



IV models of ordered choice

Andrew Chesher*, Konrad Smolinski

CeMMAP & UCL, United Kingdom

ARTICLE INFO

Article history:

Available online 26 June 2011

JEL classification:

C10
C14
C50
C51

Keywords:

Endogeneity
Incomplete models
Instrumental variables
Ordered choice
Ordered probit
Set identification
Threshold crossing models

ABSTRACT

This paper studies single equation instrumental variable models of ordered choice in which explanatory variables may be endogenous. The models are weakly restrictive, leaving unspecified the mechanism that generates endogenous variables. These incomplete models are set, not point, identifying for parametrically (e.g. ordered probit) or nonparametrically specified structural functions. The paper gives results on the properties of the identified set for the case in which potentially endogenous explanatory variables are discrete. The results are used as the basis for calculations showing the rate of shrinkage of identified sets as the number of classes in which the outcome is categorised increases.

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

This paper studies single equation instrumental variable models for ordered outcomes in which explanatory variables may be endogenous. These models arise in structural econometric analysis of individuals' choices amongst ordered alternatives, or of individuals' attitudes arranged on an ordinal scale and they arise in many other settings in which data are interval censored continuous outcomes.

A common ploy when dealing with endogenous variation in a discrete response situation is to presume that the discrete response is generated in a recursive, triangular system along with the endogenous variable. Then, calling on some further restrictions, a control function method is used as the basis for identification and estimation. See for example Smith and Blundell (1986), Rivers and Vuong (1988), Blundell and Powell (2003, 2004) and Chesher (2003).¹

Unfortunately this control function strategy does not generally work when endogenous variables are discrete.² And, as explained

in Chesher (2009), the triangular model imposes strong restrictions on the process generating the endogenous variables, restrictions which may not be found plausible in many econometric settings. By contrast here we work with a limited information, single equation model which is far less restrictive in this regard.

The model requires that a scalar ordered outcome Y , with $M \geq 2$ points of support, is determined by a structural function $h(X, U)$ which is weakly monotone in scalar unobserved continuously distributed U . The observed vector of explanatory variables, X , and U may not be independently distributed. However, the model requires that U is distributed independently of instruments, Z . We call the model a *Single Equation Instrumental Variable* (SEIV) model. The SEIV model requires that the support of X does not depend on the value of U but places no other restrictions on the process generating the endogenous variables, X , and in this respect it is incomplete.

Thinking about Manski's "Law of Decreasing Credibility", Manski (2003), encourages us to take this approach. It allows one to see what is lost by relaxing the strong restrictions of the triangular control function model. It turns out that what is lost is point identification because the SEIV model is generally *set* not point identifying.

This paper focusses on models with discrete endogenous variables, having K points of support, $\{x_1, \dots, x_K\}$, and explores the identified sets the SEIV model delivers. The main results are now summarised.

Since the structural functions of an SEIV model are monotone in scalar U , there is a threshold crossing representation in which U is

* Corresponding author.

E-mail address: andrew.chesher@ucl.ac.uk (A. Chesher).

¹ The control function approach is used quite widely in applied econometric practice. STATA, *Statacorp* (2007) and LIMDEP, *Greene* (2007), are examples of widely used proprietary software suites armed with commands to conduct control function estimation of models of binary responses.

² Chesher (2005) gives partial identification results for a control function model with discrete endogenous variables.

normalised marginally uniformly distributed on the unit interval.

$$h(X, U) \equiv \begin{cases} 1 & , \quad 0 \leq U \leq h_1(X) \\ 2 & , \quad h_1(X) < U \leq h_2(X) \\ \vdots & \vdots \\ M & , \quad h_{M-1}(X) < U \leq 1. \end{cases}$$

In the discrete endogenous variable case a nonparametrically specified structural function, h , is characterised by $N = K \times (M - 1)$ parameters, denoted γ , which are the values of the $M - 1$ threshold functions at the K values of X .

Let $\mathcal{H}^0(\mathcal{Z})$ denote the set of values of γ identified by the SEIV model given $F_{YX|Z}^0$, a probability distribution for Y and X conditional on Z , when Z takes values in a set \mathcal{Z} . Each structural function is characterised by a point in the unit N -cube and $\mathcal{H}^0(\mathcal{Z})$ is a subset of that space.

The identified set delivered by a nonparametric SEIV model is shown to be a union of convex sets each defined by a system of linear equalities and inequalities. The number of sets involved can be enormous in what seem to be small scale problems. For example, when $M = K = 5$ there may be over 300 billion component sets. The result is generally not a convex set unless instruments are strong. We give examples in which the identified set is not convex and, indeed, not connected. Shape restrictions (e.g. monotonicity) or parametric restrictions can bring substantial simplification.

A system of inequalities given in Chesher (2010) defines an outer set, $\mathcal{C}^0(\mathcal{Z})$, within which the SEIV model's identified set lies. We develop expressions for these inequalities for the M outcome, discrete endogenous variable case. We propose a second system of inequalities defining a set of values of γ , $\mathcal{D}^0(\mathcal{Z})$, and show that the identified set resides in the intersection $\tilde{\mathcal{C}}^0(\mathcal{Z}) \equiv \mathcal{C}^0(\mathcal{Z}) \cap \mathcal{D}^0(\mathcal{Z})$.

When the outcome Y is binary $\mathcal{C}^0(\mathcal{Z})$ is a subset of $\mathcal{D}^0(\mathcal{Z})$ and, as shown in Chesher (2010), in that case $\mathcal{C}^0(\mathcal{Z})$ is the identified set $\mathcal{H}^0(\mathcal{Z})$. Here we show that when the endogenous variable is binary $\tilde{\mathcal{C}}^0(\mathcal{Z})$ is the identified set, however many categories there are for Y . The nonparametric SEIV model is falsifiable as the example in Appendix C demonstrates.

We examine the impact of response discreteness on the identified sets. The discrete response model studied here is a non-additive error model and the results for such models for continuous outcomes given in Chernozhukov and Hansen (2005) show that there can be point identification in SEIV models when observed responses are continuous. So it is to be expected that as the number of categories indicated by Y increases there is reduction in ambiguity and an approach to point identification.

