] = 1.$$

The situation is as follows.

$$P[Y \leq h(X, \tau)|z] = \begin{cases} \alpha_1(z)\beta_1(z) + \alpha_2(z)\beta_2(z) & , \quad 0 \leq \tau \leq \theta_1 \\ \beta_1(z) + \alpha_2(z)\beta_2(z) & , \quad \theta_1 < \tau \leq \theta_2 \\ 1 & , \quad \theta_2 < \tau \leq 1 \end{cases}$$

The inequality $P[Y \leq h(X, \tau)|Z = z] \geq \tau$ restricts the identified set because in each row above the value of the probability must at least equal all values of τ in the interval to which it relates and in particular must at least equal the *maximum* value of τ in that interval. The result is the following pair of inequalities.

$$\theta_1 \leq \alpha_1(z)\beta_1(z) + \alpha_2(z)\beta_2(z) \quad \theta_2 \leq \beta_1(z) + \alpha_2(z)\beta_2(z) \quad (11)$$

Bringing (10) and (11) together gives, for the case in which $Z = z$, the part of the identified set in which $\theta_1 \leq \theta_2$, which is defined by the following inequalities.

$$\alpha_1(z)\beta_1(z) \leq \theta_1 \leq \alpha_1(z)\beta_1(z) + \alpha_2(z)\beta_2(z) \leq \theta_2 \leq \beta_1(z) + \alpha_2(z)\beta_2(z) \quad (12)$$

The part of the identified set in which $\theta_2 \leq \theta_1$ is obtained directly by exchange of indexes, thus:

$$\alpha_2(z)\beta_2(z) \leq \theta_2 \leq \alpha_1(z)\beta_1(z) + \alpha_2(z)\beta_2(z) \leq \theta_1 \leq \alpha_1(z)\beta_1(z) + \beta_2(z) \quad (13)$$

and the identified set for the case in which $Z = z$ is the union of the sets defined by the inequalities (12) and (13). The resulting set consists of two rectangles in the unit square, one above and one below the 45° line, oriented with edges parallel to the axes. The two rectangles intersect at the point $\theta_1 = \theta_2 = \alpha_1(z) = \alpha_2(z)$.

There is one such set for each value of z in Ω_Z and the identified set for $\Theta \equiv (\theta_1, \theta_2)$ delivered by the model is the intersection of these sets. The result is not in general a connected set, comprising two disjoint rectangles in the unit square, one strictly above and the other strictly below the 45° line. However with a strong instrument and rich support one of these rectangles will not be present.

Sharpness. The identified set just derived is sharp in the sense that for every value Θ in the identified set a distribution for U given X and Z can be found which is proper and satisfies the independence restriction, $U \perp\!\!\!\perp Z$, and delivers the distribution of Y given X and Z used to define the identified set. The existence of such a distribution is now demonstrated, providing a template for such constructions in other discrete outcome problems in which endogenous variables are discrete.

Consider some value z and a value $\Theta^* \equiv \{\theta_1^*, \theta_2^*\}$ with, say, $\theta_1^* \leq \theta_2^*$, which satisfy the inequalities (12), and consider a distribution function for U given X and Z , $F_{U|XZ}^*$. The proposed distribution is piecewise uniform but other choices could be made. Define values of the proposed distribution function as follows.

$$\begin{aligned} F_{U|XZ}^*(\theta_1^*|x_1, z) &\equiv \alpha_1(z) & F_{U|XZ}^*(\theta_1^*|x_2, z) &\equiv (\theta_1^* - \beta_1(z)\alpha_1(z)) / \beta_2(z) \\ F_{U|XZ}^*(\theta_2^*|x_1, z) &\equiv (\theta_2^* - \beta_2(z)\alpha_2(z)) / \beta_1(z) & F_{U|XZ}^*(\theta_2^*|x_2, z) &\equiv \alpha_2(z) \end{aligned} \quad (14)$$

The choice of values for $F_{U|XZ}^*(\theta_1^*|x_1, z)$ and $F_{U|XZ}^*(\theta_2^*|x_2, z)$ ensures that this structure is observationally equivalent to the structure which generated the conditional probabilities that define the identified set.⁶ The proposed distribution respects the independence restriction because the implied probabilities marginal with respect to X are independent of z , as follows.

$$\begin{aligned} P[U \leq \theta_1^*|z] &= \beta_1(z)F_{U|XZ}^*(\theta_1^*|x_1, z) + \beta_2(z)F_{U|XZ}^*(\theta_1^*|x_2, z) = \theta_1^* \\ P[U \leq \theta_2^*|z] &= \beta_1(z)F_{U|XZ}^*(\theta_2^*|x_1, z) + \beta_2(z)F_{U|XZ}^*(\theta_2^*|x_2, z) = \theta_2^* \end{aligned}$$

It just remains to determine whether the proposed distribution of U given X and $Z = z$ is proper, that is has probabilities lying in the unit interval and respecting monotonicity. Both $F_{U|XZ}^*(\theta_1^*|x_1, z)$ and $F_{U|XZ}^*(\theta_2^*|x_2, z)$ lie in $[0, 1]$ by definition. The other two elements lie in the unit interval if and only if

$$\begin{aligned} \beta_1(z)\alpha_1(z) &\leq \theta_1^* \leq \beta_1(z)\alpha_1(z) + \beta_2(z) \\ \beta_2(z)\alpha_2(z) &\leq \theta_2^* \leq \beta_1(z) + \beta_2(z)\alpha_2(z) \end{aligned}$$

