Econometric Theory, **29**, 2013, 231–266.
doi:10.1017/S0266466612000321

SEMIPARAMETRIC STRUCTURAL MODELS OF BINARY RESPONSE: SHAPE RESTRICTIONS AND PARTIAL IDENTIFICATION

ANDREW CHESHER
CeMMAP
and
University College London

I study the partial identifying power of structural single-equation threshold-crossing models for binary responses when explanatory variables may be endogenous. The sharp identified set of threshold functions is derived for the case in which explanatory variables are discrete, and I provide a constructive proof of sharpness. There is special attention to a widely employed semiparametric shape restriction, which requires the threshold-crossing function to be a monotone function of a linear index involving the observable explanatory variables. The restriction brings great computational benefits, allowing calculation of the identified set of index coefficients without calculating the nonparametrically specified threshold function. With the restriction in place, the methods of the paper can be applied to produce identified sets in a class of binary response models with mismeasured explanatory variables.

1. INTRODUCTION

This paper gives new results on the identifying power of a single-equation, limited information, instrumental variable (IV) model for a binary response in a structural econometric model that allows explanatory variables to be endogenous. The model involves a scalar continuously distributed latent variable U and a threshold-crossing function $p(X)$ that depends on a vector random variable X, some of whose elements may be jointly dependently distributed with U.

In this threshold-crossing model realizations of U exceeding $p(X)$ lead to the outcome $Y = 1$; realizations less than or equal to $p(X)$ lead to $Y = 0$. The marginal distribution of U is normalized uniform on the unit interval. Realizations of (Y, X, Z) are observed, where Z is a list of instrumental variables excluded

I thank Martin Cripps, Lars Nesheim, Adam Rosen, and Richard Spady for stimulating comments and discussions and Konrad Smolinski for excellent research assistance. Some of the results of this paper were presented at seminars at Caltech, UCLA, and USC in November 2007, at the Malinvaud Seminar in Paris in December 2007, and at the inaugural Asian Econometric Theory Lecture at the SETA Conference, Kyoto, Japan, July 30, 2009. I thank participants, two referees, and a co-editor for helpful comments. I gratefully acknowledge the financial support of the UK Economic and Social Research Council through a grant (RES-589-28-0001) to the ESRC Centre for Microdata Methods and Practice (CeMMAP). Address correspondence to Andrew Chesher, University College London, Gower St., London WC1E 6BT, United Kingdom; e-mail: andrew.chesher@ucl.ac.uk.

© Cambridge University Press 2012

https://doi.org/10.1017/S0266466612000321 Published online by Cambridge University Press

from the threshold-crossing function. The structural latent variable U is restricted to be independent of Z in the sense that $\Pr[U \leq u|Z = z]$ is independent of z for all u in the unit interval and for all z in some support set $\mathcal{Z}$.

The model is partially identifying for the threshold-crossing function p and for its parameters if semiparametric or parametric restrictions are imposed. The identified set can be topologically complex. In the discrete endogenous variable case it is a union of convex sets but may not itself be convex, nor even connected.

The paper derives the sharp identified set for the case in which endogenous variables are discrete and illustrates the identified set in this case. The paper studies the impact of a very commonly imposed monotonicity restriction on the function p when there is an index restriction on the way in which explanatory variables X affect the function. This shape restriction brings the substantial computational advantage that the identified set for the index coefficients can be obtained without calculating the identified set of threshold-crossing functions, p. One possible cause of endogeneity is mismeasurement of explanatory variables. It is shown how the methods of the paper can be applied to characterize identified sets when there is additive measurement error and a monotone index restriction.

1.1. Alternative Models Built on Conditional Independence Restrictions

The leading alternative to the IV model of this paper is the widely used triangular equation system model that motivates the use of control function methods.

This model is studied in parametric contexts in Rivers and Vuong (1988) and Smith and Blundell (1986) and in semi- and nonparametric settings in, for example, Blundell and Powell (2003, 2004), Chesher (2003, 2005, 2007), Florens, Heckman, Meghir, and Vytlacil (2008), Hoderlein (2009), and Imbens and Newey (2009). There are early examples of the use of control function methods in econometrics in Hausman (1978) and Heckman (1979). STATA 10 (Statacorp, 2007) and LIMDEP 9.0 (Greene, 2007) have commands to perform parametric probit control function estimation in triangular models for binary responses.

The triangular system control function model is attractive because under some conditions it is point identifying. One requirement is that, in the absence of a fully parametric specification such as is employed in Heckman (1978), the endogenous variables are determined by a structural function, $X = g(Z, V)$, in which there is a one-to-one mapping from latent variables, V, to the endogenous variables, X, at each value of the instrumental variables, Z. This ensures that there exists a single-valued control function $g^{-1}(Z, X)$ that delivers the value of V that causes a particular value of X to arise for some value of Z. The restriction rules out cases in which there are discrete endogenous variables or V is high-dimensional such as arises when there is a random coefficients model for X.

The control function model also requires that the latent variables (U, V) be jointly distributed independently of the instrumental variables. This, along with triangularity, ensures that U and X are independent conditional on V. Blundell

https://doi.org/10.1017/S0266466612000321 Published online by Cambridge University Press

and Matzkin (2010) demonstrate that it is only under very special conditions that a nonlinear simultaneous equations model for the joint determination of Y and X satisfies the restrictions of the triangular system model.

Other types of conditional independence restrictions have been studied. For example, Lewbel (2000) studies threshold-crossing models for binary outcomes with an index restriction and the requirement that the unobservable is distributed independently of a continuously distributed explanatory variable (a "special regressor"), conditional on other explanatory and instrumental variables. The model is point identifying when the special regressor has large support. Magnac and Maurin (2007) replace the large support restriction with an alternative that delivers point identification, and Magnac and Maurin (2008) relax the support restrictions on the special regressor and obtain a partially identifying model. Vytlacil and Yildiz (2007) propose a triangular model that partially identifies average effects of a binary endogenous variable, and Shaikh and Vytlacil (2011) provide additional results for the case in which the outcome of interest is binary.

1.2. Benefits and Costs of the IV Approach

A virtue of the IV model studied in this paper is that it is less restrictive than the triangular system model and it does not rely on knowledge of a special regressor. But the IV model is essentially incomplete in the sense that it does not specify a structural function for the endogenous variables. As a result, the model is generally partially identifying for certain structural features.[1] Inference obtained using the IV model is robust to failure of some of the conditional independence restrictions commonly employed in structural binary response models, but the model does not deliver unambiguous information about some structural features.

The IV model is encompassing for some of the conditional independence-based alternatives, for example, the triangular system model and the special regressor model with instrument independence,[2] in the sense that the values of structural features that those models point identify are points in the sets identified by the IV model.

The sets identified by the IV model can be large. In this circumstance the restrictions of the conditional independence-based models are extremely informative, and, in the context of the partially identifying IV model, not falsifiable. The results delivered in this paper open the door to calculations that expose the possible fragility of inferences based on point-identifying conditional independence restrictions.

In some cases, even when the identified sets delivered by the IV model are large they can deliver useful information about the structural threshold function. For example, it will always be possible to falsify the hypothesis that the threshold function is insensitive to variation in potentially endogenous explanatory variables; calculations reported here show that, even with relatively weak instruments, hypotheses about the monotonicity of structural threshold functions and the direction of dependence can be falsifiable.

https://doi.org/10.1017/S0266466612000321 Published online by Cambridge University Press

1.3. Related Results on IV Models

Outer regions for threshold-crossing functions in IV models for scalar ordered outcomes, of which the binary outcome model is a special case, are given in Chesher (2010). The sharp identified set for a threshold-crossing function in a binary outcome model with a continuous endogenous variable is given in that paper.

This paper extends these results by providing the sharp identified set for the binary outcome model with discrete endogenous variables and giving a constructive proof of sharpness.

The paper explores the impact of particular restrictions available in the binary response model, namely that the threshold-crossing function is a monotone function of a linear index function through which the explanatory variables act. This restriction is satisfied in most of the parametric binary outcome models used in practice, for example, the probit and logit models.

Other related papers are Chesher and Smolinski (2012), which delivers the sharp identified set and a constructive proof of sharpness in an IV scalar ordered outcome model with a binary endogenous variable, and Chesher, Rosen, and Smolinski (2011) which, using tools drawn from the theory of random sets, develops sharp identified sets for random utility functions in multiple discrete choice models with instrumental variable restrictions.

2. IDENTIFYING POWER OF THE SINGLE-EQUATION IV BINARY RESPONSE MODEL

2.1. The Single-Equation Model

The value of a binary variable Y is determined as

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

Here U is an unobserved scalar continuously distributed random variable, and X is a vector random variable that may be jointly dependently distributed with U. The marginal distribution of U is normalized to be uniform on $[0, 1]$.

There are instrumental variables arranged in a vector Z. The support of (U, X, Z) is $[0, 1] \times \mathcal{X} \times \mathcal{Z}$. The model imposes the restriction that U and Z are independently distributed in the sense that the conditional distribution of U given $Z = z$ does not depend on z for values in the support set $\mathcal{Z}$.

In what follows, all probabilities are conditioned on values of Z so the instrumental variables can appear as arguments of the threshold-crossing function p. Of course for a model to deliver a nontrivial identified set it will have to include some restriction on the impact of Z on p, for example, restricting some elements of Z to be excluded from p. Then Z will appear as an argument of p when index restrictions are considered in Section 4.3, but for now, to simplify notation, the model will require all elements of Z to be excluded from p.

https://doi.org/10.1017/S0266466612000321 Published online by Cambridge University Press

The identifying power of this single-equation model is now considered. The question to be answered is: What can be known of the function p from knowledge of the probability distribution of Y and X given $Z = z$ when z varies within the set $\mathcal{Z}$ and the data generating process satisfies the restrictions of the single-equation IV model?

2.2. Identification

The single-equation IV model, denoted $\mathcal{M}$, is defined as follows.

MODEL $\mathcal{M}$. *Y is a binary random variable determined as*

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

where U is a continuously distributed scalar random variable normalized marginally Unif$(0, 1)$ and independent of Z, which is a list of instrumental variables excluded from the threshold-crossing function p and taking values in a set $\mathcal{Z}$.

In this definition independence, denoted $U \perp\!\!\!\perp Z$, signifies that $\Pr[U \in \mathcal{S} | Z = z]$ is independent of z for all sets $\mathcal{S} \subseteq [0, 1]$ and $z \in \mathcal{Z}$, allowing the possibility that Z is not a random variable. For example, its values could be purposively chosen by an experimenter.

Consider a particular data generating structure $S_0 \equiv \{p_0, F^0_{U|XZ}, F^0_{X|Z}\}$ admitted by this model. There are components: p_0, a particular threshold-crossing function; $F^0_{U|XZ}$, a particular conditional distribution function for U given X and Z; and $F^0_{X|Z}$, a particular conditional distribution function for X given Z. The index "0" distinguishes the structure under consideration (and its components) from others that might have generated the probability distributions of the observable variables.

To be admitted by the Model $\mathcal{M}$ the distribution functions $\{F^0_{U|XZ}, F^0_{X|Z}\}$, equivalently the distribution function $F^0_{UX|Z}$, must have associated with them marginal distributions for U given Z that respect the independence restriction and the uniform distribution normalization; that is,

$$F^0_{U|Z}(u|z) = u$$

for all $u \in [0, 1]$ and $z \in \mathcal{Z}$. Henceforth, "for all u" signifies for all $u \in [0, 1]$.

Let $F^0_{YX|Z}$ denote the joint distribution function of Y and X given Z generated by the structure S_0. This is determined as

$$F^0_{YX|Z}(0, x|z) = F^0_{UX|Z}(p_0(x), x|z),$$

$$F^0_{YX|Z}(1, x|z) = F^0_{X|Z}(x|z).$$

Let $\Pr_0[\cdot|Z = z]$ indicate probabilities calculated with respect to these measures.

https://doi.org/10.1017/S0266466612000321 Published online by Cambridge University Press

Observationally equivalent admissible structures, say S_*, have threshold-crossing functions p_* for which there exist distribution functions $F^*_{UX|Z}$ that respect the independence property and satisfy the condition

$$F^*_{YX|Z}(0,x|z) \equiv F^*_{UX|Z}(p_*(x),x|z) = F^0_{YX|Z}(0,x|z) \qquad \forall (x,z) \in \mathcal{X} \times \mathcal{Z}. \quad \textbf{(2)}$$

Let $\mathcal{P}_0(\mathcal{Z})$ denote the identified set of threshold-crossing functions associated with $F^0_{YX|Z}$ and the set $\mathcal{Z}$. It comprises all admissible threshold functions p_* that reside in structures for which (2) is satisfied.

Theorem 1 gives a system of inequalities that is satisfied by all functions in $\mathcal{P}_0(\mathcal{Z})$.

THEOREM 1. *A structure S_0 admitted by the Model $\mathcal{M}$ generates a distribution $F^0_{YX|Z}$. If a function p is a threshold-crossing function in a structure admitted by the Model $\mathcal{M}$ and observationally equivalent to S_0, then p satisfies the inequalities (3) and (4) for all u and all $z \in \mathcal{Z}$:*

$$c_{0l}(u,z;p) \equiv \Pr_0[Y = 0 \wedge p(X) \le u | Z = z] \le u, \quad \textbf{(3)}$$

$$c_{0u}(u,z;p) \equiv 1 - \Pr_0[Y = 1 \wedge u \le p(X) | Z = z] \ge u. \quad \textbf{(4)}$$

Here subscripts "l" and "u" indicate, respectively, lower and upper bounding probability functions. The subscript "0" indicates that a function (c_{0l} or c_{0u}) is calculated using the particular distribution functions $F^0_{YX|Z}$ generated by the structure S_0. Because there is conditioning on $Z = z$, Theorem 1 continues to hold when the threshold-crossing function p includes Z as an argument. The proof involves consideration of the requirement to locate threshold functions so that probability masses specified by the distributions $F^0_{YX|Z}$ can be accommodated.