We investigate this in the context of a model with parametrically specified structural functions such as arise in ordered probit models. We find that in the cases considered identified sets for a parameter which is a coefficient in a linear index shrink at a rate approximately equal to the inverse of the square of the number of classes in which the outcome is categorised. In the example, when Y is categorised into 10 or more classes, the SEIV model delivers identified sets which are very small indeed.

The paper is organised as follows. Section 2 gives a formal definition of the SEIV model and defines its identified set of structural functions.

Section 3 develops the main results for nonparametrically specified structural functions with discrete endogenous variables. In Section 3.1, a piecewise uniform system of conditional distributions of U given X and Z is introduced and conditions under which a structural function lies in the identified set are stated. The geometry of the identified set for nonparametrically specified structural functions is discussed in Section 3.2, the impact of shape restrictions on the complexity of the identified set is considered in Section 3.3, and systems of inequalities obeyed by values of these functions that lie in the identified set are set out in Section 3.4.

Section 4 illustrates the results using a parametrically specified model which, in the absence of endogeneity, would be a conventional ordered probit model. This section gives results on the rate of shrinkage of identified sets as the number of categories of the discrete outcome increases. It also demonstrates the impact of instrument support on the identified set. Section 5 concludes.

2. An IV model for ordered outcomes

In the SEIV model a scalar ordered outcome Y is determined by observable X , which may be a vector, and unobserved scalar U . Restriction 1 defines admissible structural functions.

Restriction 1. Y is determined by a structural function as follows:

$$Y = h(X, U) \equiv \begin{cases} 1 & , \quad h_0(X) \leq U \leq h_1(X) \\ 2 & , \quad h_1(X) < U \leq h_2(X) \\ \vdots & \vdots \\ M & , \quad h_{M-1}(X) < U \leq h_M(X) \end{cases}$$

with, for all x , $h_0(x) = 0$ and $h_M(x) = 1$ and for all x and m , $h_m(x) > h_{m-1}(x)$. U is continuously distributed and normalised to have a marginal uniform distribution on $[0, 1]$. The support of X conditional on U is independent of U .

Specifying the values of Y to be the first M integers is an innocuous normalisation because Y is an ordered outcome.

U and X are not required to be independently distributed so the model allows elements of X to be endogenous. However, U is required to be distributed independently of instrumental variables, Z , as set out in Restriction 2.

Restriction 2. There are instrumental variables Z which take values in some set \mathcal{Z} . U and Z are independently distributed in the sense that the conditional distribution function of U given $Z = z$ satisfies $F_{U|Z}(u|z) = u$ for all $u \in [0, 1]$ and $z \in \mathcal{Z}$.

Restriction 1 excludes the instrumental variables from the structural function. Now consider the identifying power of this model.

Let $F_{YX|Z}^0$ denote some distribution function of Y and X conditional on Z . Imagine a situation in which data are informative about this distribution for values of Z that lie in a set \mathcal{Z} . If this distribution function is compatible with the SEIV model, then there exists (i) a structural function h^0 with threshold functions $\{h_m^0\}_{m=1}^M$ and (ii) a distribution function $F_{UX|Z}^0$, both admitted by the SEIV model and such that the following condition holds when $h = h^0$ and $F_{UX|Z} = F_{UX|Z}^0$.

$$F_{YX|Z}^0(m, x|z) = F_{UX|Z}(h_m(x), x|z), \quad \text{for all: } z \in \mathcal{Z}, m \text{ and } x. \tag{1}$$

This relationship must hold because the threshold crossing Restriction 1 requires that Y is less than or equal to m when $X = x$ if and only if U is less than or equal to the m th threshold function $h_m(x)$.

There may be more than one admissible structure $(h, F_{UX|Z})$ satisfying (1) because it may be possible to compensate for variations in the x -sensitivity of the threshold functions $\{h_m\}_{m=1}^M$ by adjusting the u - and x -sensitivity of $F_{UX|Z}$ leaving the left-hand side of (1) unchanged while respecting the independence Restriction 2. So the model is partially identifying.

For a distribution $F_{YX|Z}^0$ let $\mathcal{H}^0(\mathcal{Z})$ denote the set of structures identified by the model comprising Restrictions 1 and 2. This is the set of structures admitted by the SEIV model and satisfying (1). The set of structural functions identified by the model, denoted $\mathcal{H}^0(\mathcal{Z})$,

is the set of structural functions h which are elements of structures lying in the identified set.

$$\mathcal{H}^0(\mathcal{Z}) \equiv \{h : \exists \text{ admissible } F_{U|X|Z} \text{ s.t. } (h, F_{U|X|Z}) \in \mathcal{S}^0(\mathcal{Z})\}.$$

The set $\mathcal{H}^0(\mathcal{Z})$ is a projection of the set $\mathcal{S}^0(\mathcal{Z})$.

This set is difficult to characterise and compute when X is continuously distributed because determining whether there exists a distribution function $F_{U|X|Z}$ that can accommodate a particular structural function may require searching across an infinite-dimensional space of functions.

However, Chesher (2010) shows that when Y is binary the identified set is determined by a system of inequalities in which the distribution function $F_{U|X|Z}$ does not appear. One of the contributions of this paper is a similar result for the case in which a scalar endogenous explanatory variable X is binary and Y takes any number of values.

When X is discrete, say with K points of support, the distribution function $F_{U|X|Z}$ can be characterised by a finite number of parameters for each value of Z and the identified set can be computed when M and K are not too large. The remainder of the paper studies the case in which the explanatory variable, X , is discrete.

3. Identified sets with discrete endogenous variables

3.1. Identification

When X is discrete and K -valued with $X \in \{x_i\}_{i=1}^K$, the threshold functions are characterised by the parameters

$$\gamma_{mi} \equiv h_m(x_i), \quad m \in \{0, \dots, M\}, \quad i \in \{1, \dots, K\} \quad (2)$$

of which $N \equiv (M - 1)K$ are free, that is not restricted to be zero or one. Define $\boldsymbol{\gamma}_i \equiv \{\gamma_{mi}\}_{m=0}^M$ and $\boldsymbol{\gamma} \equiv \{\boldsymbol{\gamma}_i\}_{i=1}^K$ with, for all $i \in \{1, \dots, K\}$, $\gamma_{0i} \equiv 0$, $\gamma_{Mi} \equiv 1$.