⁶This because for $j \in \{1, 2\}$, $\alpha_j(z) \equiv P[Y = 0|x_j, z] = P[U \leq \theta_j|X = x_j, Z = z]$.

which both hold when θ_1^* and θ_2^* satisfy the inequalities (12). The case under consideration has $\theta_1^* \leq \theta_2^*$ so if the distribution function of U given X and $Z = z$ is to be monotonic, it must be that the following inequalities hold.

$$\begin{aligned} F_{U|XZ}^*(\theta_1^*|x_1, z) &\leq F_{U|XZ}^*(\theta_2^*|x_1, z) \\ F_{U|XZ}^*(\theta_1^*|x_2, z) &\leq F_{U|XZ}^*(\theta_2^*|x_2, z) \end{aligned}$$

. Manipulating the expressions in (14) yields the result that these inequalities are satisfied if:

$$\theta_1^* \leq \alpha_1(z)\beta_1(z) + \alpha_2(z)\beta_2(z) \leq \theta_2^*$$

which is assured when θ_1^* and θ_2^* satisfy the inequalities (12). There is a similar argument for the case $\theta_2^* \leq \theta_1^*$.

This argument applies at each value $z \in \Omega_Z$ so it can be concluded that for each value Θ^* in the identified set formed by intersecting sets obtained at each $z \in \Omega_Z$ there exists a proper distribution function $F_{U|XZ}^*$ with U independent of Z which, combined with that value delivers the probabilities used to define the identified set.

Numerical example. The identified sets are illustrated using probability distributions generated by a structure in which binary $Y \equiv 1[Y^* > 0]$ and $X \equiv 1[X^* > 0]$ are generated by a triangular linear equation system which delivers values of latent variables Y^* and X^* as follows.

$$\begin{aligned} Y^* &= a_0 + a_1X + \varepsilon \\ X^* &= b_0 + b_1Z + \eta \end{aligned}$$

Latent variates ε and η are jointly normally distributed conditional on an instrumental variable Z .

$$\begin{bmatrix} \varepsilon \\ \eta \end{bmatrix} | Z \sim N \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & r \\ r & 1 \end{bmatrix} \right)$$

Let Φ denote the standard normal distribution function. The structural equation for binary Y is as follows:

$$Y = \begin{cases} 0 & , \quad 0 < U \leq p(X) \\ 1 & , \quad p(X) < U \leq 1 \end{cases}$$

with $U \equiv \Phi(\varepsilon) \sim Unif(0, 1)$ and $U \perp\!\!\!\perp Z$ and $p(X) = \Phi(-a_0 - a_1X)$ with $X \in \{0, 1\}$.

Figure 1 shows identified sets when the parameter values generating the probabilities are:

$$a_0 = 0 \quad a_1 = 0.5 \quad b_0 = 0 \quad b_1 = 1 \quad r = -0.25$$

for which

$$p(0) = \Phi(-a_0) = 0.5 \quad p(1) = \Phi(-a_0 - a_1) = 0.308$$

and z takes values in $\Omega_Z \equiv \{0, -75, -.75\}$.

Pane (a) in Figure 1 shows the identified set when $z = 0$. It comprises two rectangular regions, touching at the point $p(0) = p(1)$ but otherwise not connected.

In the upper rectangle $p(1) \geq p(0)$ and in the lower rectangle $p(1) \leq p(0)$. The dashed lines intersect at the location of $p(0)$ and $p(1)$ in the structure generating the probability distributions used to calculate the identified sets. In that structure $p(0) = 0.5 > p(1) = 0.308$ but there are observationally equivalent structures lying in the rectangle above the 45° line in which $p(1) > p(0)$.

Pane (b) in Figure 1 shows the identified set when $z = .75$ - at this instrumental value the range of values of $p(1)$ in the identified set is smaller than when $z = 0$ but the range of values of $p(0)$ is larger. Pane (c) shows the identified set when $z = -.75$ - at this instrumental value the range of values of $p(1)$ in the identified set is larger than when $z = 0$ and the range of values of $p(0)$ is smaller. Pane (d) shows the identified set (the solid filled rectangle) when all three instrumental values are available.

The identified set is the intersection of the sets drawn in Panes (a) - (c). The strength and support of the instrument in this case is sufficient to eliminate the possibility that $p(1) > p(0)$. If the instrument were stronger ($b_1 \gg 1$) the solid filled rectangle would be smaller and as b_1 increased without limit it would contract to a point. For the structure used to construct this example the model achieves “point identification at infinity” because the mechanism generating X is such that as Z passes to $\pm\infty$ the value of X becomes perfectly predictable.

Figure 2 shows identified sets when the instrument is weaker, achieved by setting $b_1 = 0.3$. In this case even when all three values of the instrument are employed there are observationally equivalent structures in which $p(1) > p(0)$.⁷

3.2. Ordered outcomes: a parametric example. In the second example Y records an ordered outcome in M classes, X is a continuous explanatory variable and there are parametric restrictions. The model used in this illustration has Y generated as in an ordered probit model with specified threshold values c_0, \dots, c_M and potentially endogenous X . The unobservable variable in a threshold crossing representation is distributed independently of Z which varies across a set of instrumental values, Ω_Z . This sort of specification might arise when studying ordered choice using a ordered probit model or when employing interval censored data to estimate a linear model, in both cases allowing for the possibility of endogenous variation in the explanatory variable. In order to allow a graphical display of the identified sets just two parameters are unrestricted in this example. In many applications there would be other free parameters, for example the threshold values.