Proof of Theorem 1. It is first shown that, when $p = p_0$, (3) and (4) hold for all u and all $z \in \mathcal{Z}$.

Consider the inequality (3) with $p = p_0$ and probabilities conditional on X and Z. For all x such that $p_0(x) > u$,

$$\Pr_0[Y = 0 \wedge p_0(X) \le u | X = x, Z = z] = 0,$$

and for all x such that $p_0(x) \le u$,

$$\begin{aligned}\Pr_0[Y = 0 \wedge p_0(X) \le u | X = x, Z = z] &= \Pr_0[Y = 0 | X = x, Z = z] \\ &= \Pr_0[U \le p_0(x) | X = x, Z = z] \\ &\le \Pr_0[U \le u | X = x, Z = z],\end{aligned}$$

and so for all x there is the inequality

$$\Pr_0[Y = 0 \wedge p_0(X) \le u | X = x, Z = z] \le \Pr_0[U \le u | X = x, Z = z]. \quad \textbf{(5)}$$

https://doi.org/10.1017/S0266466612000321 Published online by Cambridge University Press

Now consider the inequality (4) with $p = p_0$. For all x such that $u > p_0(x)$,

$$1 - \Pr_0[Y = 1 \wedge u \leq p_0(X) | X = x, Z = z] = 1,$$

and for all x such that $u \leq p_0(x)$,

$$\begin{aligned} 1 - \Pr_0[Y = 1 \wedge u \leq p_0(X) | X = x, Z = z] &= 1 - \Pr_0[Y = 1 | X = x, Z = z] \\ &= \Pr_0[U \leq p_0(x) | X = x, Z = z] \\ &\geq \Pr_0[U \leq u | X = x, Z = z], \end{aligned}$$

and so for all x there is the inequality

$$1 - \Pr_0[Y = 1 \wedge u \leq p_0(X) | X = x, Z = z]. \geq \Pr_0[U \leq u | X = x, Z = z]. \quad \textbf{(6)}$$

Taking expected value over X given $Z = z$ on the left- and right-hand sides of (5) and (6) using the distribution function $F^0_{X|Z}$ and exploiting the independence restriction $U \perp\!\!\!\perp Z$ and the uniform distribution normalization yields the inequalities (3) and (4) with $p = p_0$.

The result of the theorem now follows directly because, if some threshold-crossing function p_* is an element of a structure observationally equivalent to S_0, then it generates the same probability measure as S_0 does, so (3) and (4) hold for all u and all $z \in \mathcal{Z}$ with $p = p_*$. ■

The functions c_{0l} and c_{0u} in (3) and (4) are nondecreasing in u and satisfy inequalities

$$0 \leq c_{0l}(u, z; p) \leq F^0_{Y|Z}(0|z) \leq c_{0u}(u, z; p) \leq 1, \quad \textbf{(7)}$$

which hold for all functions p in the identified set, all $z \in \mathcal{Z}$ and u. The functions attain their lower and upper bounds as u approaches, respectively, 0 and 1. Examples are drawn in Section 4.2.

The inequalities (3) and (4) hold for all $z \in \mathcal{Z}$ so structural functions in the identified set satisfy

$$c_{0l}(u; p) \equiv \max_{z \in \mathcal{Z}} c_{0l}(u, z; p) \leq u \leq \min_{z \in \mathcal{Z}} c_{0u}(u, z; p) \equiv c_{0u}(u; p) \quad \textbf{(8)}$$

for all u, The envelope functions $c_{0u}(u; p)$ and $c_{0l}(u; p)$ are nondecreasing functions of u, and it follows from (7) that for all u and any admissible p,

$$0 \leq c_{0l}(u; p) \leq \max_{z \in \mathcal{Z}} F^0_{Y|Z}(0|z),$$

$$\min_{z \in \mathcal{Z}} F^0_{Y|Z}(0|z) \leq c_{0u}(u; p) \leq 1,$$

the bounds being approached as u passes to 0 or 1. For all functions p in the identified set $c_{0u}(u; p) \geq c_{0l}(u; p)$ for every $u \in [0, 1]$ but for functions p outside the identified set, violation of this inequality is possible.

https://doi.org/10.1017/S0266466612000321 Published online by Cambridge University Press

Structural equations for general discrete outcomes, $Y = h(X, U)$ with h monotone in scalar continuously distributed U and with U restricted to be distributed independently of instrumental variables Z and normalized $Unif(0, 1)$, are studied in Chesher (2010). It is shown there that structural functions in the identified set associated with distributions $F^0_{YX|Z}$ and a set of instrumental values $\mathcal{Z}$ satisfy the system of inequalities

$$\max_{z\in\mathcal{Z}} \Pr_0[Y < h(X,u)|Z=z] < u \leq \min_{z\in\mathcal{Z}} \Pr_0[Y \leq h(X,u)|Z=z] \quad \textbf{(9)}$$

for all u and that this system defines the sharp identified set of functions h when Y is binary and X is continuous.

The binary outcome model studied here is a special case with h defined by (1). In the binary response case the lower bound in (9) can be expressed as

$$\max_{z\in\mathcal{Z}} \Pr_0[Y = 0 \wedge p(X) < u|Z=z] < u.$$

Although the lower bounding probability here involves strong inequalities where (3) has weak inequalities, the restrictions imposed by the two lower bounds are effectively the same for discrete and continuous X. Chesher (2010) shows that (9) defines the sharp identified set of functions h when Y is binary and X is continuous. In the next section the analogous result for discrete X is provided and the relationship between (9) and the bounds (8) in the discrete X case is elucidated.

A restricted version of Model $\mathcal{M}$ may require the threshold-crossing function, p, to be a member of a parametric family of functions. Later the case in which $p(x)$ has the probit form is considered. When there are parametric restrictions the inequalities (3) and (4) sharply define the identified set of values of parameters associated with the distribution $F^0_{YX|Z}$ and Model $\mathcal{M}$.

In the parametric case it may be possible to obtain a complete characterization of the identified set, but in general this is difficult without further restriction. In econometric practice most of the parametric models that are used satisfy a "monotone index" restriction, under which the threshold-crossing function is a monotone function of a scalar index through which the explanatory variables act. Probit and logit models are leading examples.

The force of this semiparametric shape restriction is considered in Section 4. It leads to a result that allows visualization of identified sets of nonparametrically specified monotone structural functions and simple characterization of identified sets of values of index coefficients. The restriction also opens the way to an analysis of identification when explanatory variables are measured with error.

3. DISCRETE ENDOGENOUS VARIABLES

3.1. Identified Sets

The probability inequalities that appear in Theorem 1 are now given explicit representations for the case in which X is discrete. It is shown that in that case the

https://doi.org/10.1017/S0266466612000321 Published online by Cambridge University Press

inequalities of Theorem 1 define the sharp identified set of threshold-crossing functions.

Let X have support $\mathcal{X} = \{x_1, x_2, \ldots, x_K\}$, and for $k \in \{1, \ldots, K\}$ define $\gamma_k \equiv p(x_k)$, $\gamma_0 \equiv 0$, $\gamma_{K+1} \equiv 1$. Define $\gamma \equiv \{\gamma_1, \ldots, \gamma_K\}$. There may be restrictions on γ arising from consideration of the values x_k in $\mathcal{X}$ and the economic setting under study. These are not considered here, but they can easily be employed to refine the identified set developed here.[3]

For $k \in \{1, \ldots, K\}$ define conditional probabilities associated with the probability distributions $F^0_{YX|Z}$ as

$$\alpha^0_k(z) \equiv \Pr_0[Y = 0 | X = x_k, Z = z] \qquad \delta^0_k(z) \equiv \Pr_0[X = x_k | Z = z]$$

$$\pi^0_k(z) \equiv \Pr_0[Y = 0 \wedge X = x_k | Z = z] = \alpha^0_k(z)\delta^0_k(z).$$

The values $\{\delta^0_1(z), \ldots, \delta^0_K(z)\}$ are required to all be nonzero for all $z \in \mathcal{Z}$.[4] Adopt the convention that sums from 1 to 0 are zero, $\sum_{s=1}^{0}(\cdot)_s \equiv 0$.

THEOREM 2. *For a structural function p, assign indexes to the elements in the support set $\mathcal{X}$ so that $\gamma_1 \leq \gamma_2 \leq \cdots \leq \gamma_K$. The inequalities (3) and (4) of Theorem 1 obtained as u varies in* $[0, 1]$ *define a set of values of γ characterized by the inequalities*

$$\sum_{j=1}^{k} \pi^0_j(z) \leq \gamma_k \leq \sum_{j=1}^{k-1} \delta^0_j(z) + \sum_{j=k}^{K} \pi^0_j(z) \qquad \forall k \in \{1, \ldots, K\}. \tag{10}$$

If more than one element in γ is equal to γ_k, then their associated values $\pi_j(z)$ all contribute to the summations from 1 *to k on the left-hand side and from k to K on the right-hand side.*

Proof of Theorem 2. Consider a value $u \in [0, 1]$ and an index value k such that $\gamma_{k-1} < u \leq \gamma_k$. Considering the lower bounding probability (3), since

$$\{x : p(x) \leq u\} = \begin{cases} \{x_1, \ldots, x_{k-1}\}, & u < \gamma_k \\ \{x_1, \ldots, x_k\}, & u = \gamma_k, \end{cases}$$

there is, for $u \in (\gamma_{k-1}, \gamma_k]$,

$$\Pr_0[Y = 0 \wedge p(X) \leq u | Z = z] = \sum_{j=1}^{k-1} \pi^0_j(z) + \pi^0_k(z) 1[u = \gamma_k]. \tag{11}$$

Here $1[C]$ is 1 if the condition C is true and 0 otherwise.

Considering the upper bounding probability (4), since

$$\{x : u \leq p(x)\} = \{x_k, \ldots, x_K\}, \quad u \leq \gamma_k,$$

https://doi.org/10.1017/S0266466612000321 Published online by Cambridge University Press

there is, for $u \in (\gamma_{k-1}, \gamma_k]$,

$$1 - \Pr_0[Y = 1 \wedge u \leq p(X) | Z = z] = \sum_{j=1}^{k-1} \delta_j^0(z) + \sum_{j=k}^{K} \pi_j^0(z). \tag{12}$$

Setting $u = \gamma_k$ in (11) and (12) and substituting in the inequalities (3) and (4) delivers the inequalities (10).

All values of u in an interval $(\gamma_{k-1}, \gamma_k]$ deliver just the γ_k inequality in the sequence (10), and the union of the intervals obtained as k passes from 1 to K is $(0, 1]$. Setting u equal to zero in (11) and (12) does not deliver informative inequalities. Therefore, evaluating the inequalities (11) and (12) at all $u \in [0, 1]$ delivers only the K inequalities (10). ■

The lower inequality in Chesher (2010) has strong inequalities:

$$\Pr_0[Y = 0 \wedge p(X) < u | Z = z] < u, \tag{13}$$

whereas the inequality (3) has weak inequalities. When X varies continuously, this has no practical consequence. When X is discrete, there is for all $u \in (\gamma_k, \gamma_{k+1}]$

$$\{x : p(x) < u)\} = \{x : p(x) \leq \gamma_k\},$$

and so for all $u \in (\gamma_k, \gamma_{k+1}]$ (13) implies

$$\Pr_0[Y = 0 \wedge p(X) \leq \gamma_k | Z = z] < u,$$

from which it follows that

$$\Pr_0[Y = 0 \wedge p(X) \leq \gamma_k | Z = z] \leq \inf\{u : u \in (\gamma_k, \gamma_{k+1}]\} = \gamma_k,$$

which is the lower bounding inequality in (10).

A consequence of Theorem 2 is that structural functions that do not vary with x cannot lie in the identified set if $\Pr_0[Y = 0 | Z = z]$ varies at all as z varies in $\mathcal{Z}$. This is so because if $\gamma_1 = \gamma_2 = \cdots = \gamma_K$ are equal, say to some value $\bar{\gamma}$, then the left- *and* the right-hand sides of all the inequalities (10) are equal to

$$\sum_{j=1}^{K} \pi_j^0(z) = \Pr_0[Y = 0 | Z = z] \equiv \bar{\alpha}^0(z), \tag{14}$$

so the only vector γ that satisfies the inequalities has every element equal to $\bar{\alpha}^0(z) = \bar{\gamma}$. If $\bar{\alpha}^0(z)$ is not constant for variations of z in $\mathcal{Z}$, then there is no admissible constant value $\bar{\gamma}$ (recall there is the IV restriction excluding Z from the threshold-crossing function), so threshold-crossing functions $p(x)$ that do not vary with x are not contained in the identified set.[5]

Systems of inequalities for different permutations of γ are obtained by exchange of indices in the inequalities of Theorem 2.

For each permutation the set of values of γ defined by the inequalities (10) is precisely the subset of the identified set for the Model $\mathcal{M}$ associated with the permutation. This is the subject of Theorem 3.

https://doi.org/10.1017/S0266466612000321 Published online by Cambridge University Press

THEOREM 3. *Consider a value $z \in \mathcal{Z}$. For every sequence $\gamma_1 \leq \gamma_2 \leq \cdots \leq \gamma_K$ for which, at that value z, the system of inequalities (10) defines a nonempty set, there exists a distribution function $F_{U|XZ}(u|x,z)$ such that the following conditions hold for $k \in \{1,\ldots,K\}$.*

(1) Proper conditional distribution functions:

$$0 \leq F_{U|XZ}(\gamma_1|x_k,z) \leq \cdots \leq F_{U|XZ}(\gamma_K|x_k,z) \leq 1.$$

(2) Independence:[6]

$$\sum_{j=1}^{K} \delta_j^0(z) F_{U|XZ}(\gamma_k|x_j,z) = \gamma_k.$$

(3) Observational equivalence:

$$F_{U|XZ}(\gamma_k|x_k,z) = \alpha_k^0(z).$$

A constructive proof of Theorem 3 is given in the Appendix. Combining the results of Theorems 1 and 3 leads to Theorem 4.