In the discrete X case considered here an identified set of structural functions is a set of values of $\boldsymbol{\gamma}$, comprising a subset of the unit N -cube.

When determining whether a structural function characterised by a value of $\boldsymbol{\gamma}$ lies in the identified set it is sufficient to search across distribution functions which, at each value z of the instrumental variables are characterised by the following parameters.

$$\beta_{mij}(z) \equiv F_{U|XZ}(\gamma_{mi}|x_j, z), \quad m \in \{0, 1, \dots, M\}, \quad (i, j) \in \{1, \dots, K\}.$$

Let $\boldsymbol{\beta}(z)$ denote the list of values $\beta_{mij}(z)$, $m \in \{0, \dots, M\}$, $(i, j) \in \{1, \dots, K\}$ for some value z . For all $(i, j) \in \{1, \dots, K\}$ define $\beta_{0ij}(z) \equiv 0$ and $\beta_{Mij}(z) \equiv 1$. Let $\boldsymbol{\beta}(\mathcal{Z})$ denote the list of values of $\boldsymbol{\beta}(z)$ generated as z varies across \mathcal{Z} .

Values $\beta_{mij}(z)$ with $i = j$ are relevant because observational equivalence requires that if $\boldsymbol{\gamma}$ lies in the identified set then for each $z \in \mathcal{Z}$, m and i the equality

$$F_{U|XZ}(\gamma_{mi}|x_i, z) = F_{Y|XZ}^0(m|x_i, z) \quad (3)$$

must hold. This follows directly from (1) on replacing x by x_i and $h_m(x_i)$ by γ_{mi} as in (2). The conditional distribution $F_{X|Z}^0$ is identified so (3) is effectively the observational equivalence condition (1).

Values of $\beta_{mij}(z)$ with $i \neq j$ are also relevant because the independence Restriction 2 together with the uniform distribution normalisation of the marginal distribution of U requires that for each m, i and z the following condition holds.

$$E_{X|Z=z}^0[F_{U|XZ}(\gamma_{mi}|X, z)] \equiv \sum_{j=1}^K F_{U|XZ}(\gamma_{mi}|x_j, z) \Pr_0[X = x_j|Z = z] = \gamma_{mi} \quad (4)$$

Here $E_{X|Z=z}^0$ indicates expectation taken with respect to the distribution $F_{X|Z}^0$ with the conditioning variable Z taking the value z .

So, for each point x_j in the support of X the values of the conditional distribution functions, $F_{U|XZ}(u|x_j, z)$, at each value of $u \in \mathcal{Y}$ are relevant when determining whether $\boldsymbol{\gamma}$ is in the identified set. Other values of u are not relevant because they play no role in the fulfilment of the observational equivalence condition (3) or the independence condition (4).

If γ_{mi} and $\gamma_{m'i'}$ are adjacent³ values of the threshold parameters then the definition of $F_{U|XZ}$ for any values, x_j and z of the conditioning variables can be completed by connecting $F_{U|XZ}(\gamma_{mi}|x_j, z)$ and $F_{U|XZ}(\gamma_{m'i'}|x_j, z)$ with straight line segments delivering histogram-like piecewise uniform conditional probability density functions.⁴

Let

$$\delta_i(z) \equiv \Pr[X = x_i|Z = z] \quad i \in \{1, \dots, K\}$$

be conditional probabilities of X given Z .

Let \Pr_0 denote probabilities calculated using a particular distribution function $F_{YX|Z}^0$ so that

$$\delta_i^0(z) \equiv \Pr_0[X = x_i|Z = z] \quad i \in \{1, \dots, K\}$$

and define $\boldsymbol{\delta}^0(z) \equiv \{\delta_i^0(z)\}_{i=1}^K$.

Define conditional probabilities (α) and cumulative probabilities ($\bar{\alpha}$) of the outcome:

$$\alpha_{mi}^0(z) \equiv \Pr_0[Y = m|X = x_i, Z = z], \quad m \in \{0, \dots, M\}, \quad i \in \{1, \dots, K\}$$

$$\bar{\alpha}_{mi}^0(z) \equiv \sum_{n=0}^m \alpha_{ni}^0(z), \quad m \in \{0, \dots, M\}, \quad i \in \{1, \dots, K\}$$

with $\alpha_{0i}^0(z) \equiv 0$ for all i and z , and define lists of conditional probabilities as follows.

$$\alpha_i^0(z) \equiv \{\alpha_{mi}^0(z)\}_{m=0}^M \quad \alpha^0(z) \equiv \{\alpha_i^0(z)\}_{i=1}^K$$

$$\bar{\alpha}_i^0(z) \equiv \{\bar{\alpha}_{mi}^0(z)\}_{m=0}^M \quad \bar{\alpha}^0(z) \equiv \{\bar{\alpha}_i^0(z)\}_{i=1}^K$$

Consider a structure characterised by

1. $\boldsymbol{\gamma}$: a list of values of threshold functions,
2. $\boldsymbol{\beta}(\mathcal{Z})$: a list of values of conditional distribution functions of U given X and Z obtained as Z takes values in \mathcal{Z} , and,
3. $\boldsymbol{\delta}(\mathcal{Z})$: a list of values of conditional probabilities of X given $Z = z$, $\boldsymbol{\delta}(z) = \{\delta_i(z)\}_{i=1}^K$ where $\delta_i(z) \equiv \Pr[X = x_i|Z = z]$, obtained as z varies across \mathcal{Z} .

Proposition 1. A structure, $\{\boldsymbol{\gamma}, \boldsymbol{\beta}(\mathcal{Z}), \boldsymbol{\delta}(\mathcal{Z})\}$ lies in the set identified by the SEIV model associated with probabilities $\alpha^0(z)$ and $\boldsymbol{\delta}^0(z)$ and a set of instrumental values \mathcal{Z} if and only if the following three conditions hold for all $z \in \mathcal{Z}$.