The parametric model considered states that for some constant parameter value $\alpha \equiv (\alpha_0, \alpha_1)$,

$$Y = h(X, U; \alpha) \quad U \perp\!\!\!\perp Z \in \Omega_Z \quad U \sim Unif(0, 1)$$

where, for $m \in \{1, \dots, M\}$, with Φ denoting the standard normal distribution function:

$$h(X, U; \alpha) = m, \text{ if: } \Phi(c_{m-1} - \alpha_0 - \alpha_1 X) < U \leq \Phi(c_m - \alpha_0 - \alpha_1 X)$$

and $c_0 = -\infty$, $c_M = +\infty$ and c_1, \dots, c_{M-1} are specified finite constants. The notation

⁷In supplementary material more extensive graphical displays are available.

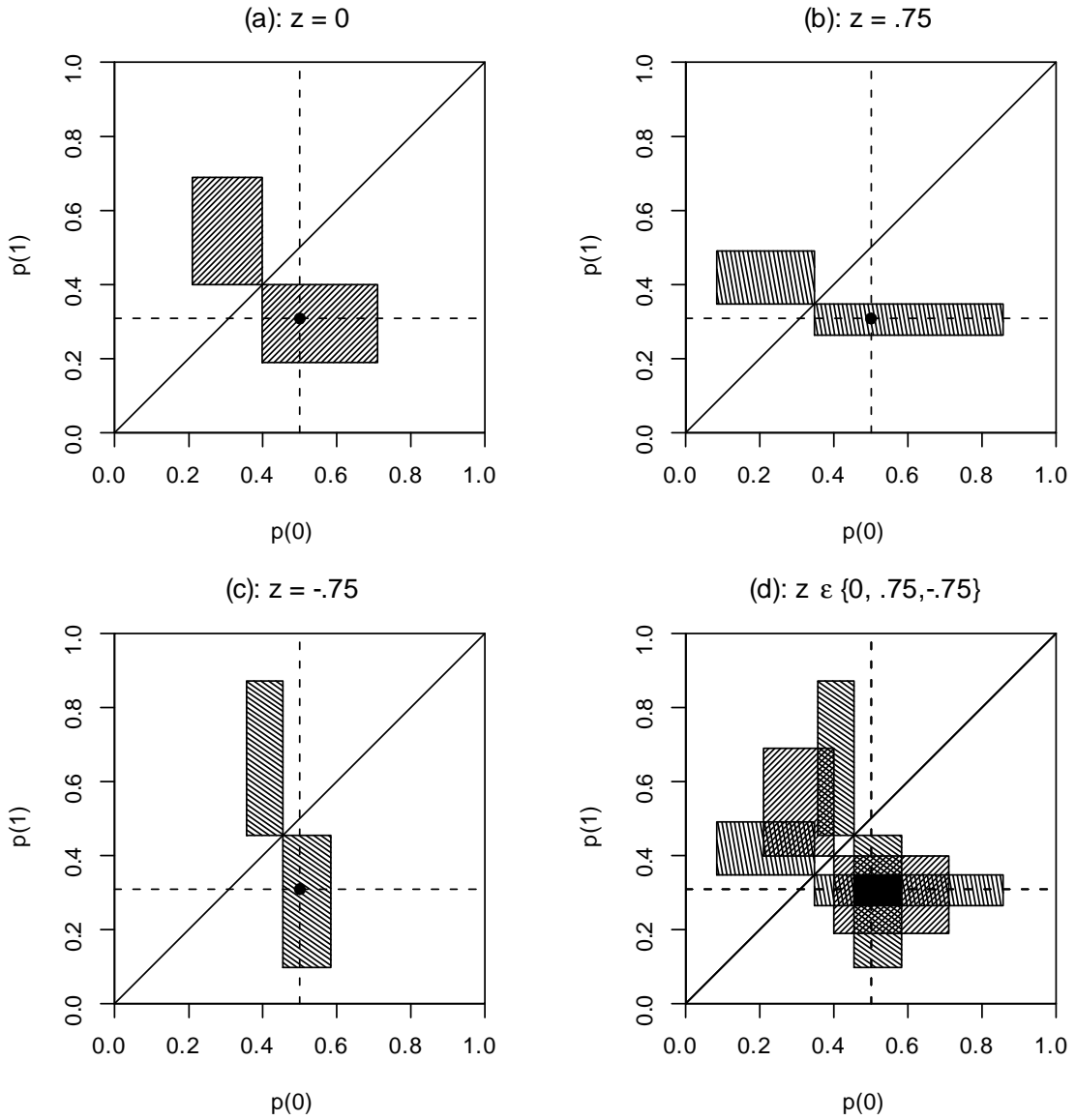


Figure 1: Identified sets with a binary outcome and binary endogenous variable as instrumental values, z , vary. Strong instrument ($b_1 = 1$). Dotted lines intersect at the values of $p(0)$ and $p(1)$ in the distribution generating structure. Panes (a) - (c) show identified sets at each of 3 values of the instrument. Pane (d) shows the intersection (solid area) of these identified sets. The instrument is strong enough and has sufficient support to rule out the possibility $p(1) > p(0)$.

