cemmap
centre for microdata methods and practice

# 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

Institute for Fiscal Studies, 7 Ridgmount Street, London WC1E 7AE **Tel:** 44 (0)20 7291 4800 **Email:** info@cemmap.ac.uk **www.cemmap.ac.uk**

**Instrumental Variable Models for Discrete Outcomes**


Andrew Chesher*
CeMMAP and UCL



Revised July 3rd 2009



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. Bounds on identified sets are derived for a general nonparametric model and sharp set identification is demonstrated in the binary outcome case. 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 and càglàd 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 $\Omega$. $X$ may be endogenous in the sense that $U$ and $X$ may not be independently distributed. This is a single


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

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.

There could be parametric restrictions. For example the function $h(X,U)$ could be specified to be the structural function associated with a probit or a logit model with endogenous $X$, in the latter case:

$$h(X,U) = 1\left[U > \left(1+\exp(X'\beta)\right)^{-1}\right] \qquad U \sim Unif(0,1)$$

with $U$ potentially jointly dependent with $X$ but independent of instrumental variables $Z$ which are excluded from $h$. The results of this paper apply in this case. Until now instrumental variables analysis of binary outcome models has been confined to linear probability models.

The central result of this paper is that the single equation IV model *set identifies* the structural function $h$. Parametric restrictions on the structural function do not typically secure point identification although they may reduce the extent of identified sets.

Underpinning the identification results are the following inequalities:

$$\text{for all } \tau \in (0,1) \text{ and } z \in \Omega: \quad \begin{array}{l} \Pr_a[Y \leq h(X,\tau)|Z=z] \geq \tau \\ \Pr_a[Y < h(X,\tau)|Z=z] < \tau \end{array} \tag{1}$$

which hold for any structural function $h$ which is an element of an admissible structure that generates the probability measure indicated by $\Pr_a$.

In the binary outcome case these inequalities *sharply* define the identified set of structural functions for the probability measure under consideration in the sense that all functions $h$, and only functions $h$, that satisfy these inequalities for all $\tau \in (0,1)$ and all $z \in \Omega$ are elements of the observationally equivalent admissible structures which generate the probability measure $\Pr_a$.

When $Y$ has more than two points of support the model places restrictions on structural functions additional to those that come from (1) and the inequalities define an *outer region*[1], that is a set within which lies the set of structural functions identified by the model. Calculation of the sharp identified set seems infeasible when $X$ is continuous or discrete with many points of support without additional restrictions. Similar issues arise in some of the models of oligopoly market entry discussed in Berry and Tamer (2006).

When the outcome $Y$ is continuously distributed (in which case $h$ is strictly monotonic in $U$) both probabilities in (1) are equal to $\tau$ 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 a discrete outcome and endogenous explanatory variables use a control function approach. This is rooted in a more restrictive complete, triangular model which can be point identifying but the model's restrictions are not always applicable. There is a brief discussion in Section 4 and a detailed

[1] This terminology is borrowed from Beresteanu, Molchanov and Molinari (2008).

comparison with the single equation instrumental variable model in Chesher (2009).

A few papers take a single equation IV approach to endogeneity in parametric count data models, basing identification on moment conditions.[2] Mullahy (1997) and Windmeijer and Santos Silva (1997) consider models 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 $\beta$ but the fine details of the functional form restrictions are influential in securing point identification and the approach, based as it is on a multiplicative heterogeneity specification, is not applicable 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 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.

## 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 deliver bounds on the set of structural functions identified by the IV model in the $M > 2$ outcome case and sharp identification in the binary outcome case. Section 2.3 discusses the identification results with brief comments on: the impact of support and strength of instruments and discreteness of outcome on the identified set, sharpness, and local independence restrictions.

**2.1. Model.** 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 some 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 $z \in \Omega$.*

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

[2] See the discussion in Section 11.3.2 of Cameron and Trivedi (1988).

as follows:

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

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 $Y$ increasing as $U$ ascends through thresholds which depend on the value of the explanatory variables, $X$, but not on $Z$.

Restriction D2 requires that the conditional distribution of $U$ given $Z = z$ be invariant with respect to $z$ for variations within $\Omega$. If $Z$ is a random variable and $\Omega$ 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 $\Omega$ 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 $Z = z \in \Omega$, 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 $S^a \equiv \{h^a, F^a_{UX|Z}\}$ with structural function $h^a$ delivers a conditional distribution for $(Y, X)$ given $Z$, $F^a_{YX|Z}$, as follows.

$$F^a_{YX|Z}(y_m, x|z) = F^a_{UX|Z}(p^a_m(x), x|z), \qquad m \in \{1, \ldots, M\} \tag{3}$$

Here the functions $\{p^a_m(x)\}_{m=0}^M$ 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 $z \in \Omega$. 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 $p^a_m(x)$ can be offset by altering the sensitivity of $F^a_{UX|Z}(u, x|z)$ to variations in $u$ and $x$ so that the left hand side of (3) is left unchanged.

Crucially the independence restriction D2 places limits on the variations in the functions $p^a_m(x)$ that can be so compensated and results in the model having *nontrivial set identifying power.* A pair of probability inequalities place limits on the structural functions which lie in the set identified by the model. They are the subject of the following Theorem.

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

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

**Proof of Theorem 1**. *For all $x$ each admissible $h^a(x,u)$ is càglàd for variations in $u$, and so for all $x$ and $\tau \in (0,1)$:*

$$\{u : h^a(x,u) \leq h^a(x,\tau)\} \supseteq \{u : u \leq \tau\}$$

$$\{u : h^a(x,u) < h^a(x,\tau)\} \subset \{u : u \leq \tau\}$$

*which lead to the following inequalities which hold for all $\tau \in (0,1)$ and for all $x$ and $z$.*

$$\Pr_a[Y \leq h^a(X,\tau)|X=x, Z=z] \geq F^a_{U|XZ}(\tau|x,z)$$

$$\Pr_a[Y < h^a(X,\tau)|X=x, Z=z] < F^a_{U|XZ}(\tau|x,z)$$

*Let $F^a_{X|Z}$ be the distribution function of $X$ given $Z$ associated with $F^a_{YX|Z}$. 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 $F^a_{U|Z}(\tau|z)$ which is equal to $\tau$ for all $z \in \Omega$ and $\tau \in (0,1)$ under the conditions of model $\mathcal{D}$.*

**2.2. Identification.** Consider the model $\mathcal{D}$, a structure $S_a = \{h^a, F^a_{UX|Z}\}$ admitted by it, and the set $\tilde{\mathcal{S}}_a$ 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 $\tilde{\mathcal{S}}_a$. Let $F^a_{YX|Z}$ be the joint distribution function of $(Y,X)$ given $Z$ delivered by the observationally equivalent structures in the set $\tilde{\mathcal{S}}_a$.