THEOREM 4. *A structure S_0 admitted by the Model $\mathcal{M}$ generates a distribution $F_{YX|Z}^0$. Define the set of functions $\tilde{\mathcal{P}}_0(\mathcal{Z})$:*

$$\tilde{\mathcal{P}}_0(\mathcal{Z}) = \{p : \max_{z\in\mathcal{Z}} c_{0l}(u,z;p) \leq u \leq \min_{z\in\mathcal{Z}} c_{0u}(u,z;p), \forall u \in [0,1]\},$$

where $c_{0l}(u,z;p)$ and $c_{0u}(u,z;p)$ are defined, respectively, in equations (3) and (4). The set $\tilde{\mathcal{P}}_0(\mathcal{Z})$ is the identified set of threshold-crossing functions, that is $\tilde{\mathcal{P}}_0(\mathcal{Z}) = \mathcal{P}_0(\mathcal{Z})$.

Proof of Theorem 4. Theorem 1 implies that all functions p in the identified set satisfy

$$c_{0l}(u,z;p) \leq u \leq c_{0u}(u,z;p)$$

for all $u \in [0,1]$ and for each $z \in \mathcal{Z}$. Since only functions p that satisfy these inequalities for all $z \in \mathcal{Z}$ at each value of u are in the identified set, $\mathcal{P}_0(\mathcal{Z}) \subseteq \tilde{\mathcal{P}}_0(\mathcal{Z})$. Theorem 3 states that for each $z \in \mathcal{Z}$ and each function p that satisfies these inequalities there is a distribution function $F_{U|XZ}(u|x,z)$ that combined with $F_{X|Z}^0(x|z)$ delivers the probability distribution $F_{YX|Z}^0(y,x|z)$. It follows that $\mathcal{P}_0(\mathcal{Z}) \supseteq \tilde{\mathcal{P}}_0(\mathcal{Z})$ and combining results, $\tilde{\mathcal{P}}_0(\mathcal{Z}) = \mathcal{P}_0(\mathcal{Z})$. ■

Each admissible permutation of the elements of γ delivers a system of linear inequalities (10) that defines a convex subset of the unit K-cube.[7] In general there are up to $K!$ permutations, so the identified set is the union of as many as $K!$ convex sets. However, in practice some of the component sets may be empty, and shape restrictions can rule some permutations inadmissible. The next section illustrates.

https://doi.org/10.1017/S0266466612000321 Published online by Cambridge University Press

3.2. Illustration

In this section identified sets are visualized for a case in which discrete X has three points of support. There are then three structural features of interest: $\gamma_1 \equiv p(x_1)$, $\gamma_2 \equiv p(x_2)$, and $\gamma_3 \equiv p(x_3)$. They take values in the unit 3-cube, and identified sets can be portrayed in perspective 3D displays. Define $\gamma \equiv \{\gamma_1, \gamma_2, \gamma_3\}$.

The probability distributions for Y and X given Z used in this example are generated using Gaussian triangular structures as follows.

$$Y = 1[X + W > 0] \qquad X = \begin{cases} -1, & -\infty < X^* \leq -0.5 \\ 0, & -0.5 < X^* \leq 0.5 \\ 1, & 0.5 < X^* \leq +\infty \end{cases}$$

$$X^* = b_1 Z + V \qquad \begin{bmatrix} W \\ V \end{bmatrix} \perp\!\!\!\perp Z \sim N_2(\mathbf{0}, \mathbf{I}_2).$$

The instrumental variable Z takes values in the set

$$\mathcal{Z} = \{-1.5, -1.0, -0.5, 0.0, +0.5, +1.0, +1.5\}.$$

The coefficient b_1 is varied across the interval $[0.15, 1.45]$, generating a sequence of graphs that show how the strength of the instrument affects the identified set.

This structure is of the sort admitted by a simultaneous equations model with dummy endogenous variable studied in Heckman (1978). Heckman's (1978) model is parametric and Gaussian and is point identifying. In the absence of the fully parametric restriction, the model would not be point identifying, as the endogenous explanatory variable is not continuous, as explained in Chesher (2005), where some partial identification results are provided.

The structure is triangular, and X is exogenous because the covariance of Gaussian W and V is zero. This information is not embodied in the single-equation IV nonparametric model whose set-identifying power is now studied.

Figure 1 shows the identified set when $b_1 = 0.15$. It is the union of six convex sets, one lying in each of the orthoschemes of the unit 3-cube. The orthoschemes of the unit cube are the six right tetrahedra each of which is a set of values of γ that satisfy a particular ordering, for example,[8]

$$\mathcal{O}_{123} \equiv \{\gamma : \gamma_1 \leq \gamma_2 \leq \gamma_3\}.$$

The orthoschemes have intersections only at their faces and along the ray of equality. For example, the orthoschemes $\mathcal{O}_{123}$ and $\mathcal{O}_{213}$ have nonempty intersection $\{\gamma : \gamma_1 = \gamma_2 \leq \gamma_3\}$. The set $\{\gamma : \gamma_1 = \gamma_2 = \gamma_3\}$ is the ray of equality. It is a subset of all the orthoschemes and is their common intersection.

In Figure 1 the instrument is very weak, and the identified set has elements in each of the six orthoschemes. Any ordering of the elements of γ could have

https://doi.org/10.1017/S0266466612000321 Published online by Cambridge University Press

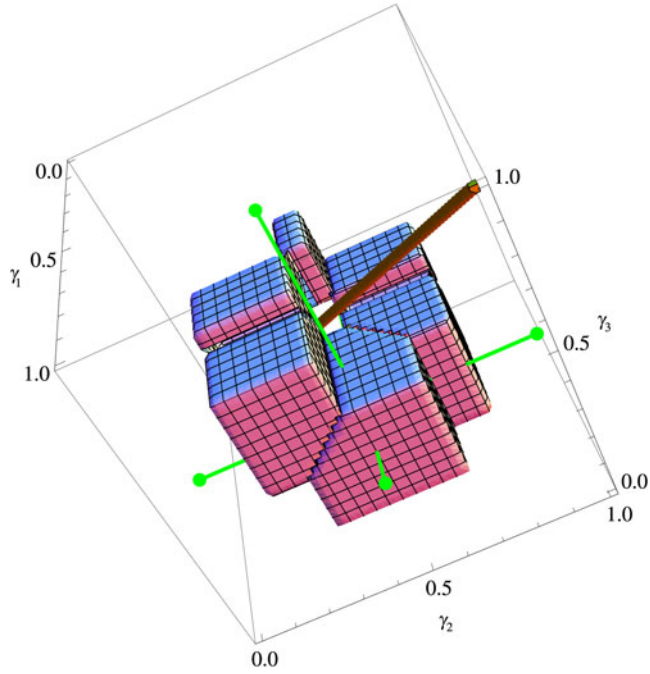


FIGURE 1. Binary Y, ternary X: identified set for $\{\gamma_1, \gamma_2, \gamma_3\}$, $b_1 = 0.15$.

produced the probability distributions used in the computations. In the structure generating the probabilities used in this exercise,

$$\gamma = \{0.691, 0.500, 0.308\}.$$

This point in the unit 3-cube is at the intersection of the three green rays that lie parallel to the axes in the various figures.

The ray of equality that connects the points $\{0, 0, 0\}$ and $\{1, 1, 1\}$ is drawn brown. Notice that it does not intersect the identified set at all. This is because the instrument, although weak, does affect $\bar{\alpha}^0(z) = \Pr_0[Y = 0|Z = z]$.[9] At this value of the coefficient on the instrumental variable, the hypothesis that $p(x)$ does not depend on x is falsifiable. The identified set is not convex, but it is connected.

In the structure generating Figure 2, $b_1 = 0.25$ and this increase in value is sufficient to remove a set occupying one of the orthoschemes, specifically $\mathcal{O}_{123}$. At this value of the coefficient on the instrumental variable, the hypothesis that $p(x)$ is monotone increasing is falsifiable.[10] The identified set remains connected and has no intersection with the ray of equality.

Increasing the value of b_1 to 0.45 (see Figure 3) causes the identified set to become disconnected. The sets lying in orthoschemes $\mathcal{O}_{132}$ and $\mathcal{O}_{213}$ have no intersection with each other nor with other parts of the identified set. They have

https://doi.org/10.1017/S0266466612000321 Published online by Cambridge University Press

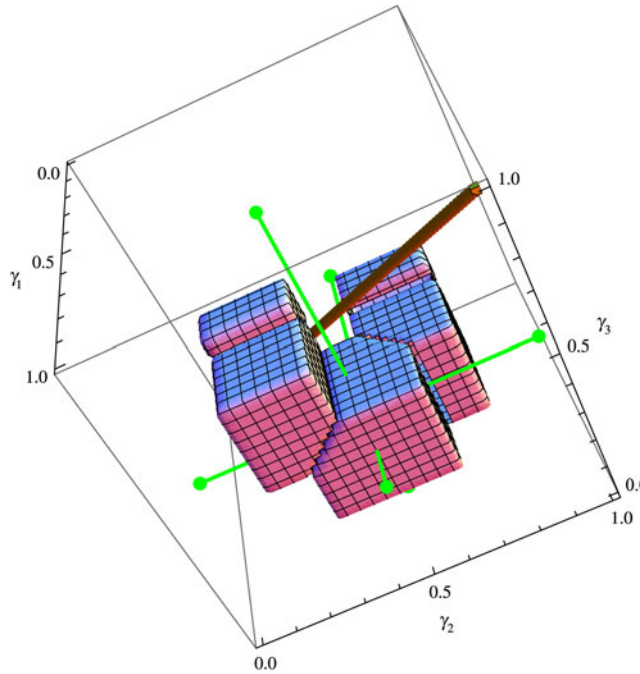


FIGURE 2. Binary Y, ternary X: identified set for $\{\gamma_1, \gamma_2, \gamma_3\}$, $b_1 = 0.25$.

small volumes and on increasing b_1 further to 0.55 they disappear; see Figure 4. At this value of b_1 the identified set lies in just three of the orthoschemes, $\mathcal{O}_{312}$, $\mathcal{O}_{321}$, and $\mathcal{O}_{231}$. The value of γ in the structures generating probabilities for this example lies in $\mathcal{O}_{321}$. At this value of b_1 the identified set is connected.

Increasing the value of b_1 further reduces the volumes of the remaining three components of the identified set. With $b_1 = 1.25$ the volumes are very small, and Figure 5 shows that the identified set is again disconnected. With $b_1 = 1.45$ the identified set becomes connected again, occupying only the orthoscheme O_{321}. As Figure 6 shows, the identified set is now arranged on and very close to the ray connecting

$$\{0.691, 0.308, 0.308\} \quad \text{and} \quad \{0.691, 0.691, 0.308\},$$

and γ_1 and γ_3 are essentially point identified, by the model. The value of γ_2 remains set identified, and further increases in b_1 do nothing to alter this.

In this example allowing the set of instrumental values to cover the interval $[-1.5, +1.5]$ more densely fails to deliver point identification of γ_2. The parameters γ_1 and γ_3 are associated with extreme values of X, and the approach to point identification as b_1 increases is of the sort referred to as identification at infinity in, for example, Heckman (1990). In the absence of a parametric specification of

https://doi.org/10.1017/S0266466612000321 Published online by Cambridge University Press

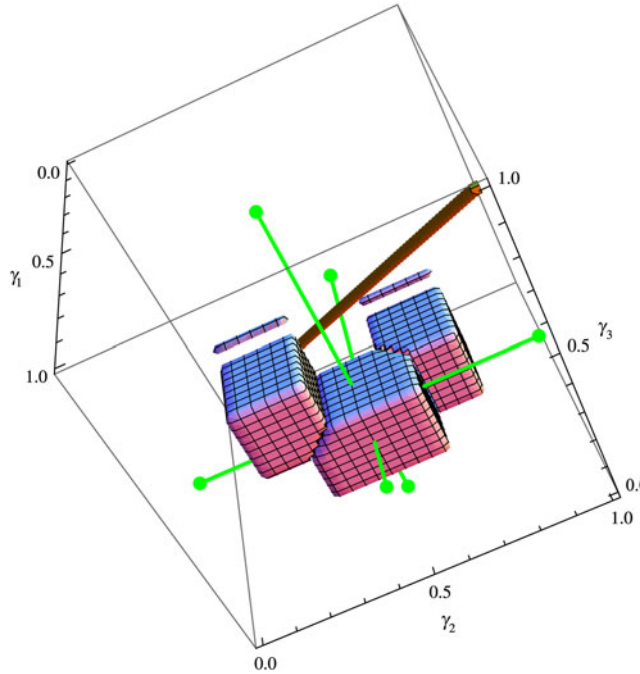


FIGURE 3. Binary Y, ternary X: identified set for $\{\gamma_1, \gamma_2, \gamma_3\}$, $b_1 = 0.45$.

$p(x)$, this mechanism fails to deliver point identification of γ_2, which is associated with a value of X in the interior of its support.

Many of the features of the identified set seen in this example appear when X has more points of support. When X has K points of support, γ lies in a unit K-cube and the identified set is the union of up to $K!$ convex sets, each lying in one of the $K!$ orthoschemes of the unit K-cube. The identified set may be disconnected and is generally not convex. The ray of equality does not intersect the identified set if $\Pr_0[Y = 0|Z = z]$ varies with z.