I1. Observational equivalence. For $m \in \{1, \dots, M\}$ and $i \in \{1, \dots, K\}$

$$\beta_{mii}(z) = \bar{\alpha}_{mi}^0(z) \quad \delta_i(z) = \delta_i^0(z)$$

I2. Independence. For $m \in \{1, \dots, M\}$ and $i \in \{1, \dots, K\}$

$$\sum_{j=1}^K \delta_j(z) \beta_{mij}(z) = \gamma_{mi}. \quad (5)$$

I3. Proper conditional distributions. For $(m, n) \in \{1, \dots, M\}$ and $(i, j, k) \in \{1, \dots, K\}$ if $\gamma_{mi} \leq \gamma_{nj}$ then $\beta_{mik}(z) \leq \beta_{njk}(z)$.

³ If there is no $\gamma_{st} \in \boldsymbol{\gamma}$ such that $\gamma_{mi} < \gamma_{st} < \gamma_{m'i'}$ then γ_{mi} and $\gamma_{m'i'}$ are adjacent.

⁴ Using straight line segments ensures that the independence condition:

$$E_{X|Z=z}^0[F_{U|XZ}(u|X, z)] = u$$

holds for all $u \in (0, 1)$ and $z \in \mathcal{Z}$.

Proof of Proposition 1. The SEIV model only admits distributions satisfying the monotonicity and independence conditions I3 and I2 so all structures in the set identified by the SEIV model must satisfy I2 and I3. Considering the set of structures that satisfy I2 and I3, it is precisely those that deliver the probability distribution $F_{Y|X|Z}^0$ for all values of Z in \mathcal{Z} that comprise the identified set.

Eq. (5) arises in I2 because on the left-hand side is the value of $\Pr[U \leq \gamma_{mi}|Z = z]$ associated with the distribution of U given X and Z characterised by $\beta(z)$ which the independence condition of the SEIV model requires to be equal to $\Pr[U \leq \gamma_{mi}]$. This probability is equal to γ_{mi} because the marginal distribution of U is normalised $Unif(0, 1)$ in the SEIV model.

In the observational equivalence condition I1, $\beta_{mi}(z)$ is the value of $\Pr[U \leq \gamma_{mi}|X = x_i, Z = z]$ associated with the distribution of U given X and Z characterised by $\beta(z)$. All structures and only structures that have this value equal to $\Pr_0[Y \leq m|X = x_i, Z = z_i]$ and $\delta_i(z) = \Pr_0[X = x_i|Z = z]$ deliver the probability distribution function $F_{Y|X|Z}^0$. \square

In Section 3.4, the identified set of values of γ is characterised as a set of values satisfying a system of inequalities. First, some aspects of the geometry of the identified set are examined.

3.2. Geometry of the identified set

When determining whether a particular value of γ lies in the identified set, the ordering of the elements of γ is crucial in determining whether there exist distribution functions which satisfy condition I3.

There are

$$L \equiv (K(M - 1))! / ((M - 1)!)^K \tag{6}$$

admissible orderings of the N elements of γ which are not restricted to be zero or one.⁵ For example, when $M = 3$ and $K = 2$, there are 6 of the possible 24 orderings that are admissible. The 18 inadmissible orderings have $\gamma_{11} > \gamma_{21}$ or $\gamma_{12} > \gamma_{22}$ or both.

Let l index the admissible orderings of γ . For each $l \in \{1, \dots, L\}$ define sets $\mathcal{S}_l^0(z)$ and $\mathcal{H}_l^0(z)$ as follows.

$$\mathcal{S}_l^0(z) \equiv \{(\gamma, \beta(z), \delta(z)) : \gamma \text{ is in order } l \text{ and } (\gamma, \beta(z), \delta(z)) \text{ respects I1–I3}\}$$

$$\mathcal{H}_l^0(z) \equiv \{\gamma : \gamma \text{ is in order } l \text{ and } \exists (\beta(z), \delta(z)) \text{ s.t. } (\gamma, \beta(z), \delta(z)) \in \mathcal{S}_l^0(z)\}.$$

The set $\mathcal{S}_l^0(z)$ is the set comprising those structures admitted by the SEIV model that have γ in order l and deliver the distribution $F_{Y|X|Z}^0$ for $Z = z$. The set $\mathcal{H}_l^0(z)$ is the projection of this set onto the component γ , that is onto the structural function.

Since for any ordering, l , conditions I1–I3 comprise a system of linear equalities and inequalities, each set $\mathcal{S}_l^0(z)$ is convex, or empty. It follows, from consideration of the Fourier–Motzkin elimination algorithm⁶, that the set $\mathcal{H}_l^0(z)$ is also defined by a system of linear equalities and inequalities, so it is also convex or empty.

The identified set of values of γ in order l obtained as z takes all values in the set of instrumental values \mathcal{Z} , denoted $\mathcal{H}_l^0(\mathcal{Z})$, is the following intersection of the sets $\mathcal{H}_l^0(z)$:

$$\mathcal{H}_l^0(\mathcal{Z}) \equiv \bigcap_{z \in \mathcal{Z}} \mathcal{H}_l^0(z)$$

⁵ There are $(K(M - 1))!$ permutations of the free elements of γ . Amongst these only 1 in each $(M - 1)!$ have a sequence γ_i in ascending order and there are K such sequences to be considered so only 1 in each $((M - 1)!)^K$ have all these sequences in ascending order.

⁶ See Ziegler (2007).

which is convex or empty. The identified set of values of γ of all orders is the union of the sets $\mathcal{H}_l^0(\mathcal{Z})$, as follows.

$$\mathcal{H}^0(\mathcal{Z}) = \bigcup_{l=1}^L \mathcal{H}_l^0(\mathcal{Z}).$$

Thus the identified set of values of γ , that is the identified set of structural functions, is a union of convex sets but that union may not itself be convex, indeed it may not even be connected. In the example in Section 4 there are a number of cases in which the identified set is disconnected.