The model $\mathcal{D}$ set identifies the structural function generating $F^a_{YX|Z}$ - it must be one of the structural functions in the set $\tilde{\mathcal{H}}_a$. The inequalities (4) constrain this set as follows: all structural functions in the identified set $\tilde{\mathcal{H}}_a$ satisfy the inequalities (4) when they are calculated using the probability distribution $F^a_{YX|Z}$, conversely no admissible function that violates one or other of the inequalities at any value of $z$ or $\tau$ can lie in the identified set. Thus the inequalities (4) in general define an outer region within which $\tilde{\mathcal{H}}_a$ lies. This is the subject of Theorem 2.

When the outcome $Y$ is binary the inequalities do define the identified set, that is, all and only functions that satisfy the inequalities (4) lie in the identified set $\tilde{\mathcal{H}}_a$. This is the subject of Theorem 3. There is a discussion of sharp identification in the case when $Y$ has more than two points of support in Section 2.3.3.

**Theorem 2**. *Let $S_a$ be a structure admitted by the model $\mathcal{D}$ and delivering the distribution function $F^a_{YX|Z}$. 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 $F^a_{YX|Z}$. The following inequalities are satisfied.*

$$\text{For all } z \in \Omega \text{ and } \tau \in (0,1): \quad \begin{cases} \Pr_a[Y \leq h^*(X,\tau)|Z=z] \geq \tau \\ \\ \Pr_a[Y < h^*(X,\tau)|Z=z] < \tau \end{cases} \tag{5}$$

**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 $z \in \Omega$ 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, $F^*_{YX|Z} = F^a_{YX|Z}$ and the inequalities (5) follow on substituting "$\Pr_a$" for "$\Pr_*$".*

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

**Corollary**. *If the inequalities (5) are violated for any $(z,\tau) \in \Omega \times (0,1)$ then $h^* \notin \tilde{H}_a$.*

The consequence of these results is that for any probability measure $F^a_{YX|Z}$ generated by an admissible structure the set of functions that satisfy the inequalities (5) contains all members of the set of structural functions $\tilde{\mathcal{H}}_a$ identified by the model $\mathcal{D}$. When the outcome $Y$ is binary the sets are identical, a sharpness result which follows from the following Theorem.

**Theorem 3** *If $Y$ is binary and $h^*(x,u)$ satisfies the restrictions of the model $\mathcal{D}$ and the inequalities (5) 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 $F^a_{YX|Z}$.*

A proof of Theorem 3 is given in the Annex. The proof is constructive. For a given distribution $F^a_{YX|Z}$ and each value of $z \in \Omega$ and each structural function $h^*$ satisfying the inequalities (5) 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 $F^a_{YX|Z}$ at that value of $z$.

### 2.3. Discussion.

**2.3.1. Intersection bounds.** Let $\tilde{\mathcal{I}}_a(z)$ be the set of structural functions satisfying the inequalities (5) for all $\tau \in (0,1)$ at a value $z \in \Omega$. Let $\tilde{\mathcal{H}}_a(z)$ denote the set of structural functions identified by the model at $z \in \Omega$, that is $\tilde{\mathcal{H}}_a(z)$ contains the structural functions which lie in those structures admitted by the model that deliver the distribution $F^a_{YX|Z}$ for $Z = z$. When $Y$ is binary $\tilde{\mathcal{I}}_a(z) = \tilde{\mathcal{H}}_a(z)$ and otherwise $\tilde{\mathcal{I}}_a(z) \supseteq \tilde{\mathcal{H}}_a(z)$.

The identified set of structural functions, $\tilde{\mathcal{H}}_a$, defined by the model given a distribution $F^a_{YX|Z}$ is the *intersection* of the sets $\tilde{\mathcal{H}}_a(z)$ for $z \in \Omega$, and because for each $z \in \Omega$, $\tilde{\mathcal{I}}_a(z) \supseteq \tilde{\mathcal{H}}_a(z)$ the identified set is a subset of the set defined by the intersection of the inequalities (5), thus:

$$\tilde{\mathcal{H}}_a \subseteq \tilde{\mathcal{I}}_a = \left\{ h^* : \text{for all } \tau \in (0,1) \quad \left( \begin{array}{c} \min\limits_{z\in\Omega} \Pr_a[Y \leq h^*(X,\tau)|Z=z] \geq \tau \\ \\ \max\limits_{z\in\Omega} \Pr_a[Y < h^*(X,\tau)|Z=z] < \tau \end{array} \right) \right\} \tag{6}$$

with $\tilde{\mathcal{H}}_a = \tilde{\mathcal{I}}_a$ when the outcome is binary.

The set $\tilde{\mathcal{I}}_a$ can be estimated by calculating (6) using an estimate of the distribution of $F_{YX|Z}$. Chernozhukov, Lee and Rosen (2008) give results on inference in the presence of intersection bounds. There is an illustration in Chesher (2009).

**2.3.2. Strength and support of instruments.** It is clear from (6) that the support of the instrumental variables, $\Omega$, 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, $x_1, \ldots, x_K$, 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 $m \in \{1, \ldots, M\}$:[3]

$$p_m(x_{k^*}) = P[Y \leq m|Z = z^*]. \tag{7}$$

**2.3.3. Sharpness.** The inequalities of Theorem 1 define the identified set when the outcome is binary. When $Y$ has more than two points of support there may exist admissible functions that satisfy the inequalities but do not lie in the identified set. This happens when for a function, that satisfies the inequalities, say $h^*$, it is not possible to find an admissible distribution function $F^*_{UX|Z}$ which, when paired with $h^*$, delivers the "observed" distribution function $F^a_{YX|Z}$. In the three or more outcome case it is not possible, without further restriction, to characterise the identified set of structural functions using inequalities involving only the structural function; the distribution of observable variables, $F_{UX|Z}$, must feature as well. Section 3.1.3 gives an example based on a 3 outcome model.

When $X$ is continuous it is not feasible to compute the identified set without

[3]This 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 rediction 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 (7) on substituting $p = p_m(x_{k^*})$.

additional restrictions because in that case $F_{UX|Z}$ is infinite dimensional . A similar situation arises in the oligopoly entry game studies in Ciliberto and Tamer (2009). Some progress is possible when $X$ is discrete but if there are many points of support for $Y$ and $X$ then computations are infeasible without further restriction. Chesher and Smolinski (2009) give some results using parametric restrictions.

**2.3.4. Discreteness of outcomes.** The degree of discreteness in the distribution of $Y$ affects the extent of the identified set. The difference between the two probabilities in the inequalities (4) which delimit the identified set 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.

However, 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.

**2.3.5. 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 $\tau$, and $z \in \Omega$. It is straightforward to show that, with this amendment to the model, Theorems 1 and 2 hold for $\tau \in \tau_L$ from which results on set identification of $h(\cdot, \tau)$ for $\tau \in \tau_L$ can be developed.