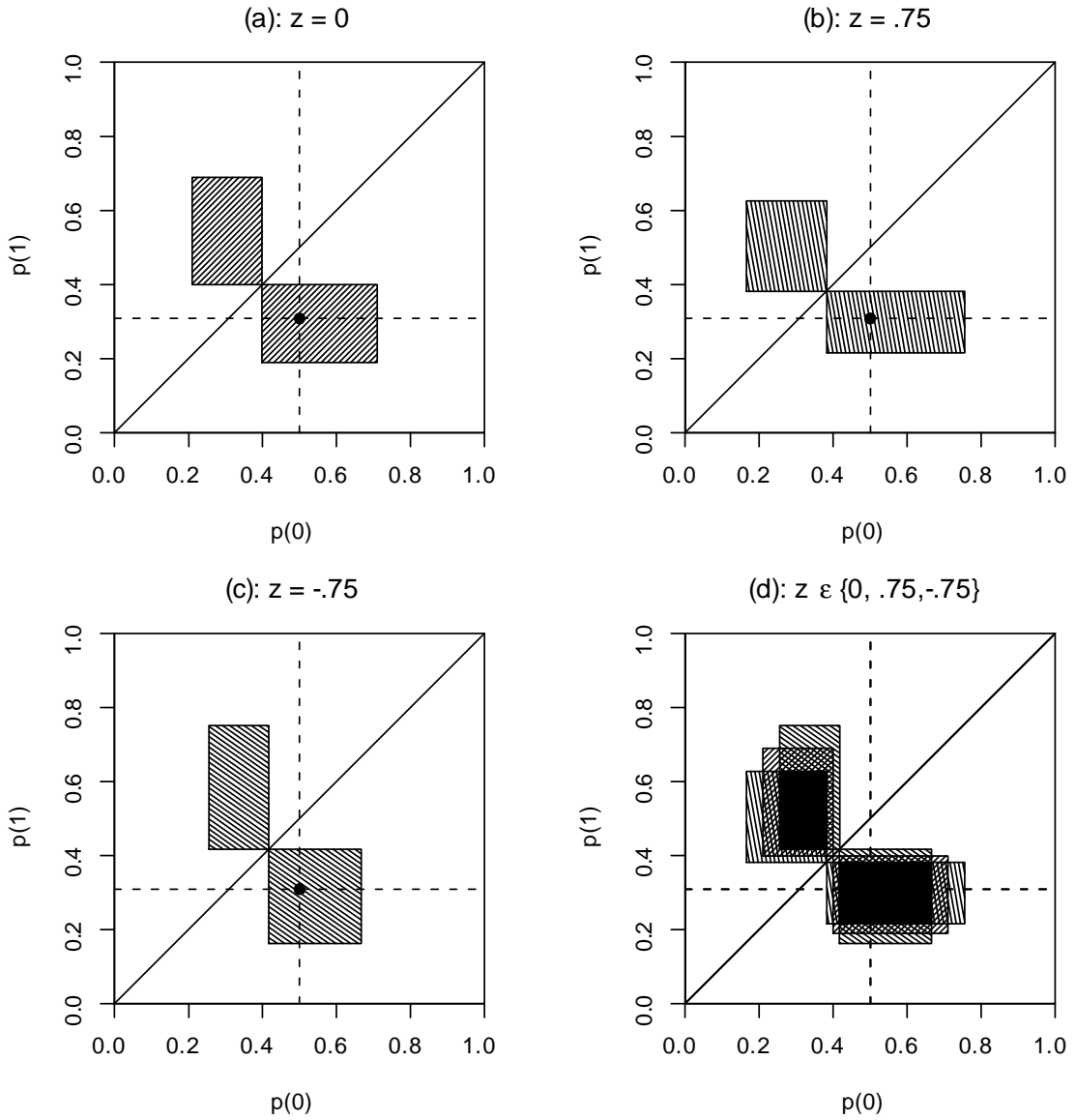


Figure 2: Identified sets with a binary outcome and binary endogenous variable as instrumental values, z , vary. Weak instrument ($b_1 = 0.3$). Dotted lines intersect at the values of $p(0)$ and $p(1)$ in the distribution generating structure. Panes (a) - (c) show identified sets at each of 3 values of the instrument. Pane (d) shows the intersection (solid area) of these identified sets. The instrument is weak and there are observationally equivalent structures in which $p(1) > p(0)$.