4. MONOTONICITY AND INDEX RESTRICTIONS

Almost all parametric models for binary outcomes used in practice require the threshold-crossing function to be a monotone function of a scalar index function of explanatory variables. Probit and logit models are leading examples. The impact of such a monotone index restriction in a semiparametric single equation IV model with endogenous explanatory variables is now considered

First, the case in which discrete or continuous X is scalar is considered.[11] The threshold function is specified as $p(x)$ with p strictly monotone in x. There is no restriction on the direction of the dependence on x, although this is easy to

https://doi.org/10.1017/S0266466612000321 Published online by Cambridge University Press

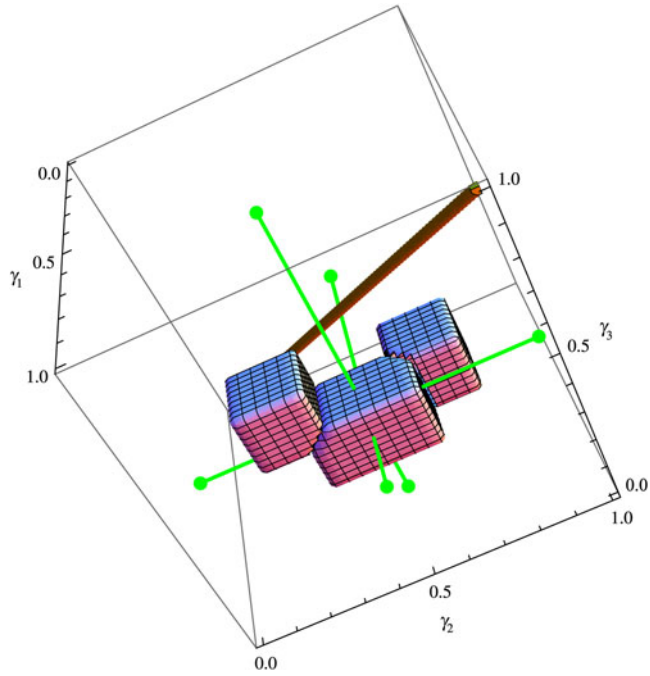


FIGURE 4. Binary Y, ternary X: identified set for $\{\gamma_1, \gamma_2, \gamma_3\}$, $b_1 = 0.55$.

incorporate. The identified set of threshold functions is shown to comprise all monotone functions that lie between one of two pairs of bounding functions: one pair increasing, the other pair decreasing. These functions are shown to be simple functionals of the joint distribution of the binary outcome and the endogenous variable conditional on the instrumental variables.

When instruments are not strong, the identified set can contain both increasing and decreasing functions, but not functions that are insensitive to variations in x. In a sense then the identified set of structural functions may not be connected. The results are illustrated using a probability measure generated by a Gaussian triangular system, and the impact of imposing parametric restrictions is considered.

The identified set of threshold functions is the intersection of sets determined by pairs of upper and lower bounding functions. Each distinct value of the instrumental variables generates a pair of bounds. The monotonicity restriction is falsifiable because for a particular probability distribution of (Y, X) given Z the bounding functions may intersect, in which case there is no monotone threshold function in the IV model's identified set.

Attention is then turned to models in which X may be a vector. Now Z is allowed to appear in the structural function, possibly subject to some exclusion restrictions. The models considered have threshold functions of the form $p(X'\alpha + Z'\delta)$ with the function p monotone. The identified set comprises a set

https://doi.org/10.1017/S0266466612000321 Published online by Cambridge University Press

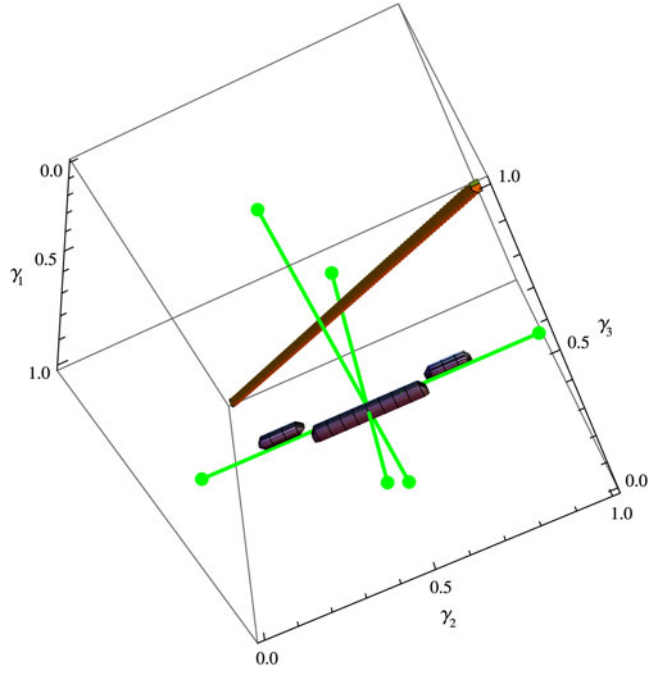


FIGURE 5. Binary Y, ternary X: identified set for $\{\gamma_1, \gamma_2, \gamma_3\}$, $b_1 = 1.25$.

of parameter values, each element of which is associated with a set of monotone functions, p. For each value of (α, δ) in the identified set, bounding functions are derived that define the set of functions associated with (α, δ). Under the monotonicity restriction, there is no need to consider particular alternative functions p when developing the identified set of index coefficients. This substantially simplifies the computation and estimation of that identified set.

4.1. Monotone Threshold Functions with Scalar X

Let X be scalar and let p^{-1} denote the inverse function of p.[12] If the threshold function is restricted to be monotone, then

$$\{x : p(x) \leq u\} = \begin{cases} \{x : x \leq p^{-1}(u)\}, & p \text{ increasing} \\ \{x : x \geq p^{-1}(u)\}, & p \text{ decreasing,} \end{cases}$$

so the inequalities (3) and (4) can be written as follows.

- Increasing p.

$$c_{0l}^{\uparrow}(u, z; p) = \Pr_0[Y = 0 \wedge X \leq p^{-1}(u)|z] \leq u \tag{15}$$

$$c_{0u}^{\uparrow}(u, z; p) = 1 - \Pr_0[Y = 1 \wedge X \geq p^{-1}(u)|z] \geq u \tag{16}$$

https://doi.org/10.1017/S0266466612000321 Published online by Cambridge University Press

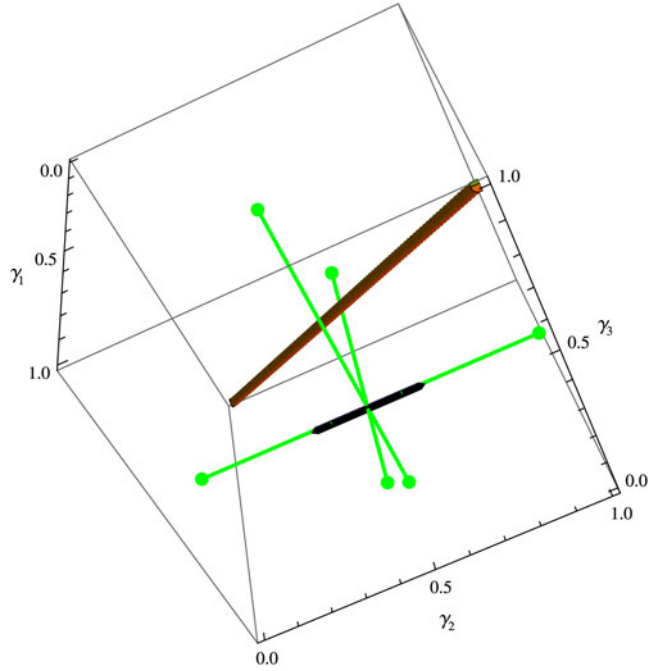


FIGURE 6. Binary Y, ternary X: identified set for $\{\gamma_1, \gamma_2, \gamma_3\}$, $b_1 = 1.45$.

- Decreasing p.

$$c_{0l}^{\downarrow}(u, z; p) = \Pr_0[Y = 0 \wedge X \geq p^{-1}(u)|z] \leq u \tag{17}$$

$$c_{0u}^{\downarrow}(u, z; p) = 1 - \Pr_0[Y = 1 \wedge X \leq p^{-1}(u)|z] \geq u \tag{18}$$

Evaluating these functions at $u = p(\sigma)$ with $\sigma \in \mathcal{X}$ moves the threshold function out of the bounding function. For example,

$$c_{0l}^{\uparrow}(p(\sigma), z; p) \equiv d_{0l}^{\uparrow}(\sigma, z) = \Pr_0[Y = 0 \wedge X \leq \sigma|z] \leq u = p(\sigma),$$

and the following inequalities are satisfied under the monotonicity restriction for all $\sigma \in \mathcal{X}$ and $z \in \mathcal{Z}$ by all functions p in the identified set, and only by these functions.

- Increasing p.

$$d_{0l}^{\uparrow}(\sigma, z) \equiv \Pr_0[Y = 0 \wedge X \leq \sigma|z] \leq p(\sigma) \tag{19}$$

$$d_{0u}^{\uparrow}(\sigma, z) \equiv 1 - \Pr_0[Y = 1 \wedge X \geq \sigma|z] \geq p(\sigma) \tag{20}$$

https://doi.org/10.1017/S0266466612000321 Published online by Cambridge University Press

- Decreasing p.

$$d_{0l}^{\downarrow}(\sigma, z) \equiv \Pr_0[Y = 0 \wedge X \geq \sigma \,|z] \leq p(\sigma) \tag{21}$$

$$d_{0u}^{\downarrow}(\sigma, z) \equiv 1 - \Pr_0[Y = 1 \wedge X \leq \sigma \,|z] \geq p(\sigma) \tag{22}$$

It is very convenient to have the threshold-crossing function outside the bounding functions because the functions can be derived or estimated just once and then compared with any candidate threshold function.

Define envelope functions as

$$d_{0l}^{\uparrow}(\sigma) \equiv \max_{z \in \mathcal{Z}} d_{0l}^{\uparrow}(\sigma, z) \qquad d_{0u}^{\uparrow}(\sigma) \equiv \min_{z \in \mathcal{Z}} d_{0u}^{\uparrow}(\sigma, z), \tag{23}$$

$$d_{0l}^{\downarrow}(\sigma) = \max_{z \in \mathcal{Z}} d_{0l}^{\downarrow}(\sigma, z) \qquad d_{0u}^{\downarrow}(\sigma) \equiv \min_{z \in \mathcal{Z}} d_{0u}^{\downarrow}(\sigma, z). \tag{24}$$

The arguments set out above lead to the following theorem.

THEOREM 5. *In the Model $\mathcal{M}$ let the threshold-crossing function p be a monotone function of a scalar argument. The identified set of threshold functions is the union of two sets of functions,*

$$\left\{p : d_{0l}^{\uparrow}(\sigma) \leq p(\sigma) \leq d_{0u}^{\uparrow}(\sigma), \forall \sigma \in \mathcal{X}\right\} \cup \left\{p : d_{0l}^{\downarrow}(\sigma) \leq p(\sigma) \leq d_{0u}^{\downarrow}(\sigma), \forall \sigma \in \mathcal{X}\right\}.$$

One of these sets may be empty, and this will tend to happen when instruments are strong with rich support, as illustrated shortly. Prior restrictions may eliminate one of the sets by restricting the threshold function to be, say, increasing.

If a model further restricts p to lie in a parametric family, then only parameter values leading to functions in the family that lie within the union of the sets just defined fall in the identified set of parameter values. Parametric probit restrictions are considered shortly in an illustrative example.

The functions $d_{0l}^{\uparrow}$ and $d_{0u}^{\uparrow}$ are increasing in σ, and the functions $d_{0l}^{\downarrow}$ and $d_{0u}^{\downarrow}$ are decreasing in σ. It is easy to show that the functions obey the following inequalities with left- and right-hand bounds achieved as σ approaches, respectively, $-\infty$ and $+\infty$:

$$0 \leq d_{0l}^{\uparrow}(\sigma, z) \leq \Pr_0[Y = 0|z] \leq d_{0u}^{\uparrow}(\sigma, z) \leq 1, \tag{25}$$

$$1 \geq d_{0u}^{\downarrow}(\sigma, z) \geq \Pr_0[Y = 0|z] \geq d_{0l}^{\downarrow}(\sigma, z) \geq 0, \tag{26}$$

and that

$$\left.\begin{matrix}\lim_{\sigma \to -\infty} d_{0u}^{\uparrow}(\sigma) \\ \lim_{\sigma \to +\infty} d_{0u}^{\downarrow}(\sigma)\end{matrix}\right\} = \min_{z \in \mathcal{Z}} \Pr_0[Y = 0|z] \leq \max_{z \in \mathcal{Z}} \Pr_0[Y = 0|z] = \begin{cases}\lim_{\sigma \to +\infty} d_{0l}^{\uparrow}(\sigma) \\ \lim_{\sigma \to -\infty} d_{0l}^{\downarrow}(\sigma)\end{cases}.$$

The bounding functions are illustrated in the next section.

https://doi.org/10.1017/S0266466612000321 Published online by Cambridge University Press

4.2. Illustration

These results are illustrated using probability measures generated by a triangular Gaussian structure that satisfies the restrictions of the single-equation IV model. The structural function for binary Y has a probit form with an endogenous explanatory variable. This choice makes the calculation of the bounding functions easy, it highlights the relative power of the control function model that would be point identifying in this case, and it places us in familiar applied econometrics territory.[13]

The structure has binary Y recording whether Y^* is positive. Latent Y^* and X are generated by structures with linear equations and jointly Gaussian latent variables,

$$Y = 1(Y^* > 0) \qquad Y^* = a_0 + a_1 X + W \qquad X = b_0 + b_1 Z + V,$$

$$\begin{bmatrix} W \\ V \end{bmatrix} \perp\!\!\!\perp Z \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & s_{wv} \\ s_{wv} & s_{vv} \end{bmatrix} \right).$$

The joint distribution of Y^* and X given $Z = z$ is $N(\mu(z), \Sigma)$.

$$\mu(z) = \begin{bmatrix} a_0 + a_1 b_0 + a_1 b_1 z \\ b_0 + b_1 z \end{bmatrix} \qquad \Sigma = \begin{bmatrix} 1 + 2a_1 s_{wv} + a_1^2 s_{vv} & s_{wv} + a_1 s_{vv} \\ \sigma_{wv} + a_1 s_{vv} & s_{vv} \end{bmatrix}.$$

The probabilities that appear in the inequalities (19)–(22) are bivariate normal orthant probabilities.[14] The threshold function for the structures employed in this example is $p(x) = \Phi(-a_0 - a_1 x)$, where Φ is the standard normal distribution function.