## 3. ILLUSTRATIONS AND ELUCIDATION

This Section illustrates results of the paper with two examples. 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 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 second example employs a restrictive 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 $U$ is continuously distributed, normalised Uniform on $(0,1)$ and restricted to be distributed independently of instrumental variables $Z$. The model is as follows.

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

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.

$$\pi_1(z) \equiv P[Y = 0 \cap X = x_1|z] \qquad \pi_2(z) \equiv P[Y = 0 \cap X = x_2|z]$$

$$\beta_1(z) \equiv P[X = x_1|z] \qquad \beta_2(z) \equiv P[X = x_2|z]$$

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

**3.1.1. The identified set.** First, expressions are developed for the probabilities that appear in the inequalities (4) which, in this binary outcome case, define the identified set. With these in hand it is straightforward to characterise the identified set. The ordering of $\theta_1$ and $\theta_2$ 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.

$$P[Y < h(X,\tau)|z] = P[Y = 0 \cap p(X) < \tau|z] \tag{8}$$

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 (8) is zero. In the second case $p(X) < \tau$ only if $X = x_1$ and the probability (8) is therefore

$$P[Y = 0 \cap X = x_1|z] = \pi_1(z).$$

In the third case $p(X) < \tau$ whatever value $X$ takes and the probability (8) is therefore

$$P[Y = 0|z] = \pi_1(z) + \pi_2(z).$$

The situation is as follows.

$$P[Y < h(X,\tau)|z] = \begin{cases} 0 & , \quad 0 \leq \tau \leq \theta_1 \\ \pi_1(z) & , \quad \theta_1 < \tau \leq \theta_2 \\ \pi_1(z) + \pi_2(z) & , \quad \theta_2 < \tau \leq 1 \end{cases}$$

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 $\tau$ in the interval to which it relates and in particular must not exceed the *minimum* value of $\tau$ in that interval. The result is the following pair of inequalities.

$$\pi_1(z) \leq \theta_1 \qquad \pi_1(z) + \pi_2(z) \leq \theta_2 \tag{9}$$

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.

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

Again 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

$$P[Y \leq h(X,\tau)|z] = \pi_1(z) + \pi_2(z)$$

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

$$P[Y \leq h(X,\tau)|z] = \beta_1(z) + \pi_2(z)$$

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.

$$P[Y \leq h(X,\tau)|z] = \begin{cases} \pi_1(z) + \pi_2(z) & , \quad 0 \leq \tau \leq \theta_1 \\ \beta_1(z) + \pi_2(z) & , \quad \theta_1 < \tau \leq \theta_2 \\ 1 & , \quad \theta_2 < \tau \leq 1 \end{cases}$$

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 $\tau$ in the interval to which it relates and in particular must at least equal the *maximum* value of $\tau$ in that interval. The result is the following pair of inequalities.

$$\theta_1 \leq \pi_1(z) + \pi_2(z) \qquad \theta_2 \leq \beta_1(z) + \pi_2(z) \tag{10}$$

Bringing (9) and (10) 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.

$$\pi_1(z) \leq \theta_1 \leq \pi_1(z) + \pi_2(z) \leq \theta_2 \leq \beta_1(z) + \pi_2(z) \tag{11}$$

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

$$\pi_2(z) \leq \theta_2 \leq \pi_1(z) + \pi_2(z) \leq \theta_1 \leq \pi_1(z) + \beta_2(z) \tag{12}$$

and the identified set for the case in which $Z = z$ is the union of the sets defined by the inequalities (11) and (12). 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 $\theta_1 = \theta_2 = \pi_1(z) + \pi_2(z)$.

There is one such set for each value of $z$ in $\Omega$ 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.

**3.1.2 Sharpness.** The set just derived is precisely the identified set - that is, for every value $\Theta$ in the 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 set. The existence of such a distribution is now demonstrated.

Consider some value $z$ and a value $\Theta^* \equiv \{\theta_1^*, \theta_2^*\}$ with, say, $\theta_1^* \leq \theta_2^*$, which satisfies the inequalities (11), 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.

$$\begin{array}{ll} F^*_{U|XZ}(\theta_1^*|x_1, z) \equiv \pi_1(z)/\beta_1(z) & F^*_{U|XZ}(\theta_1^*|x_2, z) \equiv (\theta_1^* - \pi_1(z))/\beta_2(z) \\ F^*_{U|XZ}(\theta_2^*|x_1, z) \equiv (\theta_2^* - \pi_2(z))/\beta_1(z) & F^*_{U|XZ}(\theta_2^*|x_2, z) \equiv \pi_2(z)/\beta_2(z) \end{array} \tag{13}$$

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.[4] The proposed distribution respects the independence restriction because the implied probabilities marginal with respect to $X$ are independent of $z$, as follows.

$$P[U \leq \theta_1^*|z] = \beta_1(z) F^*_{U|XZ}(\theta_1^*|x_1, z) + \beta_2(z) F^*_{U|XZ}(\theta_1^*|x_2, z) = \theta_1^*$$

$$P[U \leq \theta_2^*|z] = \beta_1(z) F^*_{U|XZ}(\theta_2^*|x_1, z) + \beta_2(z) F^*_{U|XZ}(\theta_2^*|x_2, z) = \theta_2^*$$

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

$$\pi_1(z) \leq \theta_1^* \leq \pi_1(z) + \beta_2(z)$$

$$\pi_2(z) \leq \theta_2^* \leq \beta_1(z) + \pi_2(z)$$

[4]This is 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 $\theta_1^*$ and $\theta_2^*$ satisfy the inequalities (11). 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.

$$\begin{array}{rcl} F^*_{U|XZ}(\theta_1^*|x_1, z) & \leq & F^*_{U|XZ}(\theta_2^*|x_1, z) \\ F^*_{U|XZ}(\theta_1^*|x_2, z) & \leq & F^*_{U|XZ}(\theta_2^*|x_2, z) \end{array}$$

Manipulating the expressions in (13) yields the result that these inequalities are satisfied if:

$$\theta_1^* \leq \pi_1(z) + \pi_2(z) \leq \theta_2^*$$

which is assured when $\theta_1^*$ and $\theta_2^*$ satisfy the inequalities (11). There is a similar argument for the case $\theta_2^* \leq \theta_1^*$.

This argument above applies at each value $z \in \Omega$ so it can be concluded that for each value $\Theta^*$ in the set formed by intersecting sets obtained at each $z \in \Omega$ 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 sets.