$h(X, U; \alpha)$ makes explicit the dependence of the structural function on the parameter α

For a conditional probability function $F_{Y|XZ}$ and a conditional density $f_{X|Z}$ and some value α the probabilities in (6) are:

$$\Pr[Y \leq h(X, \tau; \alpha) | Z = z] = \sum_{m=1}^M \int_{\{x: h(x, \tau; \alpha) = m\}} F_{Y|XZ}(m|x, z) f_{X|Z}(x|z) dx \quad (15)$$

$$\Pr[Y < h(X, \tau; \alpha) | Z = z] = \sum_{m=2}^M \int_{\{x: h(x, \tau; \alpha) = m\}} F_{Y|XZ}(m-1|x, z) f_{X|Z}(x|z) dx \quad (16)$$

In the numerical calculations the conditional distribution of Y and X given $Z = z$ is generated by a structure of the following form.

$$Y^* = a_0 + a_1 X + W \quad X = b_0 + b_1 Z + V$$

$$\begin{bmatrix} W \\ V \end{bmatrix} | Z \sim N \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & s_{uv} \\ s_{uv} & s_{vv} \end{bmatrix} \right)$$

$$Y = m, \text{ if: } c_{m-1} < Y \leq c_m, \quad m \in \{1, \dots, M\}$$

Here $c_0 \equiv -\infty$, $c_M \equiv \infty$ and c_1, \dots, c_{M-1} are the specified finite constants employed in the definition of the structure and in the parametric model whose identifying power is being considered.

The probabilities in (15) and (16) are calculated for each choice of α by numerical integration.⁸ Illustrative calculations are done for 5 and 11 class specifications with thresholds chosen as quantiles of the standard normal distribution at equispaced probability levels. For example in the 5 class case the thresholds are $\Phi^{-1}(p)$ for $p \in \{.2, .4, .6, .8\}$, that is $\{-.84, -.25, .25, .84\}$. The instrumental variable ranges over the interval $\Omega_Z \equiv [-1, 1]$, the parameter values employed in the calculations are:

$$a_0 = 0, \quad a_1 = 1, \quad b_0 = 0, \quad s_{uv} = 0.6, \quad s_{vv} = 1$$

and the value of b_1 is set to 1 or 2 to allow comparison of identified sets as the strength of the instrument varies.⁹

Figure 3 shows the identified set for the intercept and slope coefficients, α_0 and α_1 in a 5 class model. The dark shaded set is obtained when the instrument is relatively strong ($b_1 = 2$). This set lies within the identified set obtained when the instrument is relatively weak ($b_1 = 1$). Figure 4 shows identified sets (shaded) for these weak and strong instrument scenarios when there are 11 classes rather than 5. The 5 class

⁸The `integrate` procedure in R (Ihaka and Gentleman (1996)) was used to calculate probabilities. Intersection bounds over $z \in \Omega_Z$ were obtained as in (7) using the R function `optimise`. The resulting probability inequalities were inspected over a grid of values of τ at each value of α considered, a value being classified as out of the identified set as soon as a value of τ was encountered for which there was violation of one or other of the inequalities (7). I am grateful to Konrad Smolinski for developing and programming a procedure to efficiently track out the boundary of the identified sets.

⁹Equivalently as the support of the instrument varies because varying b_1 from 1 to 2 has the same effect as holding b_1 fixed at 1 and varying Ω_Z from $[-1, 1]$ to $[-2, 2]$.

sets are shown in outline. The effect of reducing the discreteness of the outcome is substantial and there is still a substantial reduction in the extent of the identified set as the instrument is strengthened.

4. CONCLUDING REMARKS

It has been shown that, when outcomes are discrete, single equation IV models do not point identify the structural function that delivers the discrete outcome. The models have been shown to have partial identifying power and sharp set identification results have been obtained. Identified sets tend to be smaller when instrumental variables are strong and have rich support and when the discrete outcome has rich support. Imposing parametric restrictions reduces the extent of the identified sets but in general parametric restrictions do not deliver point identification of the values of parameters.

To secure point identification of structural functions more restrictive models are required. For example, specifying recursive structural equations for endogenous explanatory variables and restricting all latent variates and instrumental variables to be jointly independently distributed produces a triangular system model which can be point identifying.¹⁰ This is the control function approach studied in Blundell and Powell (2004), Chesher (2003) and Imbens and Newey (2003). The restrictions of the triangular model rule out full simultaneity (Koenker (2005), Section 8.8.2) such as arises in the simultaneous entry game model of Tamer (2003). An advantage of the single equation IV approach set out in this paper is that it allows an equation-by-equation attack on such simultaneous equations models for discrete outcomes, avoiding the need to deal directly with the coherency and completeness issues they pose.

The weak restrictions imposed in the single equation IV model lead to partial identification of deep structural objects which complements the many developments in the analysis of point identification of the various average structural features studied in for example Heckman and Vytlacil (2005) and their many other works referenced there. This paper's set identification results for structural functions together with the methods set out here for construction of distributions of latent variates leads to a complete characterisation of the distinct observationally equivalent structures admitted by the single equation IV model. With this to hand a next step in this research agenda is to examine the identifiability of a wide variety of alternative structural features.

There are a number of immediate extensions of the results resented here. For example the analysis can be extended to the *vector* discrete outcome case such as arises in the study of panel data. Here endogeneity may arise because of the presence of "individual effects" which may be correlated with observed explanatory variables. Consider a case in which there are T discrete outcomes each determined by a structural equation as follows:

$$Y_t = h_t(X, U_t), \quad t = 1, \dots, T$$

¹⁰But not when endogenous variables are discrete, Chesher (2005).