When instruments are strong or there are highly informative additional restrictions (for example parametric restrictions) the sets $\mathcal{H}_l(\mathcal{Z})$ may be empty for all but one value of l and then the identified set is convex. Otherwise, the identified set may be very irregular and complex, composed of the union of a very large number of convex sets. With M and K as low as 4 the value of L is 369, 600 and as M or K increases the value of L quickly becomes astronomical.

Shape restrictions bring simplification as shown now.

3.3. Shape restrictions

3.3.1. Complete separation

In this case, there is the following restriction.

$$\min\{h_m(x_1), \dots, h_m(x_K)\} \geq \max\{h_{m-1}(x_1), \dots, h_{m-1}(x_K)\} \text{ for all } m.$$

Such completely separated threshold functions must arise when X has no effect on the structural function and will arise when the effect of X is weak. There are $K!$ arrangements of the elements of γ associated with each threshold function and $M - 1$ threshold functions so there are

$$L = (K!)^{M-1}$$

admissible orderings of the elements of γ under the complete separation restrictions. This result presumes that no elements in γ are equal. The value is an upper bound on the number of admissible orderings once ties are allowed. Henceforth, counts of orderings are conducted assuming no ties.

3.3.2. Monotonicity

Now suppose the threshold functions are all restricted to be monotone in a scalar explanatory variable X , with a common direction of dependence. This restriction would arise in a typical parametric ordered probit model.

We consider the case in which the direction of dependence is specified.⁷ The problem of finding the number of admissible orderings of γ under this restriction can be recast as the problem of finding the number of ways of filling a $(M - 1) \times K$ matrix with the integers $\{1, 2, \dots, N\}$ such that the array increases both across columns and across rows. With $K = 2$ this is the Catalan number $\frac{1}{M} \binom{2(M-1)}{(M-1)}$. The restriction of monotonicity with respect to X brings an M -fold reduction in the number of admissible orderings in this case.

These row and column ascending matrices are rectangular Young Tableaux and the hook length formula of Frame et al. (1954) delivers the following expression for the number of admissible orderings of γ under the monotonicity restriction.

$$L = \frac{(K(M - 1))!}{\prod_{m=1}^{M-1} \prod_{k=1}^K (M + K - m - k)} \tag{7}$$

Table 1 shows the number of admissible orderings (L) for values of M and K up to 5 with no shape restriction (None), the

⁷ Leaving this unspecified doubles the number of admissible arrangements.

Table 1
Number of admissible orderings of γ with no shape restrictions (None), monotonicity (M) with respect to scalar X and complete separation (CS).

M	Restriction	K			
		2	3	4	5
2	None	2	6	24	120
	M	1	1	1	1
	CS	2	6	24	120
3	None	6	90	2,520	113,400
	M	2	5	14	42
	CS	4	36	576	14,400
4	None	20	1,680	369,600	168,168,000
	M	5	42	462	6,006
	CS	8	216	13,824	1,728,000
5	None	70	34,650	6,306,300	305,540,235,000
	M	14	462	24,024	1,662,804
	CS	16	1,296	331,776	207,360,000

monotonicity restriction (M) and the complete separation restriction (CS). The restrictions bring substantial reductions in the number of admissible orderings but there remain very large numbers for only moderately large values of M and K .

3.3.3. Single- and twin-peakedness

Now suppose Y is binary so there is just one threshold function, h_1 . Let X be scalar and consider restrictions that limit the number of turning points of $h_1(x)$ as x passes through its ordered points of support denoted by

$$x_{[1]} < x_{[2]} < \dots < x_{[K]}.$$

Let $\gamma_{[i]}$ denote $h_1(x_{[i]})$.

When $h_1(x)$ is unrestricted there are $K!$ admissible orderings of elements of γ .⁸ When $h_1(x)$ is restricted to be monotone in x there are 2 admissible orderings. Monotone orderings have

$$\gamma_{[1]} < \gamma_{[2]} < \dots < \gamma_{[K]}$$

or

$$\gamma_{[1]} > \gamma_{[2]} > \dots > \gamma_{[K]}.$$

We now consider numbers of admissible orderings of elements of γ when h_1 is restricted to be (i) single-peaked and (ii) at most twin-peaked. Denote these respectively $c_1(K)$ and $c_2(K)$, and note that these are equal to the number of permutations of the integers $\{1, \dots, K\}$ which are respectively single- and at most twin-peaked.

Definition 1. A single-peaked permutation of $\{1, 2, \dots, K\}$ is a permutation in which either (i) elements are increasing, or (ii) elements are decreasing, or (iii) elements are first increasing and then decreasing.

Note that permutations that are monotone are classified as single-peaked, their peak is at one or the other end of the permutation.

Definition 2. An at most twin-peaked permutation of $\{1, 2, \dots, K\}$ is a permutation that is either single-peaked or has elements that: either (i) first decrease and then rise with at most one subsequent decrease, or (ii) first rise and then fall with at most one subsequent increase.

We drop the qualifier “at most” in most of the discussion that follows. One or both of the peaks in a twin-peaked permutation may appear at the start or the end of the permutation. Note that the

definitions designate all single-peaked permutations twin-peaked. Here are examples of twin-peaked permutations when $K = 5$:

$$\gamma_{[1]} < \gamma_{[2]} < \gamma_{[3]} < \gamma_{[4]} < \gamma_{[5]}$$

$$\gamma_{[1]} < \gamma_{[2]} > \gamma_{[3]} > \gamma_{[4]} > \gamma_{[5]}$$

$$\gamma_{[1]} > \gamma_{[2]} > \gamma_{[3]} < \gamma_{[4]} < \gamma_{[5]}$$

$$\gamma_{[1]} < \gamma_{[2]} > \gamma_{[3]} < \gamma_{[4]} > \gamma_{[5]}$$

and here is a permutation for $K = 5$ that is *not* twin-peaked

$$\gamma_{[1]} > \gamma_{[2]} < \gamma_{[3]} > \gamma_{[4]} < \gamma_{[5]}$$

there being 3 peaks, at $\gamma_{[1]}$, $\gamma_{[3]}$ and $\gamma_{[5]}$.

Here are two propositions with proofs giving the number of single- and twin-peaked permutations of the first K integers.