**3.1.3 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.

$$Y^* = a_0 + a_1 X + \varepsilon$$

$$X^* = b_0 + b_1 Z + \eta$$

Latent variates $\varepsilon$ and $\eta$ are jointly normally distributed conditional on an instrumental variable $Z$.

$$\left[\begin{array}{c} \varepsilon \\ \eta \end{array}\right] |Z \sim N\left(\left[\begin{array}{c} 0 \\ 0 \end{array}\right], \left[\begin{array}{cc} 1 & r \\ r & 1 \end{array}\right]\right)$$

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

$$Y = \left\{ \begin{array}{ccrcl} 0 & , & 0 < & U & \leq p(X) \\ 1 & , & p(X) < & U & \leq 1 \end{array} \right.$$

with $U \equiv \Phi(\varepsilon) \sim Unif(0,1)$ and $U \perp\!\!\!\perp Z$ and $p(X) = \Phi(-a_0 - a_1 X)$ with $X \in \{0, 1\}$.

Figure 1 shows identified sets when the parameter values generating the probabilities are: $a_0 = 0$, $a_1 = 0.5$, $b_0 = 0$, $b_1 = 1$, $r = -0.25$, for which:

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

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

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 $p(1) > p(0)$.[5]

**3.2. Three valued outcomes.** When the outcome has more than two points of support the inequalities of Theorem 1 define an outer region within which the set of structural functions identified by the model lies. This is demonstrated in a *three* outcome case:

$$Y = h(X,U) \equiv \begin{cases} 0 \quad , & 0 < \quad U \quad \leq p_1(X) \\ 1 \quad , & p_1(X) < \quad U \quad \leq p_2(x) \\ 2 \quad , & p_2(X) < \quad U \quad \leq 1 \end{cases} , \qquad U \perp\!\!\!\perp Z \in \Omega, \qquad U \sim Unif(0,1)$$

with $X$ binary, taking values in $\{x_1, x_2\}$ as before.

The structural features whose identification is of interest are now:

$$\theta_{11} \equiv p_1(x_1) \quad \theta_{12} \equiv p_1(x_2) \quad \theta_{21} \equiv p_2(x_1) \quad \theta_{22} \equiv p_2(x_2)$$

and the probabilities about which data are informative are:

$$\begin{array}{ll} \pi_{11}(z) \equiv P[Y = 0 \cap X = x_1|z] & \pi_{12}(z) \equiv P[Y = 0 \cap X = x_2|z] \\ \pi_{21}(z) \equiv P[Y \leq 1 \cap X = x_1|z] & \pi_{22}(z) \equiv P[Y \leq 1 \cap X = x_2|z] \end{array} \tag{14}$$

and $\beta_1(z)$ and $\beta_2(z)$ as before.

Consider putative values of parameters which fall in the following order.

$$\theta_{11} < \theta_{12} < \theta_{22} < \theta_{21}$$

[5]In supplementary material more extensive graphical displays are available.

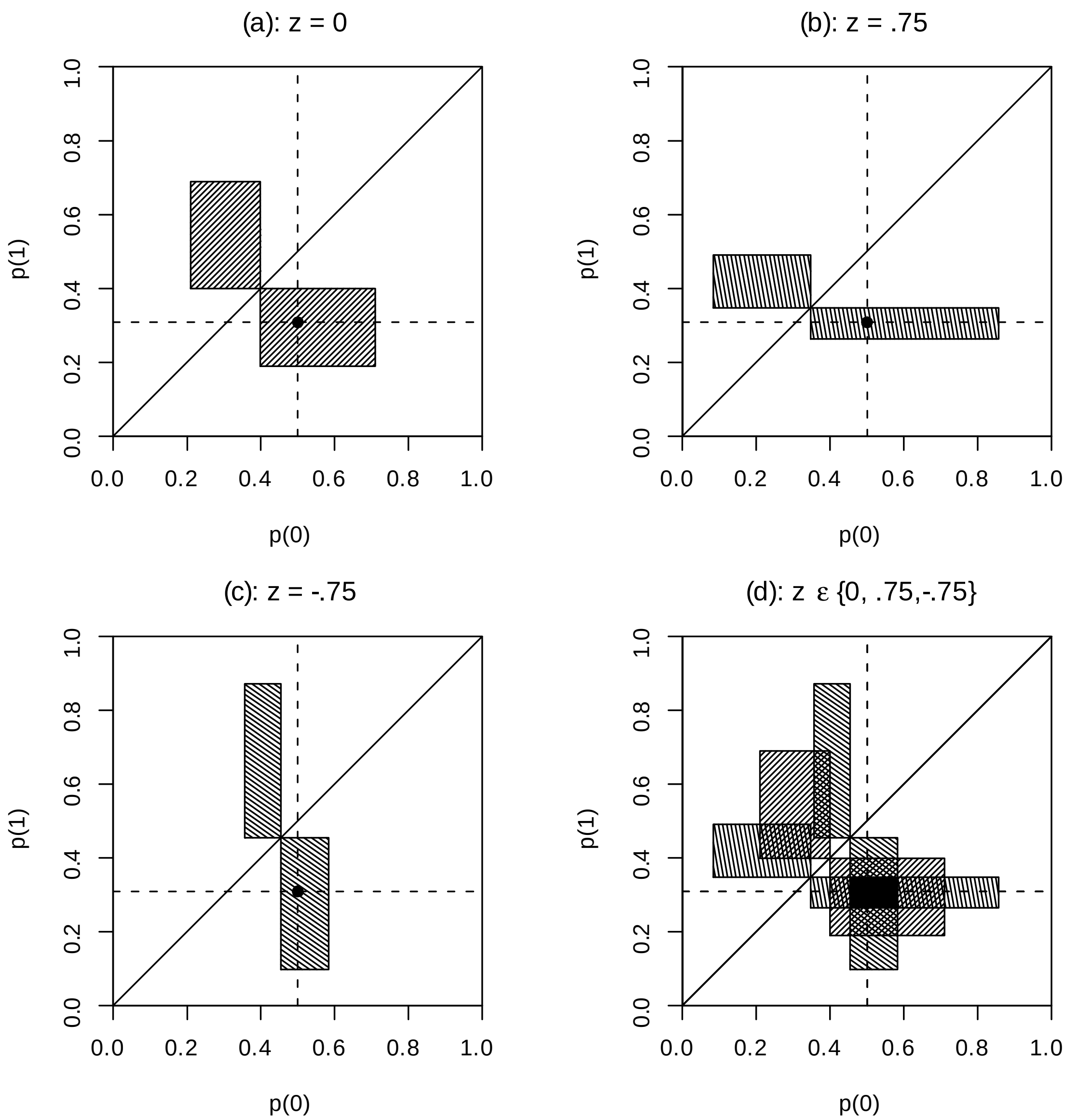


Figure 1: Identified sets with a binary outcome and binary endogenous variable as instrumental values, $z$, vary. Strong instrument ($b_1 = 1$). Dotted lines intersect at the values of $p(0)$ and $p(1)$ in the distribution generating structure. Panes (a) - (c) show identified sets at each of 3 values 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 $p(1) > p(0)$.