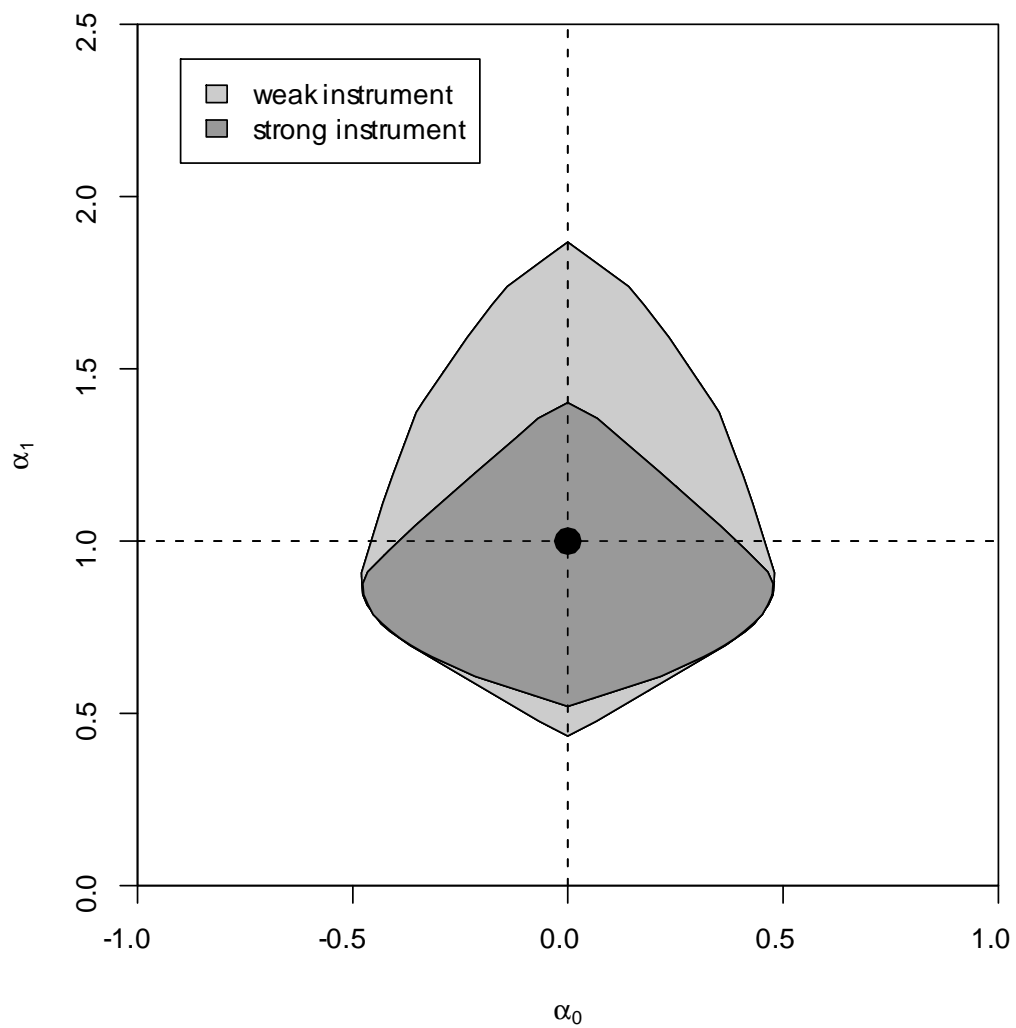


Figure 3: Identified sets for intercept, α_0 , and slope coefficient, α_1 , in a 5 class ordered probit model with endogenous explanatory variable. The dashed lines intersect at the values of a_0 and a_1 used to generate the distributions employed in this illustration.