The identifying power of the following nonparametric model is considered.

$$Y = h(X, U) = \begin{cases} 0, & 0 \le U \le p(X) \\ 1, & p(X) < U \le 1 \end{cases} \qquad U \perp\!\!\!\perp Z, \qquad p \text{ monotone.} \qquad \textbf{(27)}$$

Graphs show the bounding functions (19)–(22) varying with σ for 10 values of the instrument z equally spaced in $[-1, 1]$. The functions are calculated using the probability measure generated by Gaussian triangular structures defined above with $a_0 = b_0 = 0$, $a_1 = -1$, $s_{wv} = 0.5$, $s_{vv} = 1$. At these parameter values the structural threshold function is the standard normal distribution function.

In Figure 7 the value of b_1, the coefficient on the instrumental variable in the equation for endogenous X, is 0.3. The upper pane shows the increasing bounding functions (19)–(20); the lower pane shows the decreasing functions (21)–(22).

The envelope bounding functions (23) and (24) are obtained at each value of σ as the maximum of the lower bounding functions and the minimum of the upper bounding functions. They are drawn as dashed red lines. The identified set of structural threshold functions comprises all increasing functions that pass between the upper and lower envelope bounding functions in the upper pane and all

https://doi.org/10.1017/S0266466612000321 Published online by Cambridge University Press

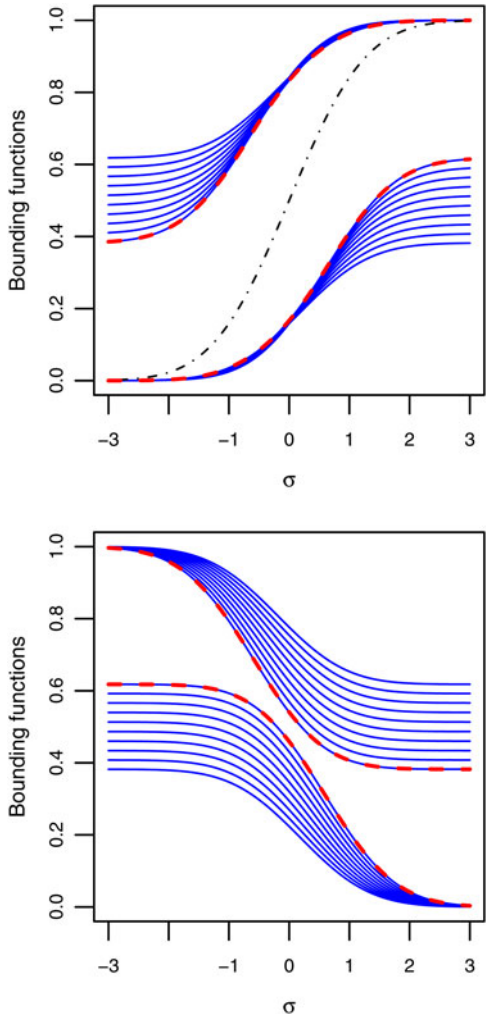


FIGURE 7. Bounding functions (blue) at 10 values of $z \in [-1, 1]$ with tight bounds (dashed red), between which lie the monotone functions in the identified set. A relatively weak instrument, with $b_1 = 0.3$.

decreasing functions that pass between the upper and lower envelope bounding functions in the lower pane.

The structural threshold function in the Gaussian triangular structure used to generate the probability measure employed in these calculations is the increasing dashed line passing between the upper and lower bounding functions in the upper pane. Any monotone increasing (decreasing) function passing between the red dashed lines in the upper (lower) pane in Figure 7, together with a suitably chosen

https://doi.org/10.1017/S0266466612000321 Published online by Cambridge University Press

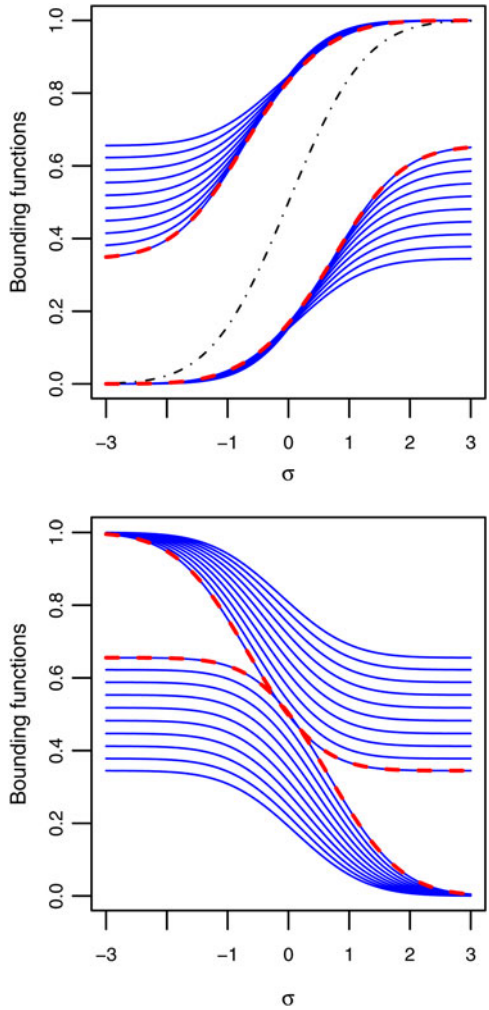


FIGURE 8. Bounding functions (blue) at 10 values of $z \in [-1, 1]$ with tight bounds (dashed red), between which lie the monotone functions in the identified set. A slightly stronger instrument than in Figure 1, with $b_1 = 0.4$. The set of decreasing functions (lower pane) is empty because the tight bonds intersect.

(typically non-Gaussian) distribution for U and X given $Z = z \in \mathcal{Z}$, generates the same probability measure.

When the power of the instrument is increased by setting the parameter $b_1 = 0.4$, the identified set is reduced as shown in Figure 8. The envelope bounding functions in the lower pane now intersect, and no decreasing function can pass between these functions. The instrument is now sufficiently strong to eliminate all monotone decreasing functions from the identified set.

https://doi.org/10.1017/S0266466612000321 Published online by Cambridge University Press

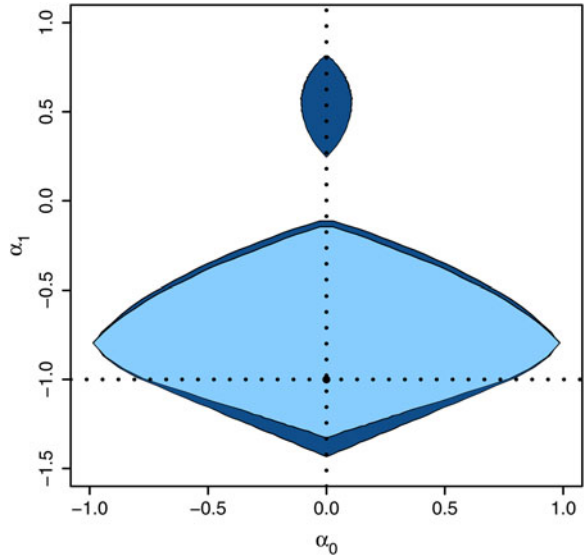


FIGURE 9. Identified sets in a parametric probit model for the structures delivering Figure 7 ($b_1 = 0.3$, dark and light blue) and Figure 8 ($b_1 = 0.4$, light blue only).

When this nonparametric model is augmented with parametric restrictions, the identified set is reduced to the subset of the identified set of nonparametric functions containing only those functions that are members of the family of functions specified in the parametric model. To illustrate, consider the identifying power of the following parametric probit model as in (27) with $p(X) = \Phi(-\alpha_0 - \alpha_1 X)$ when Y and X are determined by the structure used to produce Figure 7. At these parameter values the structural threshold-crossing function is $\Phi(X)$.

The identified set of parameter values comprises the set of values of (α_0, α_1) that deliver functions $\Phi(-\alpha_0 - \alpha_1 X)$ that lie between the envelope upper and lower bounding functions graphed (red dashed) in Figure 7. These identified sets are drawn in Figure 9, which shows two cases. The light blue colored region in the lower part of the figure is the identified set when the coefficient on the instrumental variable in the equation for X is $b_1 = 0.4$, for which the bounding functions are shown in Figure 8. On reducing b_1 to 0.3, the value used to generate Figure 7, the identified set expands by the amount colored dark blue, and it becomes disconnected, with a small region in the upper part of the figure where $\alpha_1 > 0$.

4.3. Monotone Index Restriction

Now consider models in which there is a monotone index restriction, namely that for all values, x and z, of X and Z the threshold-crossing function can be written as $p(\alpha' x + \delta' z)$ for some constant finite dimensional vectors α and δ, where p is

https://doi.org/10.1017/S0266466612000321 Published online by Cambridge University Press

a monotone function. Here X can be nonscalar, and Z is allowed to appear in the threshold function. The monotone index binary outcome IV model is

$$Y = \begin{cases} 0\,, & 0 < U \leq p(\alpha' X + \delta' Z) \\ 1\,, & p(\alpha' X + \delta' Z) < U \leq 1 \end{cases}, \quad U \perp\!\!\!\perp Z, \quad p \text{ monotone.}$$

There will typically be a restriction excluding some components of Z from the index. There will also be a normalization; for example, one might set equal to 1 an element of α or δ corresponding to a variable whose coefficient is restricted to be nonzero.

Consider a threshold function $p(\alpha' x + \delta' z)$ that lies in the identified set associated with a structure S_0 that delivers conditional distributions $F^0_{Y X|Z}$ for values of Z in the support set $\mathcal{Z}$. As before, let $\Pr_0$ indicate probabilities calculated using these measures.

Analogous to (15) and (16) there is, for increasing p

$$c^{\uparrow}_{0l}(u, z; h) = \Pr_0[Y = 0 \wedge \alpha' X + \delta' Z \leq p^{-1}(u)|z] \leq u, \tag{28}$$

$$c^{\uparrow}_{0u}(u, z; h) = 1 - \Pr_0[Y = 1 \wedge \alpha' X + \delta' Z \geq p^{-1}(u)|z] \geq u, \tag{29}$$

with inequalities reversed in the definitions of events when p is decreasing.

Continuing along the lines taken in Section 4.1 there are, on substituting $u = p(\sigma)$, the following inequalities:

- Increasing p.

$$d^{\uparrow}_{0l}(\sigma, z; \alpha, \delta) \equiv \Pr_0[Y = 0 \wedge \alpha' X \leq \sigma - \delta' z|z] \leq p(\sigma) \tag{30}$$

$$d^{\uparrow}_{0u}(\sigma, z; \alpha, \delta) \equiv 1 - \Pr_0[Y = 1 \wedge \alpha' X \geq \sigma - \delta' z|z] \geq p(\sigma) \tag{31}$$

- Decreasing p.

$$d^{\downarrow}_{0l}(\sigma, z; \alpha, \delta) \equiv \Pr_0[Y = 0 \wedge \alpha' X \geq \sigma - \delta' z|z] \leq p(\sigma) \tag{32}$$

$$d^{\downarrow}_{0u}(\sigma, z; \alpha, \delta) \equiv 1 - \Pr_0[Y = 1 \wedge \alpha' X \leq \sigma - \delta' z|z] \geq p(\sigma) \tag{33}$$

A threshold-crossing function, characterized by (p, α, δ) with monotone p, lies in the identified set if and only if one of these pairs of inequalities holds at each $\sigma \in \mathcal{X}$ for all $z \in \mathcal{Z}$. At each value of σ, it is the largest and smallest values of the respectively lower and upper bounding probabilities that are relevant. Defining envelope functions

$$d^{\uparrow}_{0l}(\sigma; \alpha, \delta) \equiv \max_{z \in \mathcal{Z}} d^{\uparrow}_{0l}(\sigma, z; \alpha, \delta) \qquad d^{\uparrow}_{0u}(\sigma; a, \delta) \equiv \min_{z \in \mathcal{Z}} d^{\uparrow}_{0u}(\sigma, z; \alpha, \delta)$$

$$d^{\downarrow}_{0l}(\sigma; \alpha, \delta) \equiv \max_{z \in \mathcal{Z}} d^{\downarrow}_{0l}(\sigma, z; \alpha, \delta) \qquad d^{\downarrow}_{0u}(\sigma; \alpha, \delta) \equiv \min_{z \in \mathcal{Z}} d^{\downarrow}_{0u}(\sigma, z; \alpha, \delta),$$

https://doi.org/10.1017/S0266466612000321 Published online by Cambridge University Press

there are the following inequalities:

$$\text{Increasing } p: \quad d_{0l}^{\uparrow}(\sigma;\alpha,\delta) \le p(\sigma) \le d_{0u}^{\uparrow}(\sigma;a,\delta), \tag{34}$$

$$\text{Decreasing } p: \quad d_{0l}^{\downarrow}(\sigma;\alpha,\delta) \le p(\sigma) \le d_{0u}^{\downarrow}(\sigma;\alpha,\delta), \tag{35}$$

which, under the monotone index restriction, hold for all $\sigma \in \mathcal{X}$ and all, and only, structural functions (p,α,δ) in the identified set.