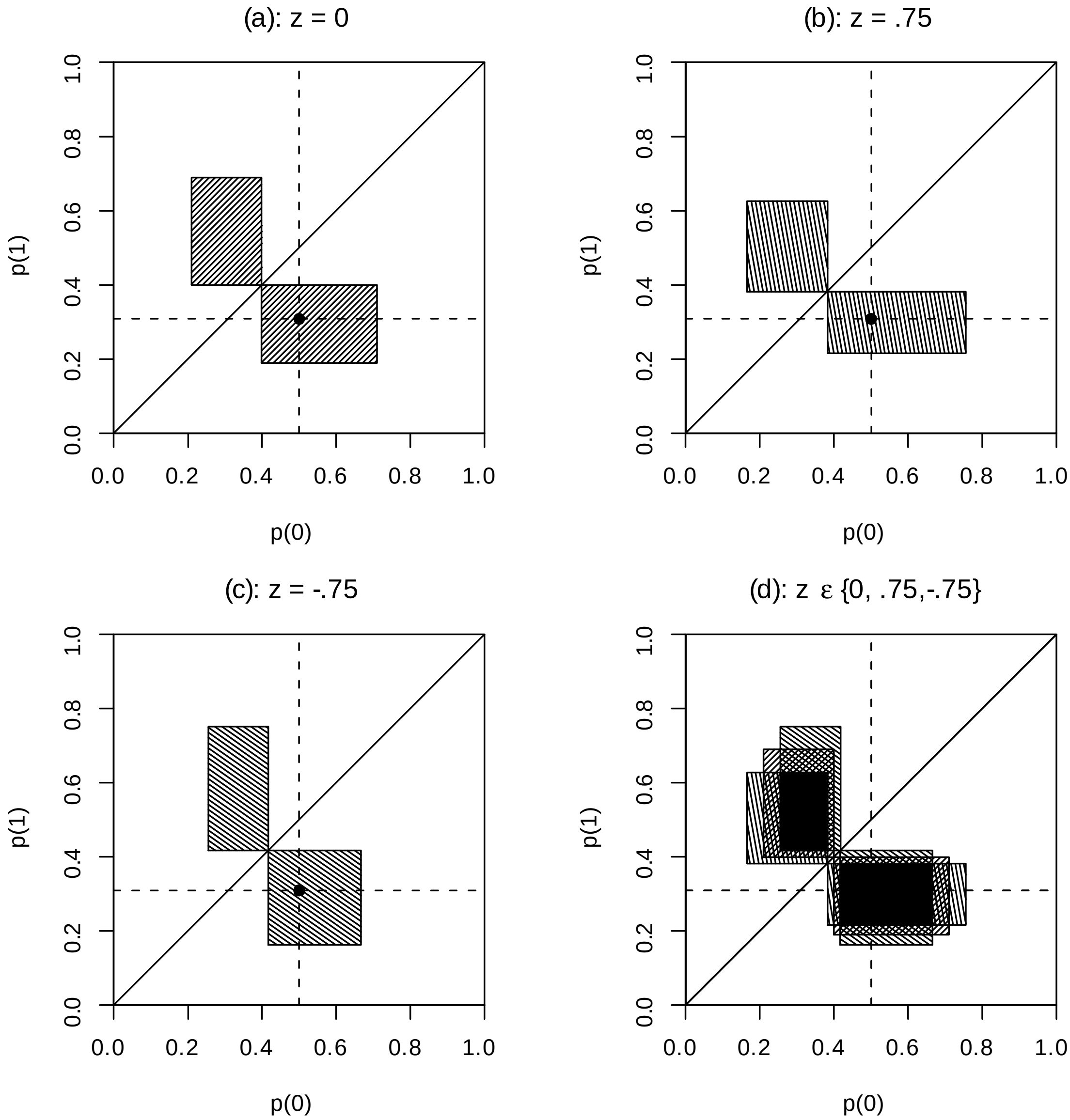


Figure 2: Identified sets with a binary outcome and binary endogenous variable as instrumental values, $z$, vary. Weak instrument ($b_1 = 0.3$). Dotted lines intersect at the values of $p(0)$ and $p(1)$ in the distribution generating structure. Panes (a) - (c) show identified sets at each of 3 values 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 $p(1) > p(0)$.

The inequalities of Theorem 1 place the following restrictions on the $\theta$'s.

$$\pi_{11}(z) < \theta_{11} \leq \pi_{11}(z) + \pi_{12}(z) < \theta_{12} \leq \pi_{21}(z) + \pi_{12}(z) \tag{15a}$$
$$\pi_{11}(z) + \pi_{22}(z) < \theta_{22} \leq \pi_{21}(z) + \pi_{22}(z) < \theta_{21} \leq \pi_{21}(z) + \beta_2(z) \tag{15b}$$

However, when determining whether it is possible to construct a proper distribution $F_{UX|Z}$ exhibiting independence of $U$ and $Z$ and delivering the probabilities (14) it is found that the following inequality is required to hold

$$\theta_{22} - \theta_{12} \geq \pi_{22}(z) - \pi_{12}(z)$$

and this is not implied by the inequalities (15).

This inequality *and* the inequality

$$\theta_{21} - \theta_{11} \geq \pi_{21}(z) - \pi_{11}(z)$$

are required when the ordering $\theta_{11} < \theta_{21} < \theta_{12} < \theta_{22}$ is considered. However in the case of the ordering $\theta_{11} < \theta_{12} < \theta_{21} < \theta_{22}$ the inequalities of Theorem 1 guarantee that both of these inequalities hold. So, if there were the additional restriction that this latter ordering prevails then the inequalities of Theorem 1 would define the identified set.[6]

**3.3. 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, $\Omega$. 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 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 $\alpha \equiv (\alpha_0, \alpha_1)$,

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

where, for $m \in \{1, \dots, M\}$, with $\Phi$ 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

[6] There are six feasible permutations of the $\theta$'s of which three are considered in this Section, the other three being obtained by exchange of the second index.

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

For a conditional probability function $F_{Y|XZ}$ and a conditional density $f_{X|Z}$ and some value $\alpha$ the probabilities in (4) are:

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

$$\Pr[Y < h(X, \tau; \alpha)|Z = z] = \sum_{m=2}^{M} \int_{\{x\,:\,h(x,\tau;\alpha)=m\}} F_{Y|XZ}(m-1|x, z) f_{X|Z}(x|z)dx \qquad (17)$$

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

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

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

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

Here $c_0 \equiv -\infty$, $c_M \equiv \infty$ and $c_1, \ldots, c_{M-1}$ 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 (16) and (17) are calculated for each choice of $\alpha$ by numerical integration.[7] 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 $\Omega \equiv [-1, 1]$ , 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, equivalently the support of the instrument, varies.