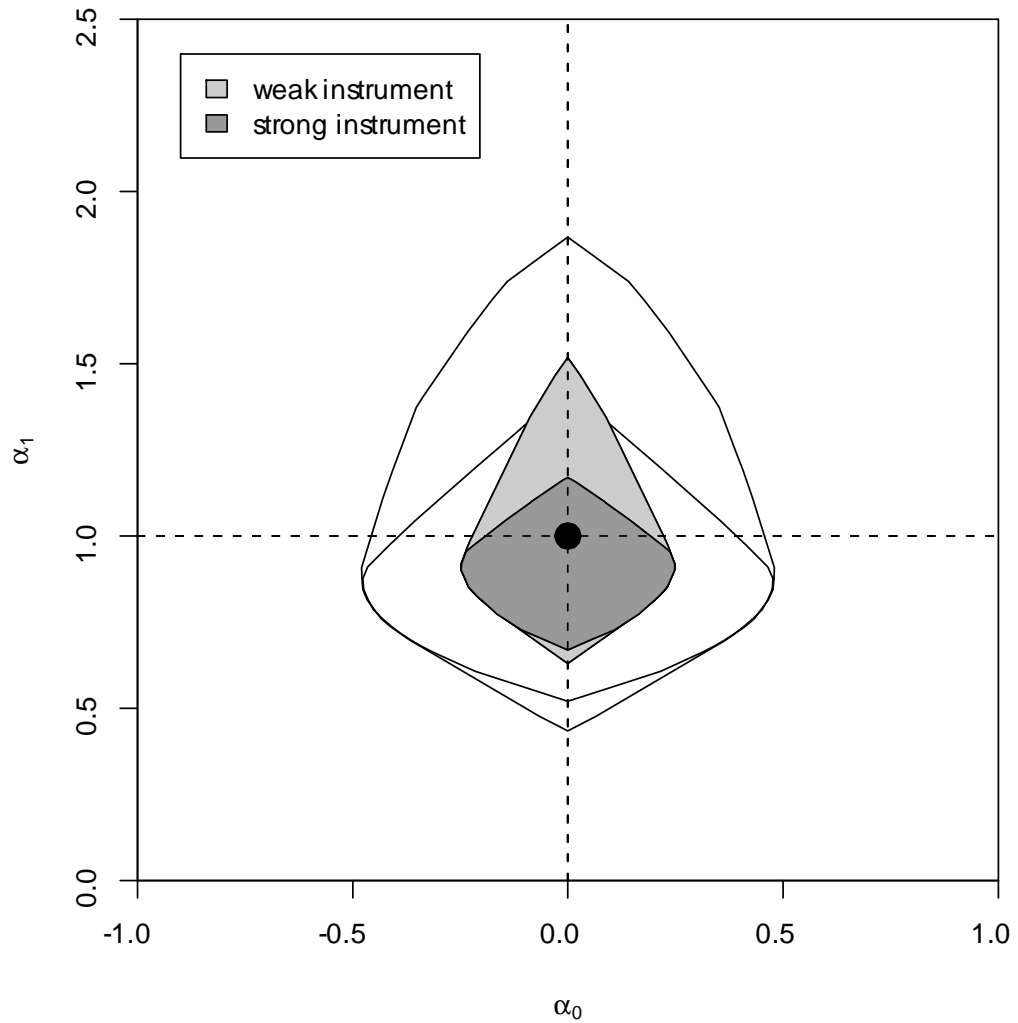


Figure 4: Identified sets for intercept, α_0 , and slope coefficient, α_1 , in a 11 class ordered probit model with endogenous explanatory variable. Identified sets for the 5 class model displayed in Figure 3 are shown in outline. The dashed lines intersect at the values of a_0 and a_1 used to generate the distributions employed in this illustration.

in which each function h_t is weakly increasing in U_t and each U_t is a scalar random variable normalised marginally $Unif(0, 1)$ and $U \equiv \{U_t\}_{t=1}^T$ and instrumental variables $Z \in \Omega_Z$ are independently distributed. In practice there will often be cross equation restrictions; for example requiring each function h_t to be determined by a common set of parameters.

Define $h \equiv \{h_t\}_{t=1}^T$ and $\tau \equiv \{\tau_t\}_{t=1}^T$ and:

$$C(\tau) \equiv \Pr \left[\bigcap_{t=1}^T (U_t \leq \tau_t) \right]$$

which is a *copula* since the components of U have marginal uniform distributions. An argument along the lines of that used in Section 2.1 leads to the following pair of inequalities which hold for all $\tau \in [0, 1]^T$ and $z \in \Omega_Z$.

$$\Pr \left[\bigcap_{t=1}^T (Y_t \leq h_t(X, \tau_t)) | Z = z \right] \geq C(\tau)$$

$$\Pr \left[\bigcap_{t=1}^T (Y_t < h_t(X, \tau_t)) | Z = z \right] < C(\tau)$$

Let \Pr_a indicate probabilities taken with respect to a distribution function $F_{YX|Z}^a$ generated by an admissible structure $S^a \equiv \{h^a, F_{UX|Z}^a\}$. Consider an admissible structure $S^* \equiv \{h^*, F_{UX|Z}^*\}$ and let C^* be the copula (for U) associated with the distribution function $F_{UX|Z}^*$. Arguing as in Section 2.2 it can be shown that S^* is observationally equivalent to S^a for $z \in \Omega_Z$ if and only if the following inequalities hold for all $\tau \in [0, 1]^T$ and $z \in \Omega_Z$.

$$\Pr_a \left[\bigcap_{t=1}^T (Y_t \leq h_t^*(X, \tau_t)) | Z = z \right] \geq C^*(\tau)$$

$$\Pr_a \left[\bigcap_{t=1}^T (Y_t < h_t^*(X, \tau_t)) | Z = z \right] < C^*(\tau)$$

These inequalities define an identified set of structural functions and copula functions associated with S^a . The set may be reduced in extent by restrictions on admissible copulas.

Some other extensions of the results arise on relaxing restrictions embodied in the model considered so far. For example it is straightforward to generalise to the case in which exogenous variables appear in the structural function. In the binary outcome case additional heterogeneity, W , independent of instruments Z , can be introduced if there is a monotone index restriction, that is if the structural function has the form $h(X\beta, U, W)$ with h monotonic in $X\beta$ and in U . This allows extension to measurement error models in which observed $\tilde{X} = X + W$. This can be further extended to the general discrete outcome case if a monotone index restriction holds for all threshold functions.