The identified set I_0 associated with a structure S_0 comprises all (p,α,δ) with monotone p for which one of the inequalities (34) and (35) holds for all $\sigma \in \mathcal{X}$.

The identified set can be characterized as follows. There are two coupled components: a set of values of the finite dimensional parameters, α and δ, denoted $I_0^{\alpha\delta}$, and for each element of this set, a nonempty set of monotone functions $I_0^p(\alpha,\delta)$. This set of monotone functions is the union of two sets:

$$I_0^p(\alpha,\delta) = A_0^{\uparrow}(\alpha,\delta) \cup A_0^{\downarrow}(\alpha,\delta),$$

one, $A_0^{\uparrow}(\alpha,\delta)$, containing no decreasing functions, and the other, $A_0^{\downarrow}(\alpha,\delta)$, containing no increasing functions. These sets of functions are defined as

$$A_0^{\uparrow}(\alpha,\delta) \equiv \{p : d_{0l}^{\uparrow}(\sigma;\alpha,\delta) \le p(\sigma) \le d_{0u}^{\uparrow}(\sigma;\alpha,\delta) \qquad \forall \sigma \in \mathcal{X}\}, \tag{36}$$

$$A_0^{\downarrow}(\alpha,\delta) \equiv \{p : d_{0l}^{\downarrow}(\sigma;\alpha,\delta) \le p(\sigma) \le d_{0u}^{\downarrow}(\sigma;\alpha,\delta) \qquad \forall \sigma \in \mathcal{X}\}. \tag{37}$$

One, but not both of these sets may be empty.

Considering the inequalities (4.3), the functions $d_{0l}^{\uparrow}$ and $d_{0u}^{\uparrow}$ are increasing functions of σ so there is at least one increasing function p satisfying these inequalities for all σ if and only if there is no value of σ at which the functions intersect.[15] Similarly, there is at least one decreasing function p satisfying the inequalities (35) if and only if there is no value of σ at which the bounding functions in (35) intersect.

The arguments set out above lead to the following theorem.

THEOREM 6. *In the Model $\mathcal{M}$ let the threshold-crossing function p be a monotone function of a scalar index $\alpha' X + \delta' Z$. The identified set of values of (α,δ) is the union of two sets of values,*

$$I_0^{\alpha\delta} = \left\{\alpha,\delta : \min_{\sigma\in\mathcal{X}} \left(d_{0u}^{\uparrow}(\sigma;\alpha,\delta) - d_{0l}^{\uparrow}(\sigma;\alpha,\delta)\right) \ge 0\right\}$$
$$\cup \left\{\alpha,\delta : \min_{\sigma\in\mathcal{X}} \left(d_{0u}^{\downarrow}(\sigma;\alpha,\delta) - d_{0l}^{\downarrow}(\sigma;\alpha,\delta)\right) \ge 0\right\}.$$

Associated with each element $(\alpha,\delta) \in I_0^{\alpha\delta}$ there is an identified set of monotone functions p, the union of two sets,

$$\left\{p : d_{0l}^{\uparrow}(\sigma;\alpha,\delta) \le p(\sigma) \le d_{0u}^{\uparrow}(\sigma;a,\delta),\ \forall \sigma \in \mathcal{X}\right\}$$
$$\cup \left\{p : d_{0l}^{\downarrow}(\sigma;\alpha,\delta) \le p(\sigma) \le d_{0u}^{\downarrow}(\sigma;\alpha,\delta),\ \forall \sigma \in \mathcal{X}\right\}.$$

https://doi.org/10.1017/S0266466612000321 Published online by Cambridge University Press

The monotonicity restriction delivers big computational benefits because it allows the identified set of index coefficient values to be characterized without calculating the identified set of threshold-crossing functions.

4.4. Illustration: Specification

The probability measures used in this illustration are, as earlier, generated by triangular structures with one endogenous variable.

$$Y = 1[-a_1 X - Z_2 + W > 0] \qquad X = b_0 + b_1 Z_1 + b_2 Z_2 + V$$

$$\begin{bmatrix} W \\ V \end{bmatrix} \perp\!\!\!\perp Z \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & s_{wv} \\ s_{wv} & s_{vv} \end{bmatrix} \right).$$

There are two exogenous variables, $Z = (Z_1, Z_2)$, and Z_1 is excluded from the structural function for Y. The coefficient on the included exogenous variable, Z_2, is normalized to -1. There is

$$Y = 1[U > \Phi(a_1 X + Z_2)],$$

where $U \equiv \Phi(W) \sim Unif(0, 1)$, and so the structural function determining Y is

$$Y = \begin{cases} 0\,, & 0 < U \leq \Phi(a_1 X + Z_2) \\ 1\,, & \Phi(a_1 X + Z_2) < U \leq 1, \end{cases}$$

while the semiparametric model whose identifying power is considered has the restrictions

$$Y = \begin{cases} 0\,, & 0 < U \leq p(a_1 X + Z_2) \\ 1\,, & p(a_1 X + Z_2) < U \leq 1 \end{cases} \qquad U \perp\!\!\!\perp Z, \qquad p \text{ monotone.}$$

To calculate the identified set, the joint distribution of Y^* and $a_1 X$ given $Z = (z_1, z_2)$ is required. Here a_1 is a trial value for inclusion in the identified set $I_0^{a\delta}$. The distribution is $N(\mu, \Sigma)$ with parameters

$$\mu \equiv \begin{bmatrix} a_0 + a_1 b_0 + a_1 b_1 z_1 + (a_1 b_2 - 1) z_2 \\ a_1 (b_0 + b_1 z_1 + b_2 z_2) \end{bmatrix},$$

$$\Sigma \equiv \begin{bmatrix} 1 + 2a_1 s_{wv} + a_1^2 s_{vv} & a_1 (s_{wv} + a_1 s_{vv}) \\ a_1 (s_{wv} + a_1 s_{vv}) & a_1^2 s_{vv} \end{bmatrix}.$$

Given this distribution, it is straightforward to compute the bounding functions (30), (31), (32), and (33) as bivariate normal orthant probabilities. The envelope bounding functions that appear in (34) and (35) are obtained by finding minimum and maximum values for $z \equiv (z_1, z_2) \in \mathcal{Z} \equiv \mathcal{Z}_1 \times \mathcal{Z}_2$. In this illustration $\mathcal{Z}_1 = [-2, 2]$ and two intervals $\mathcal{Z}_2$ are considered: $[-2, 2]$ and $[-3, 3]$.

https://doi.org/10.1017/S0266466612000321 Published online by Cambridge University Press

The parameter values used in the illustrative calculations are

$$a_1 = 1 \qquad b_0 = 0 \qquad b_1 \in [0.17, 1.5] \qquad b_2 = 0 \qquad s_{wv} = 0.5 \qquad s_{vv} = 1.$$

The coefficient b_2 is zero, so in this illustration X is uncorrelated with Z_2. The variable Z_2 effectively provides a scale against which the impact of endogenous X on the index is measured. If Z_2 were not present, for example because it exhibited no variation at all or because d_2 were actually zero, then the model would not have any identifying power for α_1 without further restriction on the threshold function. This suggests that identified sets will be smaller when Z_2 exhibits more variation.

In the structures and at the parameter values employed in this illustration, the structural threshold function $p(\alpha_1 X + Z_2)$ is $\Phi(X + Z_2)$. So the probability distributions used in this exercise are generated by a monotone increasing threshold function with α_1 taking the value 1.

4.5. Illustration: Results

The identified sets are presented in Table 1. For small values of b_1 (the value of the coefficient on Z_1 in the equation for endogenous X) the identified sets are not connected; there is an interval containing negative values of α_1 and an interval containing positive values, including of course the value $\alpha_1 = 1$. Intervals containing negative, respectively positive, values are associated with only monotone decreasing, respectively increasing, threshold functions.

TABLE 1. Identified sets for α_1 for a sequence of values of b_1 and two ranges of values of z_2. At small values of b_1, the set is the union of two disjoint sets, one containing negative values associated with monotone decreasing threshold functions, one containing positive values associated with monotone increasing threshold functions. L and U are, respectively, lower and upper limits of intervals.

	$z_2 \in [-2, 2]$				$z_2 \in [-3, 3]$			
	$p \downarrow$		$p \uparrow$		$p \downarrow$		$p \uparrow$	
b_1	L	U	L	U	L	U	L	U
0.170	−2.38	−0.38	0.11	8.69	−2.38	−0.21	0.09	8.14
0.175	−2.21	−0.41	0.11	8.41	−2.16	−0.26	0.10	7.90
0.190	−1.57	−0.49	0.12	7.67	−1.69	−0.37	0.11	7.26
0.250			0.17	5.72			0.15	5.51
0.500			0.33	3.15			0.31	2.94
0.750			0.45	2.82			0.44	2.21
1.000			0.54	2.83			0.54	1.83
1.500			0.60	2.87			0.66	1.75
2.000			0.60	2.80			0.70	1.77

https://doi.org/10.1017/S0266466612000321 Published online by Cambridge University Press

The value $\alpha_1 = 0$ and values close to zero do not lie in the identified set. This is because in the structure that generates the probability measure in this illustration the distribution of the outcome Y conditional on Z does depend on the value of Z.

For values of b_1 larger than around 0.2, the identified set is connected, containing only positive values of α_1, associated with monotone increasing threshold functions. The size of the identified set decreases as the value of b_1 increases. That reduction reduces as b_1 increases. Substantial further reductions in the size of the identified set can only be achieved by increasing the predictive power of the instrument, that is by reducing s_{vv}. As anticipated, identified sets are smaller when the range of Z_2, the exogenous variable in the index in the structural equation, is large.

4.6. Application

These identification results motivate estimators of identified sets. These may be calculated because they are of interest in their own right or to gain information on the robustness of estimates of parameters that are point identified by more restrictive models.

For example, in Blundell and Powell (2004) there are control function-based estimates of point-identified index parameters in a semiparametric model in which the outcome of interest is a binary measure of labor force participation of married male heads of households, the potentially endogenous variable is other household income, and the instrumental variables are measures of the education of the male (included in the outcome equation) and the education of his spouse and benefit income. This point-identifying model is a restricted version of the set-identifying IV model used here. Comparing an estimate of the identified set developed in this paper with the control function-based point estimate gives an idea of the information content of the control function restriction.

The same approach could be taken to investigate the sensitivity of the estimates in Rothe (2009), in which the outcome of interest is a binary indicator of married couples' home ownership and the explanatory variable of interest is observed household income. This is believed to be endogenous because it is an error-contaminated measure of permanent household income that the economic model proposes as the determinant of home ownership. Exogenous variables included in the outcome equation are the age of the male household member and a binary indicator for the presence of children. The excluded instrumental variables are the wife's employment status and education level.

Regarding estimation, a component of the identified set of index coefficients, for example $A_0^{\uparrow}(\alpha, \delta)$ defined in (36), can be estimated as the set of values of (α, δ) that deliver a zero value of the function

$$\hat{D}_0^{\uparrow}(\alpha, \delta) \equiv \max\left(\max_{\sigma \in \mathcal{X}}\left(\hat{d}_{0l}^{\uparrow}(\sigma; \alpha, \delta) - \hat{d}_{0h}^{\uparrow}(\sigma; \alpha, \delta)\right), 0\right),$$

https://doi.org/10.1017/S0266466612000321 Published online by Cambridge University Press

where $\hat{d}_{0l}^{\uparrow}(\sigma;\alpha,\delta)$ and $\hat{d}_{0h}^{\uparrow}(\sigma;\alpha,\delta)$ are estimates of the envelope functions (34) and (35). When the estimated component of the identified set is nonempty, this is the set of values of (α,δ) that minimize the function $\hat{D}_0^{\uparrow}(\alpha,\delta)$.

With data $\{Y_i, X_i, Z_i\}_{i=1}^N$ there are kernel estimators of the envelope functions, for example

$$\hat{d}_{0l}^{\uparrow}(\sigma;\alpha,\delta) \equiv \max_{z\in\mathcal{Z}} \left\{ \sum_{i=1}^{N} B_i(\sigma,z;\alpha,\delta) K_\rho(z-Z_i) \Big/ \sum_{i=1}^{N} K_\rho(z-Z_i) \right\}, \tag{38}$$

where K_ρ is a kernel function with bandwidth parameter ρ and binary variables B_i are defined as

$$B_i(\sigma,z;\alpha,\delta) \equiv 1[Y_i = 0 \wedge \alpha' X_i \leq \sigma - \delta' z].$$

The approach to reducing bias caused by the nonlinear max operator in (38) proposed in Chernozhukov, Lee, and Rosen (2011) could be employed. Tests of hypotheses on index coefficients can be performed using the method for identified sets defined by conditional moment inequalities set out in that paper.

5. CONCLUDING REMARKS

An incomplete single-equation IV threshold-crossing model for a binary response is generally partially, not point, identifying for the threshold function even when the function is parametrically restricted. In this paper sharp identified sets have been characterized for the case in which endogenous variables are discrete and a constructive proof of sharpness has been provided.

Most parametric models for binary outcomes satisfy a monotone index restriction under which the threshold function is a monotone function of a linear index constructed from explanatory variables. It has been shown that in a semiparametric model this shape restriction considerably simplifies the characterization of the identified set, leading to substantial computational benefits. In particular, the identified set of index coefficients can be obtained without calculating the set of threshold-crossing functions, placing semiparametric IV estimation of index coefficients in binary response models on the same computational footing as obtains in parametric IV estimation.

The results of this paper can be used to develop identified sets in monotone index binary response models with mismeasured explanatory variables. Consider a threshold-crossing model with a monotone index restriction as in Section 4.3,

$$Y = \begin{cases} 0\,, & 0 < U \leq p(\alpha' X + \delta' Z) \\ 1\,, & p(\alpha' X + \delta' Z) < U \leq 1 \end{cases}, \qquad U \perp\!\!\!\perp Z, \qquad p \text{ monotone},$$

and suppose that $\tilde{X} \equiv X + V$ is observed in place of X where V is measurement error and (U, V) and Z are independently distributed. Considering the case

https://doi.org/10.1017/S0266466612000321 Published online by Cambridge University Press

in which p is increasing, define $W \equiv p^{-1}(U) + \alpha' V$, let $q(\cdot)$ denote the distribution function of W, and define $\tilde{U} \equiv q(W)$, which is uniformly distributed on $[0, 1]$.