Figure 3 shows the set defined by the inequalities of Theorem 1 for the intercept and slope coefficients, $\alpha_0$ and $\alpha_1$ 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 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 sets are shown in outline. The effect of reducing the

[7]The `integrate` procedure in R (Ihaka and Gentleman (1996)) was used to calculate probabilities. Intersection bounds over $z \in \Omega_Z$ were obtained as in (6) using the R function `optimise`. The resulting probability inequalties were inspected over a grid of values of $\tau$ at each value of $\alpha$ considered, a value being classified as out of the identified set as soon as a value of $\tau$ was encountered for which there was violation of one or other of the inequaltites (6). I am grateful to Konrad Smolinski for developing and programming a procedure to efficiently track out the boundaries of the sets.

discreteness of the outcome is substantial and there is a substantial reduction in the extent of the set as the instrument is strengthened.

The sets portrayed here are outer regions which contain the sets identified by the model. The identified sets are computationally challenging to produce in this continuous endogenous variable case. Chesher and Smolinski (2009) investigate feasible procedures based on discrete approximations.

## 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 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 the outcome and 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.[8] This is the control function approach studied in Blundell and Powell (2004), Chesher (2003) and Imbens and Newey (2009). 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).

There are a number of interesting extensions. For example the analysis can be extended to the *multiple* discrete outcome case such as arises in the study of panel data. Consider a model for $T$ discrete outcomes each determined by a structural equation as follows:

$$Y_t = h_t(X, U_t), \qquad t = 1, \ldots, T$$

where each function $h_t$ is weakly increasing and càglàd for variations 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 $Z \in \Omega$ 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.

[8]But not when endogenous variables are discrete, Chesher (2005).

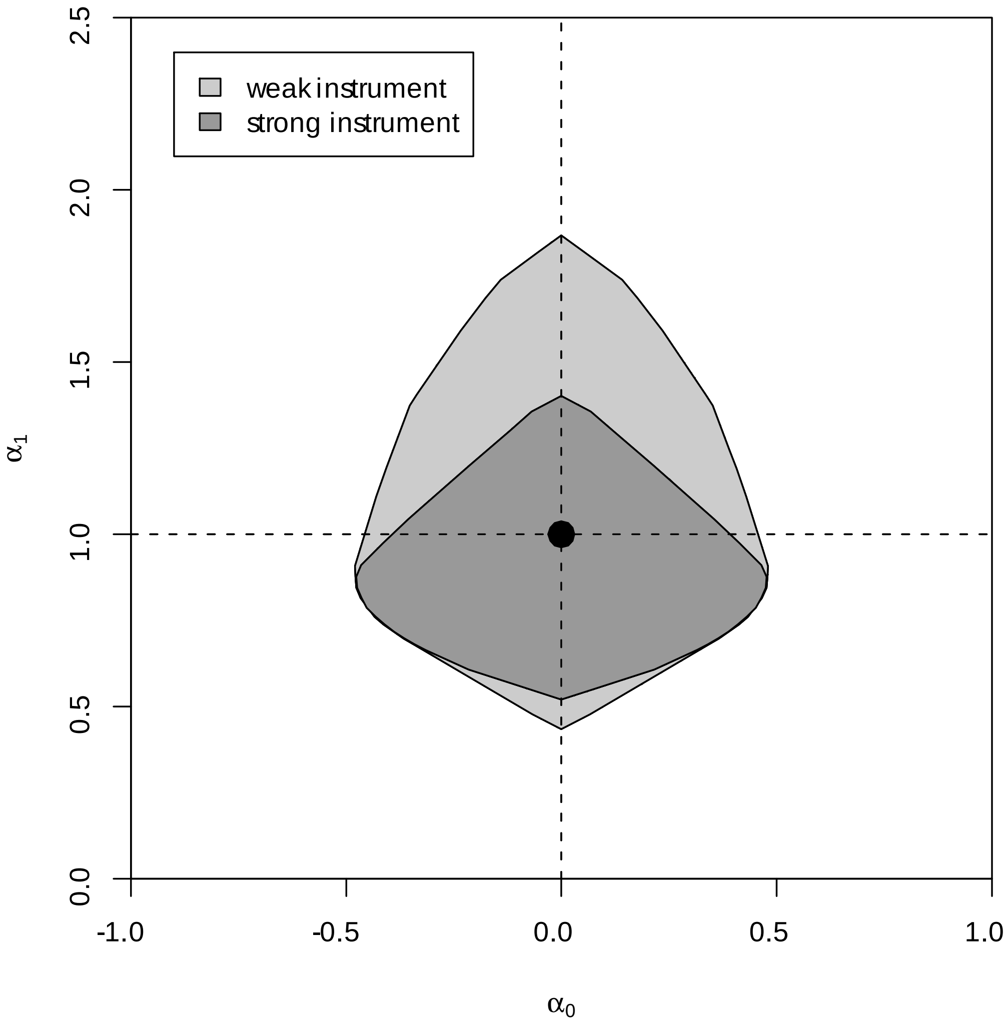


Figure 3: Outer regions within which lie identified sets for an intercept, $\alpha_0$, and slope cofficient, $\alpha_1$, in a 5 class ordered probit model with endogenous explanatory variable. The dashed lines intersect at the values of $a_0$ and $a_1$ used to generate the distributions employed in this illustration.