Proposition 2. The number of single-peaked permutations of $\{1, 2, \dots, K\}$ is $c_1(K) = 2^{(K-1)}$.

Proof of Proposition 2.⁹ First observe that whatever the value of $c_1(K - 1)$, the value of $c_1(K)$ is $2c_1(K - 1)$. This is so because for each single-peaked permutation of length $\{1, 2, \dots, K - 1\}$ there are exactly two single-peaked permutations of $\{1, 2, \dots, K\}$ constructed by placing the integer K to the left or to the right of the peak in the permutation with $K - 1$ elements. Placing the value K anywhere else in the permutation with $K - 1$ elements would introduce a second peak. Since $c_1(1) = 1$, it follows that $c_1(K) = 2^{K-1}$. □

Proposition 3. The number of at most twin-peaked permutations of $\{1, 2, \dots, K\}$ is

$$c_2(K) = 2^{(K-1)} \times (1 + 2^{K-2} - K/2).$$

Proof of Proposition 3. We first show that there is the following recursion.

$$c_2(K) = (K - 4) \times 2^{(K-2)} + 4 \times c_2(K - 1), \quad c_2(1) \equiv 1. \quad (8)$$

Consider permutations of $\{1, 2, \dots, K - 1\}$. Amongst these, by Proposition 2, there are $2^{(K-2)}$ that are single-peaked. Each of these can have the term K added at any one of K positions to deliver an at most twin-peaked sequence. So, associated with the single-peaked permutations of $\{1, 2, \dots, K - 1\}$ there are $c_{21}(K) = K \times 2^{(K-2)}$ at most twin-peaked permutations of $\{1, 2, \dots, K\}$.

There remain to be considered $c_2(K - 1) - 2^{(K-2)}$ twin-peaked but not single-peaked permutations of $\{1, 2, \dots, K - 1\}$. Each of these can have the term K added at any one of 4 positions, immediately before or after each of the two peaks. Any other placement would produce a third peak. So, associated with the twin-peaked but not single-peaked permutations of $\{1, 2, \dots, K - 1\}$ there are $c_{22}(K) = 4 \times (c_2(K - 1) - 2^{(K-2)})$ at most twin-peaked permutations of $\{1, 2, \dots, K\}$.

Inserting the term K into a permutation of $\{1, 2, \dots, K - 1\}$ that is not at most twin-peaked cannot deliver a twin-peaked permutation of $\{1, 2, \dots, K\}$.

The recursion is then obtained on adding $c_{21}(K)$ and $c_{22}(K)$ and simplifying.

On repeated substitution in (8) there is

$$c_2(K) = 4^{K-1} + 2^{K-2} \sum_{s=0}^{K-2} (K - 4 - s)2^s$$

and the result follows on using the following.

$$\sum_{s=0}^n 2^s = 2^{n+1} - 1 \quad \sum_{s=0}^n s2^s = 2(1 - 2^n + n2^n) \quad \square$$

⁸ We work supposing that all elements of γ are distinct. The counts of admissible arrangements reported are upper bounds on counts when there are ties.

⁹ We are grateful to Richard Spady for suggesting an inductive approach.

Table 2
Number of admissible orderings of γ in a model for binary Y with no restriction and with monotone, single- and twin-peakedness restrictions.

K	Restriction			
	Monotone	Single-peaked	Twin-peaked	None
2	2	2	2	2
3	2	4	6	6
4	2	8	24	24
5	2	16	104	120
6	2	32	448	720
7	2	64	1,888	5,040
8	2	128	7,808	40,320
9	2	256	31,872	362,880
10	2	512	129,024	3,628,800

Table 2 shows the number of admissible orderings of γ in a single threshold model without restrictions and under the monotone, single- and twin-peakedness conditions. Relaxing a monotonicity restriction to allow single-peakedness increases the number of admissible orderings by a factor 2^{K-2} . Further relaxation allowing twin peaks increases the number of admissible orderings by the same order of magnitude. When K is large even the relatively weak twin-peakedness restriction brings substantial reductions.

3.3.4. Two or more endogenous variables

Staying with the single threshold case, now suppose that there are two endogenous variables, X_1 and X_2 , with respectively K_1 and K_2 points of support. We will assume throughout that each value of X_1 can arise paired with each value of X_2 . There are then $(K_1 K_2)!$ admissible orderings of γ . If the threshold function $h_1(X_1, X_2)$ is restricted to be monotone (say increasing) in one of the variables, say X_1 , then there are $(K_1 K_2)! / (K_1!)^{K_2}$ admissible orderings, obtained from (6) on replacing $M - 1$ by K_1 . If the threshold function is restricted to be monotone in both variables (with directions of dependence specified) then on applying the hook length formula as above there are

$$L = \frac{(K_1 K_2)!}{\prod_{k_1=1}^{K_1} \prod_{k_2=1}^{K_2} (K_1 + K_2 + 1 - k_1 - k_2)}$$

admissible orderings of γ .

Finally suppose that an index restriction is imposed, for example that the single threshold function has the form $h_1(X_1 + X_2)$. Now we must consider how many elements are there in γ . Denoting this by N and staying with the single threshold case there is just one admissible ordering if monotonicity is imposed, $N!$ admissible orderings with no restriction on h_1 and $\sum_{i=1}^N (N - i + 1)^{i-1}$ under a single-peakedness condition.

Some subtleties arise on considering the value of N . Consider the case in which both variables have K points of support. The support of $X_1 + X_2$ contains at most K^2 distinct values but it can contain less. For example, if the support of the two variables are identical the support of $X_1 + X_2$ contains at most $K(K + 1)/2$ distinct values and if the support of each variable is a sequence of K consecutive integers then $X_1 + X_2$ takes $2K - 1$ distinct values. The impact of index restrictions on the complexity of the identified set depends critically on the nature of the support of the variables that contribute to the index.

3.4. Characterisation of the identified set

In this section a set of values of γ , $\tilde{\mathcal{C}}^0(\mathcal{Z})$, is defined as the intersection of two sets, $\mathcal{C}_0(\mathcal{Z})$ and $\mathcal{D}_0(\mathcal{Z})$. These sets are defined below. There then follow three propositions which set out the relationship between $\tilde{\mathcal{C}}^0(\mathcal{Z})$ and the identified set, $\mathcal{H}^0(\mathcal{Z})$.