Written in terms of $\tilde{U}$, $\tilde{X}$, and q, the model for Y with measurement error and observable $\tilde{X}$ is then

$$Y = \begin{cases} 0\,, & 0 < \tilde{U} \leq q(\alpha' \tilde{X} + \delta' Z) \\ 1\,, & q(\alpha' \tilde{X} + \delta' Z) < \tilde{U} \leq 1 \end{cases}, \quad \tilde{U} \perp\!\!\!\perp Z, \quad q \text{ monotone},$$

which is identical to the original monotone index model for Y with unobserved X except that the threshold function is now q rather than the original p and X is replaced by the observed $\tilde{X}$.[16] The analysis of Section 4.3 applies and delivers the sharp identified set for (α, δ). In some applications the endogeneity of the observed explanatory variable may arise only because of measurement error. The analysis of this paper applies then and also when there is additionally endogeneity in the unobserved error-free explanatory variables.

The monotone index restriction applies in many parametric models for nonbinary ordered outcomes, for example ordered probit models and many count data models. Consider a threshold-crossing model for an M-valued ordered outcome with $Y = m$ if and only if $p_{m-1}(x) < U \leq p_m(x)$, where x is a list of explanatory variables. For each value $m \in \{1, \ldots, M-1\}$, there is a threshold-crossing model for the binary outcome $Y_m \equiv 1[Y > m]$ in which the threshold function is $p_m(x)$ and the methods of this paper apply to each of these models. With semiparametric or parametric restrictions, there may be common parameters in the $M-1$ threshold-crossing functions, and there may be monotone index restrictions. The identified sets developed using the methods of this paper can be intersected to deliver outer regions. Sharp set identification would require simultaneous consideration of the $M-1$ binary responses, which are all driven by a common latent variable, but the monotone index restriction would not be so easy to bring into play in such an analysis. The outer regions obtained using the methods of this paper could be useful in their own right or as a step on the way to calculating the sharp identified set.

When the predictive power of instrumental variables for the endogenous variable is not great, the identified sets delivered by the IV model can be large. In this situation the identifying power of the additional restrictions embodied in the triangular model that motivates control function estimation or in special regressor restrictions is very substantial. This is all to the good if those restrictions are plausible in the application being considered. But if there is doubt about their validity, then it is prudent to consider the sets identified by the single-equation IV model studied here. Even when the IV model's identified sets are large, they can carry useful information on the sign of the effect of potentially endogenous explanatory variables.

https://doi.org/10.1017/S0266466612000321 Published online by Cambridge University Press

NOTES

1. Chesher and Rosen (2012) explore the relationship between model completeness, coherence, and point identification.

2. A special regressor model with instrument independence requires U and the instruments Z to be independently distributed in place of the uncorrelatedness restriction $E[UZ] = 0$ proposed in Lewbel (2000).

3. In the absence of any further restrictions on γ, the support set $\mathcal{X}$ could be normalized and set equal to $\{1, 2, \ldots, K\}$.

4. This restriction could be dropped at the cost of complicating the notation.

5. It follows that the hypothesis that the $p(x)$ does not depend on endogenous x can be examined using a standard test for lack of dependence of $\Pr_0[Y = 0|Z = z]$ on z. For example, if Z were discrete, a standard Pearson contingency table test could be employed.

6. This incorporates a normalization, namely that U is marginally uniformly distributed. The point is that the distribution function of U given $Z = z$ alone must be independent of z.

7. This subset is convex because each subset is an intersection of linear half planes.

8. See Coxeter (1973) for properties of orthoschemes. There is the following definition: $O_{ijk} \equiv \{\gamma : \gamma_i \leq \gamma_j \leq \gamma_k\}$.

9. See the discussion following the proof of Theorem 2 in Section 3.1.

10. In the structures generating the probabilities employed in this example, $p(x)$ is monotone decreasing, so we expect the orthoscheme in which p is monotone increasing to be one of the first to go as the instrument is strengthened.

11. The characterization of the sharp identified set given in Theorem 4 for discrete X is used. This agrees with the characterization for continuous X given in Chesher (2010).

12. For weakly monotonic functions p, define p^{-1} as

$$p \text{ increasing: } p^{-1}(u) \equiv \inf\{x : p(x) \geq u\},$$

$$p \text{ decreasing: } p^{-1}(u) \equiv \inf\{x : p(x) \leq u\},$$

and restrict increasing p to be càdlàg and decreasing p to be càglàd.

13. This structure with its linear equations and Gaussian latent variables is of the sort admitted by the triangular model underlying STATA's ivprobit command; see Statacorp (2007). The ML version of that command uses the Gaussian specification employed in this illustration.

14. The function pmvnorm in the mvtnorm package of R (R Core Development Team, 2011) is used.

15. For this purpose, a point of tangency is not a point of intersection.

16. If p is monotone decreasing, then $q(\alpha' \tilde{X} + \delta' Z)$ is replaced by $1 - q(\alpha' \tilde{X} + \delta' Z)$.

REFERENCES

Blundell R.W. & R.L. Matzkin (2010) Conditions for the existence of control functions in nonseparable simultaneous equations models. CeMMAP Working paper CWP28/10.

Blundell R.W. & J.L. Powell (2003) Endogeneity in nonparametric and semiparametric regression models. In M. Dewatripont, L.P. Hansen, and S.J. Turnovsky (eds.), *Advances in Economics and Econometrics: Theory and Applications, Eighth World Congress,* vol. II. Cambridge University Press.

Blundell, R.W. & J.L. Powell (2004) Endogeneity in semiparametric binary response models. *Review of Economic Studies* 71, 655–679.

Chernozhukov, V., S. Lee, & A. Rosen (2011) Intersection bounds: Estimation and inference. CeMMAP Working paper 34/11.

Chesher, A.D. (2003) Identification in nonseparable models. *Econometrica* 71, 1405–1441.

Chesher, A.D. (2005) Nonparametric identification under discrete variation. *Econometrica* 73, 1525–1550.

https://doi.org/10.1017/S0266466612000321 Published online by Cambridge University Press

Chesher, A.D. (2007) Identification of nonadditive structural functions. In R. Blundell, T. Persson, & W. Newey (eds.), *Advances in Economics and Econometrics, Theory and Applications, 9th World Congress,* vol III. Cambridge University Press.
Chesher, A.D. (2010) Instrumental variable models for discrete outcomes. *Econometrica* 78, 575–601.
Chesher, A.D. & A. Rosen (2012) Simultaneous equations models for discrete outcomes: Completeness, coherence and identification. In preparation.
Chesher, A.D., A. Rosen, & K. Smolinski, (2011) An instrumental variable model of multiple discrete choice. CeMMAP Working paper 39/11.
Chesher, A.D. & K. Smolinski, (2012) IV models of ordered choice. *Journal of Econometrics* 166, 33–48.
Coxeter, H.S.M. (1973) *Regular Polytopes*. Dover.
Florens, J-P., J.J. Heckman, C. Meghir, & E.J. Vytlacil (2008) Identification of treatment effects using control functions in models with continuous endogenous treatment and heterogeneous effects. *Econometrica* 76, 1191–1206.
Greene, W. (2007) *LIMDEP 9.0 Reference Guide*. Econometric Software, Inc.
Hausman, J.A. (1978) Specification tests in econometrics. *Econometrica* 46, 1251–1271.
Heckman, J.J. (1978) Dummy endogenous variables in a simultaneous equations system. *Econometrica* 46, 931–959.
Heckman, J.J. (1979) Sample selection bias as a specification error. *Econometrica* 47, 153–161.
Heckman, J.J. (1990) Varieties of selection bias. *American Economic Review* 80*, Papers and Proceedings of the 102nd Annual Meeting of the AEA,* 313–318.
Hoderlein, S. (2009) Endogenous semiparametric binary choice models with heteroskedasticity. CeMMAP Working paper CWP34/09.
Imbens, G.W. & W.K. Newey (2009) Identification and estimation of triangular simultaneous equations models without additivity. *Econometrica* 77, 1481–1512.
Lewbel, A. (2000) Semiparametric qualitative response model estimation with unknown heteroscedasticity or instrumental variables. *Journal of Econometrics* 97, 145–177.
Magnac, T. & E. Maurin (2007) Identification and information in monotone binary models. *Journal of Econometrics* 139, 76–104.
Magnac, T. & E. Maurin (2008) Partial identification in monotone binary models: Discrete regressors and interval data. *Review of Economic Studies* 75, 835–864.
R Core Development Team (2011) *R: A Language and Environment for Statistical Computing,* R Foundation for Statistical Computing.
Rivers, D. & Q. Vuong (1988) Limited information estimators and exogeneity tests for simultaneous probit models. *Journal of Econometrics* 39, 347–366.
Rothe, C. (2009) Semiparametric estimation of binary response models with endogenous regressors. *Journal of Econometrics* 153, 51–64.
Shaikh, A.M. & E.J. Vytlacil (2011) Partial identification in triangular systems of equations with binary dependent variable. *Econometrica* 79, 949–955.
Smith, R.J. & R.W. Blundell (1986) An exogeneity test for a simultaneous equation tobit model with an application to labor supply. *Econometrica* 54, 679-685.
Statacorp (2007) Stata Statistical Software: Release 10. StataCorp LP.
Vytlacil, E.J. & N. Yildiz (2007) Dummy endogenous variables in weakly separable models. *Econometrica* 75, 757–779.

APPENDIX: Proof of Theorem 3

Theorem 3 is proved by constructing a conditional distribution function with the required properties. The construction builds on a suggestion of Martin Cripps. The construction is done for a representative value $z \in \mathcal{Z}$ and a particular permutation of $\gamma = \{\gamma_k\}_{k=1}^{K}$. Without

https://doi.org/10.1017/S0266466612000321 Published online by Cambridge University Press

loss of generality, it is assumed that indices are assigned to the elements of $\mathcal{X}$ so that $\gamma_1 \leq \gamma_2 \leq \cdots \leq \gamma_K$. Assume the system of inequalities associated with this permutation given in equation (10) holds. The systems of inequalities associated with other permutations of γ are obtained by exchange of indices. Define $\gamma_0 \equiv 0$ and $\gamma_{K+1} \equiv 1$.

The notation is simplified in two respects. First, the superscript "0", which indicates a probability calculated using the probability distributions $F^0_{YX|Z}$ delivered by a structure S_0, is omitted. Second, dependence of various conditional probabilities on z is not made explicit in the notation. Thus $\delta^0_k(z)$ is written as δ_k and $\alpha^0_k(z)$ is written as α_k.

Define

$$\bar{\alpha} \equiv \Pr[Y = 0|Z = z] = \sum_{j=1}^{K} \delta_j \alpha_j.$$

For all $k \in \{0, 1, \ldots, K+1\}$ define

$$\tilde{\gamma}_k = \min(\gamma_k, \bar{\alpha}) \qquad \hat{\gamma}_k = \max(0, \gamma_k - \bar{\alpha}),$$

and note that

$$\tilde{\gamma}_k + \hat{\gamma}_k = \gamma_k,$$

and that the inequalities (10) imply the inequality

$$\gamma_1 \leq \bar{\alpha} \leq \gamma_K.$$

Define $(K+2) \times K$ arrays $[\delta_j \tilde{\beta}_{kj}]$ and $[\delta_j \hat{\beta}_{kj}]$ with elements (which depend on z) defined recursively for each $k \in \{0, 1, \ldots, K+1\}$ as follows as j ascends through the sequence $\{1, \ldots, K\}$.

$$\delta_j \tilde{\beta}_{kj} = \min\left\{\delta_j \alpha_j, \max\left\{0, \tilde{\gamma}_k - \sum_{s=1}^{j-1} \delta_s \tilde{\beta}_{ks}\right\}\right\}, \tag{A.1}$$

$$\delta_j \hat{\beta}_{kj} = \min\left\{\delta_j \left(1 - \alpha_j\right), \max\left\{0, \hat{\gamma}_k - \sum_{s=1}^{j-1} \delta_s \hat{\beta}_{ks}\right\}\right\}. \tag{A.2}$$

Define the required conditional distribution function at $u = \gamma_k$ for $k \in \{0, 1, \ldots, K+1\}$ as

$$F_{U|XZ}(\gamma_k|x_j, z) \equiv \tilde{\beta}_{kj} + \hat{\beta}_{kj}, \tag{A.3}$$

which implies $F_{U|XZ}(0|x_j, z) = 0$, $F_{U|XZ}(1|x_j, z) = 1$. The argument is as follows. Since $\gamma_0 \equiv 0$, $\tilde{\gamma}_0 = \hat{\gamma}_0 = 0$, so $\tilde{\beta}_{0j} = 0$ and $\hat{\beta}_{0j} = 0$ for all j, which yields $F_{U|XZ}(0|x_j, z) = 0$. Since $\gamma_{K+1} = 1$, $\tilde{\gamma}_{K+1} = \bar{\alpha}$ and $\hat{\gamma}_{K+1} = 1 - \bar{\alpha}$, so, for all j, $\tilde{\beta}_{K+1\,j} = \alpha_j$ and $\hat{\beta}_{K+1\,j} = 1 - \alpha_j$ which yields $F_{U|XZ}(1|x_j, z) = 1$. The distribution function is endowed with nondecreasing line segments between each successive distinct pair of elements in γ. Using linear segments delivers piecewise uniform conditional distributions of U given X and Z.

Before proceeding with the proof, it is helpful to describe the resulting arrays of conditional distribution function values.