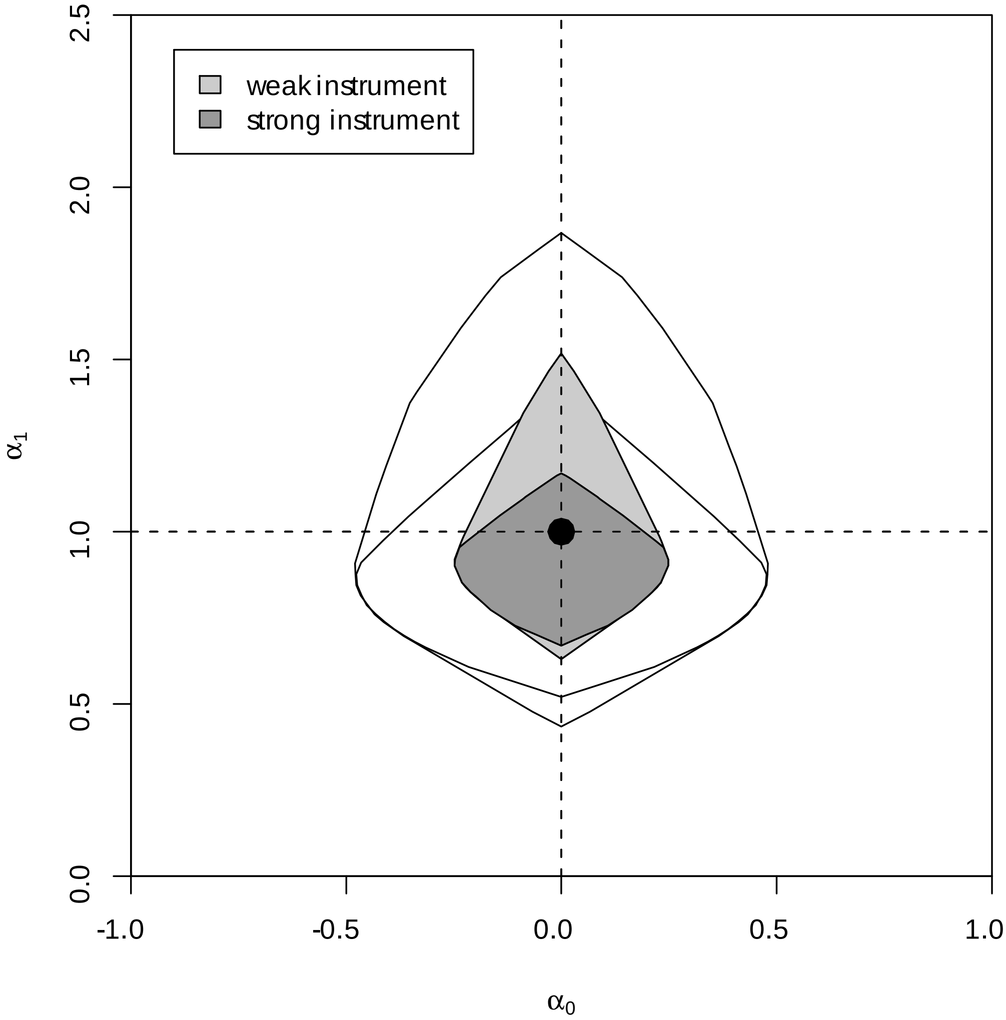


Figure 4: Outer regions within which lie identified sets for an intercept, $\alpha_0$, and slope cofficient, $\alpha_1$, in a 11 class ordered probit model with endogenous explanatory variable. Outer regions for the 5 class model displayed in Figure 3 are shown in outline. The dashed lines intersect at the values of $a_0$ and $a_1$ used to generate the distributions employed in this illustration.

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 inequalities which hold for all $\tau \in [0,1]^T$ and $z \in \Omega$.

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

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

These can be used to delimit the sets of structural function and copula combinations $\{h, C\}$ identified by the model.

Other extensions arise on relaxing restrictions maintained 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, Konrad Smolinski and Richard Spady for stimulating comments and discussions and referees for very helpful comments. The support of the Leverhulme Trust through grants to the research project *Evidence Inference and Inquiry* and 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 since June 2007 is acknowledged. This is a revised and corrected 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.

## References

Beresteanu, Arie, Molchanov, Ilya and Francesca Molimari (2008): "Sharp Identification Regions in Games," CeMMAP Working Paper, CWP15/08.

Berry, Steven and Elie Tamer (2006): "Identification in Models of Oligopoly Entry," in *Advances in Economics and Econometrics: Theory and Applications: Ninth World Congress, vol 2*, R. Blundell, W.K. Newey and T. Persson, eds, Cambridge University Press.

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.,(2009): "Single Equation Endogenous Binary Response Models," CeMMAP Working Paper, CWP16/09.

Chesher, Andrew D., and Konrad Smolinski (2009): "Set Identifying Endogenous Ordered Response Models," in preparation.

Ciliberto, Federico and Elie Tamer (2009): "Market Structure and Multiple Equilibria in Airline Markets," forthcoming in *Econometrica*.

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 (2009): "Identification and estimation of triangular simultaneous equations models without additivity," forthcoming in *Econometrica*.

Koenker, R.W., (2005): *Quantile Regression.* Econometric Society Monograph No. 38. Cambridge University Press, Cambridge.

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.

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 Theorem 3. Sharp set identification for binary outcomes

The proof proceeds by considering a structural function $h(x, u)$, that: (i) is weakly monotonic non-decreasing for variations in $u$, (ii) is characterised by a threshold function $p(x)$, and (iii) 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 such that $U$ and $Z$ are independent and with 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.

Attention is directed to 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. It is helpful to introduce some abbreviated notation. At many points dependence on $z$ is not made explicit in the notation.

Let $\Psi$ denote the support of $X$ conditional on $Z$. $Y$ is binary taking values in $\{y_1, y_2\}$. Define conditional probabilities as follows.

$$\alpha_1(x) \equiv \Pr[Y = y_1|x, z]$$

$$\alpha_1 \equiv \Pr[Y = y_1|z] = \int \alpha_1(x) dF_{X|Z}(x|z)$$

and $\alpha_2(x) \equiv 1 - \alpha_1(x)$, $\alpha_2 \equiv 1 - \alpha_1$ and note that dependence of $\Psi$, $\alpha_1(x)$, $\alpha_2(x)$, etc., on $z$ is not made explicit in the notation.

A threshold function $p(x)$ is proposed such that

$$Y = \begin{cases} y_1 \quad , & 0 \leq \ U \ \leq p(x) \\ y_2 \quad , & p(x) < \ U \ \leq 1 \end{cases}$$

and this function satisfies some inequalities to be stated. The threshold function is a continuous function of $x$ and does not depend on $z$.

Define the following functions which in general depend on $z$.

$$u_1(v) = \min(v, \alpha_1) \qquad u_2(v) = v - u_1(v)$$

Define sets as follows:

$$X(s) \equiv \{x : p(x) = s\} \qquad X[s] \equiv \{x : p(x) \leq s\}$$

and let $\phi$ denote the empty set. Define

$$s_1(v) \equiv \min_s \left\{ s : \int_{x \in X[s]} \alpha_1(x) dF_{X|Z}(x|z) = u_1(v) \right\}$$

and:

$$s_2(v) \equiv \min_s \left\{ s : \int_{x \in X[s]} \alpha_2(x) dF_{X|Z}(x|z) = u_2(v) \right\}$$

and define functions $\beta_1(v,x)$ and $\beta_2(v,x)$.

$$\beta_1(v,x) \equiv \begin{cases} \alpha_1(x) & , \quad x \in X[s_1(v)] \\ 0 & , \quad x \notin X[s_1(v)] \end{cases}$$

$$\beta_2(v,x) \equiv \begin{cases} \alpha_2(x) & , \quad x \in X[s_2(v)] \\ 0 & , \quad x \notin X[s_2(v)] \end{cases}$$

For a structural function $h(x,u)$ characterized by the threshold function $p(x)$ and for a probability measure that delivers $\alpha_1(x)$ and $F_{X|Z}$, a distribution function $F_{U|XZ}$ is defined as

$$F_{U|XZ}(u|x,z) \equiv \beta(u,x) \equiv \beta_1(u,x) + \beta_2(u,x)$$

where $z$ is the value of $Z$ upon which there is conditioning at various points in the definition of $\beta(u,x)$.