Proposition 4 states that for all M and K , $\tilde{\mathcal{C}}^0(\mathcal{Z})$ is an outer region for the identified set, that is, $\mathcal{H}^0(\mathcal{Z}) \subseteq \tilde{\mathcal{C}}^0(\mathcal{Z})$.

Proposition 5 states that $\tilde{\mathcal{C}}^0(\mathcal{Z})$ is precisely the identified set when the endogenous variable is binary ($K = 2$). The relatively long proof is contained in an Annex.

Proposition 6 states that when Y is binary ($M = 2$), $\mathcal{C}_0(\mathcal{Z}) = \tilde{\mathcal{C}}^0(\mathcal{Z})$. This reconciles the results of this paper with those in Chesher (2010) where it is shown that when Y is binary $\mathcal{C}_0(\mathcal{Z})$ is precisely the identified set.

Here are the definitions of the sets $\mathcal{C}_0(\mathcal{Z})$ and $\mathcal{D}_0(\mathcal{Z})$.

$$\mathcal{C}^0(\mathcal{Z}) \equiv \left\{ \gamma : \max_{z \in \mathcal{Z}} \sum_{i=1}^K \sum_{m=1}^{M-1} \delta_i^0(z) \alpha_{mi}^0(z) 1(\gamma_{mi} \leq \gamma_{ls}) \leq \gamma_{ls} \leq \min_{z \in \mathcal{Z}} \sum_{i=1}^K \sum_{m=1}^M \delta_i^0(z) \alpha_{mi}^0(z) 1(\gamma_{m-1,i} < \gamma_{ls}), \right. \\ \left. l \in \{1, \dots, M - 1\}, s \in \{1, \dots, K\} \right\}$$

$$\mathcal{D}^0(\mathcal{Z}) \equiv \{ \gamma : \gamma_{ni} - \gamma_{mi} \geq \max_{z \in \mathcal{Z}} (\delta_i^0(z) (\bar{\alpha}_{ni}^0(z) - \bar{\alpha}_{mi}^0(z))) \}, \\ n > m, (n, m) \in \{1, \dots, M - 1\}, i \in \{1, \dots, K\}$$

Proposition 4. $\tilde{\mathcal{C}}^0(\mathcal{Z}) \equiv \mathcal{C}^0(\mathcal{Z}) \cap \mathcal{D}^0(\mathcal{Z})$ is an outer region for the identified set, $\mathcal{H}^0(\mathcal{Z})$.

Proof of Proposition 4. It is required to show that $\mathcal{H}^0(\mathcal{Z}) \subseteq \tilde{\mathcal{C}}^0(\mathcal{Z})$. This is done by showing in part (a) that all γ in the identified set lie in $\mathcal{C}^0(\mathcal{Z})$ and then in part (b) that all γ in the identified set lie in $\mathcal{D}^0(\mathcal{Z})$.

(a). $\mathcal{H}^0(\mathcal{Z}) \subseteq \mathcal{C}^0(\mathcal{Z})$.

Chesher (2010) shows that all structural functions, h , in the set identified by the SEIV model associated with a conditional distribution function $F_{YX|Z}^0$ and a set of instrumental values \mathcal{Z} satisfy the following inequalities for all $u \in (0, 1)$ and $z \in \mathcal{Z}$.

$$\Pr_0[Y < h(X, u) | Z = z] < u \leq \Pr_0[Y \leq h(X, u) | Z = z].$$

In terms of threshold functions these inequalities are as follows.

$$\sum_{m=1}^M \Pr_0[(Y = m) \wedge (h_m(x) < u) | Z = z] < u \\ \leq \sum_{m=1}^M \Pr_0[(Y = m) \wedge (h_{m-1}(x) < u) | Z = z].$$

For the discrete endogenous variable case, there is the following representation.

$$\sum_{i=1}^K \sum_{m=1}^{M-1} \delta_i^0(z) \alpha_{mi}^0(z) 1(\gamma_{mi} < u) < u \\ \leq \sum_{i=1}^K \sum_{m=1}^M \delta_i^0(z) \alpha_{mi}^0(z) 1(\gamma_{m-1,i} < u) \tag{9}$$

Consider some arrangement of the elements of γ in which two elements, $\gamma_{kr} < \gamma_{ls}$ are adjacent so that there is no element $\gamma_{qt} \in \gamma$ satisfying $\gamma_{kr} < \gamma_{qt} < \gamma_{ls}$. Consider $u \in (\gamma_{kr}, \gamma_{ls}]$ and the right-hand inequality in (9). This inequality must hold for all u in $(\gamma_{kr}, \gamma_{ls}]$ and so must hold at the supremum of the interval which is its maximal value, γ_{ls} , and so there is the following.

$$\gamma_{ls} \leq \sum_{i=1}^K \sum_{m=1}^M \delta_i^0(z) \alpha_{mi}^0(z) 1(\gamma_{m-1,i} < \gamma_{ls}).$$

Since this inequality must hold for all $z \in \mathcal{Z}$, the minimal value achieved on the right-hand side as z varies in \mathcal{Z} is all that matters.¹⁰ This produces the upper bounding function in the definition of $\mathcal{C}^0(\mathcal{Z})$.

Now consider some arrangement of the elements of $\boldsymbol{\gamma}$ in which two elements, $\gamma_{ls} < \gamma_{pr}$ are adjacent so that there is no element γ_{qt} satisfying $\gamma_{ls} < \gamma_{qt} < \gamma_{pr}$. Consider $u \in (\gamma_{ls}, \gamma_{pr}]$ and the left-hand side of (9). This inequality must hold for all u in $(\gamma_{ls}, \gamma_{pr}]$ and so must hold as in (9) with *strong* inequalities at every value of u in the interval and so with *weak* inequalities at the *infimum* of the interval which is γ_{ls} and so there is the following inequality.

$$\sum_{i=1}^K \sum_{m=1}^{M-1} \delta_i^0(z) \alpha_{mi}^0(z) 1(\gamma_{mi} \leq \gamma_{ls}) \leq \gamma_{ls}.$$

Since this must hold for all $z \in \mathcal{Z}$, the maximal value achieved on the left-hand side as z varies in \mathcal{Z} is all that matters. This delivers the lower bounding function in the definition of $\mathcal{C}^0(\mathcal{Z})$. This concludes the demonstration that $\mathcal{H}^0(\mathcal{Z}) \subseteq \mathcal{C}^0(\mathcal{Z})$.