ACKNOWLEDGEMENTS

I thank Victor Chernozhukov, Martin Cripps, Russell Davidson, Simon Sokbae Lee, Arthur Lewbel, Charles Manski, Lars Nesheim, Adam Rosen, and Richard Spady for stimulating comments and discussions. The support of the Leverhulme Trust through a grant to the research project *Evidence Inference and Inquiry* and through a grant to the Centre for Microdata Methods and Practice (CeMMAP) is acknowledged. The support for CeMMAP given by the UK Economic and Social Research Council under grant RES-589-28-0001 is acknowledged. This is a revised version of the CeMMAP Working Paper CWP 05/07, “Endogeneity and Discrete Outcomes”. The main results of the paper were presented at an Oberwolfach Workshop on March 19th 2007. Detailed results for binary response models were given at a conference in honour of the 60th birthday of Peter Robinson at the LSE, May 25th 2007. I am grateful for comments at these meetings and at subsequent presentations of this and related papers at the European Meeting of the Econometric Society, Budapest, August 2007, CeMMAP, Cowles Foundation, CREST, EUI, the Universities of Amsterdam, Bristol, Cambridge, Illinois at Urbana-Champaign, Manchester, Nottingham, Oxford, Pennsylvania, Zurich, Caltech, Chicago Booth School of Business, Northwestern University, The Johns Hopkins University, UCLA, USC, and the CIREQ conference on Inference in Incomplete Models, Montreal, October 25th 2008.

REFERENCES

- 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.
- CAMERON, A. COLIN AND PRAVIN K. TRIVEDI (1998): *Regression Analysis of Count Data*. Econometric Society Monograph No. 30. 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 (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.
- CHERNOZHUKOV, VICTOR, LEE, SOKBAE AND ADAM ROSEN (2008): “Intersection Bounds: Estimation and Inference,” unpublished paper, presented at the cemmap - Northwestern conference on *Inference in Partially Identified Models with Applications*, London, March 27th 2008.
- 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.,(2007): “Endogeneity and Discrete Outcomes,” CeMMAP Working Paper, CWP05/07.
- HECKMAN, J.J., AND E. VYTLACIL (2005): “Structural Equations, Treatment Effects, and Econometric Policy Evaluation,” *Econometrica*, 73, 669-738.
- 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 EC² Meeting, cemmap, London, December 12th - 13th 2003.
- KOENKER, R.W., (2005): *Quantile Regression*. Econometric Society Monograph No. 38. Cambridge University Press, Cambridge.
- MOON, HYUNSIK 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.
- ROSEN, ADAM (2006): “Confidence Sets for Partially Identified Parameters that Satisfy a Finite Number of Moment Inequalities,” CeMMAP Working Paper, CWP25/06.
- 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

Proof of Lemma 1. Probability inequalities

It is shown that under restriction D1 of model \mathcal{D} there are the following inequalities which hold for all x, z and τ .

$$\Pr[Y \leq h(X, \tau) | X = x, Z = z] \geq F_{U|XZ}(\tau | x, z) \quad (\text{A1.1})$$

$$\Pr[Y < h(X, \tau) | X = x, Z = z] < F_{U|XZ}(\tau | x, z) \quad (\text{A1.2})$$

Consider the first inequality, (A1.1). Directly from the definition of Y :

$$\Pr[Y \leq h(X, \tau) | X = x, Z = z] = \Pr[h(X, U) \leq h(X, \tau) | X = x, Z = z] \quad (\text{A1.3})$$

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

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

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

$$\Pr[Y \leq h(X, p_m(x)) | X = x, Z = z] = F_{U|XZ}(p_m(x) | x, z)$$

and so

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

and since $F_{U|XZ}(t | x, z)$ 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,

$$\Pr[Y \leq h(X, \tau) | X = x, Z = z] \geq F_{U|XZ}(\tau | x, z)$$

which is the inequality (A1.1).

Consider the *second* inequality, (A1.2) and equation (A1.3). Because $h(x, \tau) = y_m$ if and only if $\tau \in (p_{m-1}(x), p_m(x)]$ there is the following.

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

Arguing as before:

$$\Pr[Y \leq h(X, p_{m-1}(x)) | X = x, Z = z] = F_{U|XZ}(p_{m-1}(x) | x, z)$$

and so

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

and because $F_{U|XZ}(\tau | x, z)$ 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

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

which is the inequality (A1.2).

Proof of Theorem 3. Sharp set identification

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

1. is weakly monotonic non-decreasing for variations in u ,
2. is characterised by threshold 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 1.¹¹

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

Let $\Omega_{X|Z}$ denote the support of X conditional on Z . Define conditional probabilities as follows.

$$\begin{aligned} \alpha_m(x) &\equiv \Pr[Y = y_m | X = x, Z = z] = F_{Y|XZ}(y_m | x, z) \\ \alpha_m &\equiv \Pr[Y = y_m | Z = z] = \int_{x \in \Omega_{X|Z}} \alpha_m(x) dF_{X|Z}(x | z) \end{aligned}$$

¹¹This method of construction builds on a method proposed for the discrete endogenous variable, binary outcome case by Martin Cripps.

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

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:

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

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

Define sets as follows. Let ϕ denote the empty set.

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

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

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

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

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.

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

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

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

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

It is now shown that $F_{U|XZ}$ is a *proper distribution function* exhibiting the *inde-*

pendence 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 non-decreasing, therefore the functions $s_m(v)$ are non-decreasing. 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 non-decreasing from which the result follows.
4. Independence. For each m and all v :

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

and so

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

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

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

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

This involves the cumulative probabilities $\bar{\alpha}_m(x)$. It is convenient to write the in-

equalities in terms of the point probabilities $\{\alpha_m\}_{m=1}^M$, as follows.

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

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.

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

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

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

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

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

equivalently

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

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 non-decreasing 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.

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

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 non-decreasing 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.