Consider functions $p$ that satisfy the inequalities of Theorem 1 which in this binary outcome case can be expressed as follows.

$$\text{for all } u \in (0,1) : \quad \int_{p(x)<u} \alpha_1(x) dF_{X|Z}(x|z) < u \leq \alpha_1 + \int_{p(x)<u} \alpha_2(x) dF_{X|Z}(x|z) \quad \text{(A1)}$$

It is now shown that:

1. for all $x$ and any $z$ the distribution function $\beta(u,x)$ is proper: (a) $\beta(0,x) = 0$, (b) $\beta(1,x) = 1$, (c) for $v' > v \qquad \beta(v',x) \geq \beta(v,x)$.
2. there is an independence property:
$$\text{for all } u \quad \int_{x \in \Psi} \beta(u,x) dF_{X|Z}(x|z) = u$$
3. if $p$ satisfies the inequalities (A1) then there is an observational equivalence property: for all $x$
$$\beta(p(x),x) = \alpha_1(x).$$

**(1a). Proper distribution:** $\beta(0,x) = 0$.

By definition $u_1(0) = u_2(0) = 0$ and so $s_1(0) = s_2(0) = 0$. Therefore $X[s_1(0)] = X[s_2(0)] = \phi$ which implies that, for all $x$, $\beta_1(0,x) = \beta_2(0,x) = 0$ and so $\beta(0,x) = 0$.

**(1b). Proper distribution:** $\beta(1,x) = 1$.

By definition $u_1(1) = \alpha_1$, so $s_1(1)$ is the smallest value of $s$ such that $X[s] = \Psi$ so $s = \max_{x \in \Psi} p(x)$. With $X[s_1(1)] = \Psi$ it is assured that $\beta_1(1,x) = \alpha_1(x)$ for all $x$.

By definition: $u_2(1) = \alpha_2$, so $s_2(1)$ is the smallest value of $s$ such that $X[s] = \Psi$. With $X[s_2(1)] = \Psi$ it is assured that $\beta_2(1,x) = \alpha_2(x)$ for all $x$. So, for all $x$, $\beta(1,x) = \alpha_1(x) + \alpha_2(x) = 1$.

**(1c). Proper distribution: nondecreasing** $\beta(u,x)$.

Since $u_1(v)$ and $u_2(v)$ are nondecreasing functions of $v$, so are $s_1(v)$ and $s_2(v)$. It follows that for $v' > v$

$$X[s_1(v')] \supseteq X[s_1(v)] \qquad X[s_2(v')] \supseteq X[s_2(v)]$$

and so for all $x$

$$\beta_1(v',x) \geq \beta_1(v,x) \qquad \beta_2(v',x) \geq \beta_2(v,x)$$

and it follows that the sum of the functions, $\beta(v,x)$, is a nondecreasing function of $v$.

**(2). Independence.**

By definition

$$\int_{x\in\Psi} \beta_1(v,x)dF_{X|Z}(x|z) = \int_{x\in X[s_1(v)]} \alpha_1(x)dF_{X|Z}(x|z) = u_1(v)$$

$$\int_{x\in\Psi} \beta_2(v,x)dF_{X|Z}(x|z) = \int_{x\in X[s_2(v)]} \alpha_2(x)dF_{X|Z}(x|z) = u_2(v)$$

and so

$$\int_{x\in\Psi} \beta(v,x)dF_{X|Z}(x|z) = u_1(v) + u_2(v) = v$$

which does not depend upon $z$.

**(3). Observational equivalence.**

This requires that for all $x$: $\beta(p(x),x) = \alpha_1(x)$ which is true if for all $x$: (a) $\beta_1(p(x),x) = \alpha_1(x)$ and (b) $\beta_2(p(x),x) = 0$. Each equation is considered in turn. The inequalities (A1) come into play.

(a) $\beta_1(p(x),x) = \alpha_1(x)$.
Since for all $u$

$$\int_{p(x)<u} \alpha_1(x)dF_{X|Z}(x|z) < u$$

there exists $\delta(u) > 0$ such that

$$\int_{p(x)<u+\delta(u)} \alpha_1(x)dF_{X|Z}(x|z) = u$$

for all $u \leq \alpha_1$.

It follows that $s_1(v) > v$ for all $v$ which implies $X[v] \subset X[s_1(v)]$ and in particular $X(v) \subset X[s_1(v)]$.

For some value of $x$, $x^*$, define $p^* \equiv p(x^*)$. Then for $p^* \in (0,1)$,

$$X(p^*) \equiv \{x : p(x) = p^*\} \subset X[p^*] \subset X[s_1(p^*)].$$

Recall that $\beta_1(v,x)$ is equal to $\alpha_1(x)$ for all $x \in X[s_1(v)]$. It has been shown that for any $p^*$, all $x$ such that $p(x) = p^*$ lie in the set $X[s_1(p^*)]$ and so $\beta_1(p^*,x^*) = \alpha_1(x^*)$ and there is the result $\beta_1(p(x),x) = \alpha_1(x)$.

(b) $\beta_2(p(x),x) = 0$.

Recall $X(v) \equiv \{x : p(x) = v\}$. It is required to show that $X(v) \cap X[s_2(v)]$ is empty for all $v$.

Since $u_2(v) = 0$ for $v \le \alpha_1$, $s_2(v) = 0$ for $v \le \alpha_1$ and $X[s_2(v)] = \phi$ for $v \le \alpha_1$. Therefore, for $v \le \alpha_1$

$$X(v) \cap X[s_2(v)] = X(v) \cap \phi = \phi.$$

From (A1) there is the following inequality.

$$\int\limits_{p(x)<v} \alpha_2(x) dF_{X|Z}(x|z) \ge v - \alpha_1$$

For $v > \alpha_1$, the constraint implies that there exists $\gamma(v) \ge 0$ such that

$$\int\limits_{p(x)<v-\gamma(v)} \alpha_2(x) dF_{X|Z}(x|z) = v - \alpha_1$$

and, since for $v > \alpha_1$

$$s_2(v) \equiv \min_s \left\{ s : \int_{p(x)<s} \alpha_2(x) dF_{X|Z}(x|z) = v - \alpha_1 \right\}$$

it follows that $s_2(v) < v$. It follows that for $v > \alpha_1$, $X(v) \cap X[s_2(v)] = \phi$ because with $s_2(v) < v$:

$$\{x : p(x) < s_2(v)\} \cap \{x : p(x) = v\} = \phi$$

Define $p^* = p(x^*)$. Then for $p^* \in (0,1)$, $X(p^*) \cap X[s_2(p^*)] = \phi$ so $\beta_2(p^*,x^*) = 0$ and there is the result, for all $x$, $\beta_2(p(x),x) = 0$.