For each value of k, as j increases, $\delta_j \tilde{\beta}_{kj}$ is assigned the value $\delta_j \alpha_j$ until a value of j is reached such that $\tilde{\gamma}_k - \sum_{s=1}^{j-1} \delta_s \alpha_s \leq \delta_j \alpha_j$. This is the value of j such that $\tilde{\gamma}_k - \sum_{s=1}^{j} \delta_s \alpha_s$

https://doi.org/10.1017/S0266466612000321 Published online by Cambridge University Press

≤ 0. At this value of j, denoted $\tilde{j}(k)$, $\delta_j \tilde{\beta}_{kj}$ is assigned the value $\tilde{\gamma}_k - \sum_{s=1}^{j-1} \delta_s \alpha_s$, which equals $\tilde{\gamma}_k - \sum_{s=1}^{j-1} \delta_s \tilde{\beta}_{ks}$, and values of $\delta_j \tilde{\beta}_{kj}$ for larger values of j are assigned the value zero. The result is that $\sum_{s=1}^{K} \delta_s \tilde{\beta}_{ks} = \tilde{\gamma}_k$. The function $\tilde{j}(k)$ has the representation

$$\tilde{j}(k) = \max \left\{ j : \tilde{\gamma}_k - \sum_{s=1}^{j-1} \delta_s \alpha_s \geq 0 \right\}. \tag{A.4}$$

Since the $\tilde{\gamma}_k$'s are a nondecreasing sequence, $\tilde{j}(k)$ is a nondecreasing function of k. It is shown below that for all j the sequence $\{\delta_j \tilde{\beta}_{kj}\}_{k=1}^{K}$ is nondecreasing.

For each value of k, as j increases, $\delta_j \hat{\beta}_{kj}$ is assigned the value $\delta_j \left(1 - \alpha_j\right)$ until a value of j is reached such that $\hat{\gamma}_k - \sum_{s=1}^{j-1} \delta_s \hat{\beta}_{ks} \leq \delta_j \left(1 - \alpha_j\right)$. This is the value of j such that $\hat{\gamma}_k - \sum_{s=1}^{j} \delta_s (1 - \alpha_s) \leq 0$. At this value of j, denoted $\widehat{j}(k)$, $\delta_j \hat{\beta}_{kj}$ is assigned the value $\hat{\gamma}_k - \sum_{s=1}^{j-1} \delta_s (1 - \alpha_s)$, which equals $\hat{\gamma}_k - \sum_{s=1}^{j-1} \delta_s \hat{\beta}_{ks}$, and values of $\delta_j \hat{\beta}_{kj}$ for larger values of j are assigned the value zero. The result is that $\sum_{s=1}^{K} \delta_s \hat{\beta}_{ks} = \hat{\gamma}_k$. The function $\widehat{j}(k)$ has the representation

$$\widehat{j}(k) = \max \left\{ j : \hat{\gamma}_k - \sum_{s=1}^{j-1} \delta_s \left(1 - \alpha_s\right) \geq 0 \right\}. \tag{A.5}$$

Since the $\hat{\gamma}_k$'s are a nondecreasing sequence, $\widehat{j}(k)$ is a nondecreasing function of k. At low values of k the value of $\hat{\gamma}_k$ can be zero, in which case $\widehat{j}(k) = 0$ and every element in $\{\delta_j \hat{\beta}_{kj}\}_{j=1}^{K}$ is zero. It is shown below that for all j, the sequence $\{\delta_j \hat{\beta}_{kj}\}_{k=1}^{K}$ is nondecreasing.

Here is an example, in which $K = 5$ with $\delta \equiv \{\delta_j\}_{j=1}^{K}$, $\alpha \equiv \{\alpha_j\}_{j=1}^{K}$ (for some value of z) and $\gamma \equiv \{\gamma_j\}_{j=1}^{K}$ take the following values:

$$\begin{aligned}
\delta &= \left[0.1\ 0.2\ 0.3\ 0.1\ 0.3\right] \\
\alpha &= \left[0.5\ 0.3\ 0.4\ 0.5\ 0.8\right] \\
\gamma &= \left[0.3\ 0.4\ 0.5\ 0.6\ 0.8\right].
\end{aligned}$$

The (7×5) arrays $[\delta_j \tilde{\beta}_{kj}]$ and $[\delta_j \hat{\beta}_{kj}]$ are

$$[\delta_j \tilde{\beta}_{kj}] = \begin{bmatrix}
0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\
0.05 & 0.06 & 0.12 & 0.05 & 0.02 \\
0.05 & 0.06 & 0.12 & 0.05 & 0.12 \\
0.05 & 0.06 & 0.12 & 0.05 & 0.22 \\
0.05 & 0.06 & 0.12 & 0.05 & 0.24 \\
0.05 & 0.06 & 0.12 & 0.05 & 0.24 \\
0.05 & 0.06 & 0.12 & 0.05 & 0.24
\end{bmatrix} \quad [\delta_j \hat{\beta}_{kj}] = \begin{bmatrix}
0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\
0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\
0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\
0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\
0.05 & 0.03 & 0.00 & 0.00 & 0.00 \\
0.05 & 0.14 & 0.09 & 0.00 & 0.00 \\
0.05 & 0.14 & 0.18 & 0.05 & 0.06
\end{bmatrix}.$$

The values of the constructed distribution functions of U conditional on $X = x \in \{x_1, \ldots, x_5\}$ (and $Z = z$) at the 7 values $\gamma_0, \ldots, \gamma_6$ are given in the columns of the (7×5) array $[\beta_{kj}]$ below, with the associated values $\gamma_0, \ldots, \gamma_6$ shown alongside.

https://doi.org/10.1017/S0266466612000321 Published online by Cambridge University Press

$$[\beta_{kj}] = \begin{bmatrix} 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.50 & 0.30 & 0.40 & 0.50 & 0.0\dot{6} \\ 0.50 & 0.30 & 0.40 & 0.50 & 0.40 \\ 0.50 & 0.30 & 0.40 & 0.50 & 0.7\dot{3} \\ 1.00 & 0.45 & 0.40 & 0.50 & 0.80 \\ 1.00 & 1.00 & 0.70 & 0.50 & 0.80 \\ 1.00 & 1.00 & 1.00 & 1.00 & 1.00 \end{bmatrix} \qquad [\gamma_k] = \begin{bmatrix} 0.0 \\ 0.3 \\ 0.4 \\ 0.6 \\ 0.6 \\ 0.8 \\ 1.0 \end{bmatrix}.$$

The proof now proceeds by showing the distribution function (A.3) is: (1) proper, (2) satisfies the independence restriction, and (3) has the observational equivalence property. It will be shown that the properness and independence conditions are satisfied by construction. Satisfaction of the observational equivalence condition relies on the elements of γ satisfying the system of inequalities (10).

1. Proper Conditional Distributions. The proposed conditional distribution functions are proper if, for all j,

$$0 \le \tilde{\beta}_{1j} \le \cdots \le \tilde{\beta}_{Kj},$$

$$0 \le \hat{\beta}_{1j} \le \cdots \le \hat{\beta}_{Kj},$$

and $\tilde{\beta}_{Kj} + \hat{\beta}_{Kj} \le 1$.

It is evident that all elements of the arrays $[\tilde{\beta}_{ij}]$ and $[\hat{\beta}_{ij}]$ are nonnegative. For all i, each element $\tilde{\beta}_{ij}$ is bounded above by α_j and each element $\hat{\beta}_{ij}$ is bounded above by $1-\alpha_j$, and so there can be no values of i and j at which $\tilde{\beta}_{ij} + \hat{\beta}_{ij}$ exceeds 1.

It is now shown that for all j and k, $\delta_j \tilde{\beta}_{kj} \le \delta_j \tilde{\beta}_{k+1\,j}$. Use is made of the fact that $\tilde{j}(k)$ is a nondecreasing function of k.

- If $j < \tilde{j}(k)$, then $j < \tilde{j}(k+1)$ so $\delta_j \tilde{\beta}_{kj} = \delta_j \tilde{\beta}_{k+1\,j} = \delta_j \alpha_j$.
- If $j > \tilde{j}(k)$, then $\delta_j \tilde{\beta}_{kj} = 0$ and, since all elements of the array $[\tilde{\beta}_{ij}]$ are nonnegative, $\delta_j \tilde{\beta}_{kj} \le \delta_j \tilde{\beta}_{k+1\,j}$.
- There remains only the possibility that $j = \tilde{j}(k)$. In this case $\delta_j \tilde{\beta}_{kj} = \tilde{\gamma}_k - \sum_{s=1}^{j-1} \delta_s \alpha_s \le \delta_j \alpha_j$.
  - If $j < \tilde{j}(k+1)$, then $\delta_j \tilde{\beta}_{k+1\,j} = \delta_j \alpha_j$ and $\delta_j \tilde{\beta}_{kj} \le \delta_j \tilde{\beta}_{k+1\,j}$.
  - Otherwise $j = \tilde{j}(k+1)$ and $\delta_j \tilde{\beta}_{k+1\,j} = \tilde{\gamma}_{k+1} - \sum_{s=1}^{j-1} \delta_s \alpha_s$, and since $\tilde{\gamma}_{k+1} \ge \tilde{\gamma}_k$, there is $\delta_j \tilde{\beta}_{kj} \le \delta_j \tilde{\beta}_{k+1\,j}$.

It is now shown that for all j and k, $\delta_j \hat{\beta}_{kj} \le \delta_j \hat{\beta}_{k+1\,j}$. Again use is made of the fact that $\widehat{j}(k)$ is a nondecreasing function of k.

- If $j < \widehat{j}(k)$, then $j < \widehat{j}(k+1)$, so $\delta_j \hat{\beta}_{kj} = \delta_j \hat{\beta}_{k+1\,j} = \delta_j (1-\alpha_j)$.
- If $j > \widehat{j}(k)$, then $\delta_j \hat{\beta}_{kj} = 0$ and, since all elements of the array $[\hat{\beta}_{ij}]$ are nonnegative, $\delta_j \hat{\beta}_{kj} \le \delta_j \hat{\beta}_{k+1\,j}$.

https://doi.org/10.1017/S0266466612000321 Published online by Cambridge University Press

- There remains only the possibility that $j = \widehat{j}(k)$. In this case, $\delta_j \hat{\beta}_{kj} = \hat{\gamma}_k - \sum_{s=1}^{j-1} \delta_s (1-\alpha_s) \leq \delta_j (1-\alpha_j)$.
  - If $j < \widehat{j}(k+1)$, then $\delta_j \hat{\beta}_{k+1\,j} = \delta_j (1-\alpha_j)$ and $\delta_j \hat{\beta}_{kj} \leq \delta_j \hat{\beta}_{k+1\,j}$.
  - Otherwise $j = \widehat{j}(k+1)$ and $\delta_j \hat{\beta}_{k+1\,j} = \tilde{\gamma}_{k+1} - \sum_{s=1}^{j-1} \delta_s (1-\alpha_s)$, and since $\hat{\gamma}_{k+1} \geq \hat{\gamma}_k$ there is $\delta_j \hat{\beta}_{kj} \leq \delta_j \hat{\beta}_{k+1\,j}$.

2. Independence. It was noted above that, for all k,

$$\sum_{j=1}^{K} \delta_j \tilde{\beta}_{kj} = \tilde{\gamma}_k \qquad \sum_{j=1}^{K} \delta_j \hat{\beta}_{kj} = \hat{\gamma}_k,$$

from which it follows that for all k,

$$\sum_{j=1}^{K} \delta_j \left(\tilde{\beta}_{kj} + \hat{\beta}_{kj} \right) = \tilde{\gamma}_k + \hat{\gamma}_k = \gamma_k,$$

as required.

3. Observational Equivalence. The observational equivalence property holds if, for all k,

$$\delta_k \tilde{\beta}_{kk} + \delta_k \hat{\beta}_{kk} = \delta_k \alpha_k.$$

The inequalities (10) can be written as

$$\sum_{j=1}^{k} \delta_j \alpha_j \leq \gamma_k \leq \bar{\alpha} + \sum_{j=1}^{k-1} \delta_j (1-\alpha_j). \tag{A.6}$$

There are two cases to consider.

First, suppose $\gamma_k \leq \bar{\alpha}$. Then $\tilde{\gamma}_k = \gamma_k$ and $\hat{\gamma}_k = 0$. From (A.6) there is, on substituting $\gamma_k = \tilde{\gamma}_k$,

$$\tilde{\gamma}_k - \sum_{j=1}^{k-1} \delta_j \alpha_j \geq \delta_k \alpha_k,$$

and so $\delta_k \tilde{\beta}_{kk} = \delta_k \alpha_k$. Since $\hat{\gamma}_k = 0$, $\delta_k \hat{\beta}_{kk} = 0$ and the result follows.

Now suppose that $\gamma_k > \bar{\alpha}$. Then $\tilde{\gamma}_k = \bar{\alpha}$, $\hat{\gamma}_k = \gamma_k - \bar{\alpha}$. Since $\bar{\alpha} - \sum_{j=1}^{k-1} \delta_j \alpha_j \geq \delta_k \alpha_k$, $\delta_k \tilde{\beta}_{kk} = \delta_k \alpha_k$. From (A.6) there is, on substituting $\gamma_k - \bar{\alpha} = \hat{\gamma}_k$,

$$\hat{\gamma}_k \leq \sum_{j=1}^{k-1} \delta_j (1-\alpha_j),$$

and so in the definition of $\delta_k \hat{\beta}_{kk}$, $\max \left\{ 0, \hat{\gamma}_k - \sum_{s=1}^{k-1} \delta_s \hat{\beta}_{ks} \right\} = 0$, so $\delta_k \hat{\beta}_{kk} = 0$ and the result follows.

https://doi.org/10.1017/S0266466612000321 Published online by Cambridge University Press