(b). $\mathcal{H}^0(\mathcal{Z}) \subseteq \mathcal{D}^0(\mathcal{Z})$.

For each $\boldsymbol{\gamma}$ in the identified set, there exists for each $z \in \mathcal{Z}$ a distribution function characterised by $\boldsymbol{\beta}(z)$ satisfying conditions I1–I3. Conditions I1 and I2 imply that

$$\gamma_{ni} = \delta_i^0(z) \bar{\alpha}_{ni}^0(z) + \sum_{j \neq i} \delta_j^0(z) \beta_{nij}(z)$$

$$\gamma_{mi} = \delta_i^0(z) \bar{\alpha}_{mi}^0(z) + \sum_{j \neq i} \delta_j^0(z) \beta_{mij}(z)$$

and on subtracting there is the following.

$$\begin{aligned} \gamma_{ni} - \gamma_{mi} &= \delta_i^0(z) (\bar{\alpha}_{ni}^0(z) - \bar{\alpha}_{mi}^0(z)) \\ &+ \sum_{j \neq i} \delta_j^0(z) (\beta_{nij}(z) - \beta_{mij}(z)). \end{aligned}$$

The properness condition I3 ensures that, since $n > m$, for each i and j , the distribution function characterised by $\boldsymbol{\beta}(z)$ has $\beta_{nij}(z) \geq \beta_{mij}(z)$ and so

$$\gamma_{ni} - \gamma_{mi} \geq \delta_i^0(z) (\bar{\alpha}_{ni}^0(z) - \bar{\alpha}_{mi}^0(z)).$$

Since this inequality must hold for all $z \in \mathcal{Z}$, the maximal value achieved on the right-hand side as z varies in \mathcal{Z} is all that matters. This delivers the lower bounding function in the definition of $\mathcal{D}^0(\mathcal{Z})$. \square

Proposition 5. When X is binary $\mathcal{H}^0(\mathcal{Z}) = \tilde{\mathcal{C}}^0(\mathcal{Z})$ no matter how many points of support Y has.

This is proved by showing that for every $\boldsymbol{\gamma}$ in $\tilde{\mathcal{C}}^0(\mathcal{Z})$ and for each $z \in \mathcal{Z}$ there exists a proper distribution of U given X and $Z = z$ characterised by $\boldsymbol{\beta}(z)$ such that Restrictions I1–I3 are respected. The proof is in [Appendix A](#).

Proposition 6. When Y is binary $\tilde{\mathcal{C}}^0(\mathcal{Z}) = \mathcal{C}^0(\mathcal{Z})$.

Proof of Proposition 6. The result follows once it is shown that when Y is binary, $\mathcal{C}^0(\mathcal{Z}) \subseteq \mathcal{D}^0(\mathcal{Z})$. When Y is binary the free elements in $\boldsymbol{\gamma}$ are γ_{li} , $i \in \{1, \dots, K\}$. The inequalities defining $\mathcal{D}^0(\mathcal{Z})$ reduce to

$$\delta_i^0(z) \alpha_{li}^0(z) \leq \gamma_{li} \leq 1 + \delta_i^0(z) (1 - \alpha_{li}^0(z)) \quad i \in \{1, \dots, K\} \quad (10)$$

and the inequalities defining $\mathcal{C}^0(\mathcal{Z})$ require that

$$\begin{aligned} \sum_{j=1}^i \delta_j^0(z) \alpha_{lj}^0(z) \leq \gamma_{li} \leq 1 + \sum_{j=i}^K \delta_j^0(z) (1 - \alpha_{lj}^0(z)) \\ i \in \{1, \dots, K\}. \end{aligned} \quad (11)$$

It is clear that (10) is satisfied if (11) is satisfied, and therefore when Y is binary $\mathcal{C}^0(\mathcal{Z}) \subseteq \mathcal{D}^0(\mathcal{Z})$. \square

[Appendix B](#) sets out the 50 inequalities that arise in a 3 outcome case with a binary endogenous variable.

The SEIV model generally set identifies structural functions. When values of endogenous variables can be predicted with high accuracy identified sets can be small. The model is falsifiable as is demonstrated in [Appendix C](#) via a simple example involving a binary outcome.

The next section studies sets identified by an ordered outcome model with a binary endogenous variable. The effect on the identified set of varying the discreteness of the outcome and the strength of the instrumental variables is investigated.

4. Discreteness and identified sets in a parametric ordered probit model

The complexity of identified sets for nonparametrically specified threshold functions increases rapidly as the number of points of support of the outcome, M , becomes large. In this section we study cases with M as large as 130 when the identified set can in principle have over 2×10^{76} convex components. Cases like this are computationally challenging, and there arises the question of representing and processing the information in the complex identified set.

We tackle this problem by focusing on parametrically specified structural functions choosing, by way of example, the form that arises in conventional ordered probit analysis. We explain how identified sets in the parameter space defined by the parametric model can be calculated using our results on nonparametric identification but without encountering computationally intractable problems. For particular cases, we show the form taken by the identified sets and the effects on the sets of changing the discreteness of the outcome and the support of the instrumental variable.

The ordered probit structural function is interesting because with exogenous explanatory variables it is widely used in econometric practice and for applied workers a first attempt at addressing potential endogeneity in explanatory variables could start with this form. Given the complexity of the identified set for nonparametrically specified structural functions it is likely that in applications some relatively low-dimensional parameterisation would be used.

4.1. Models

We now investigate the nature of the identified sets delivered by a parametric ordered probit model with a binary endogenous variable. In this model the structural function is parametrically specified, as follows.

$$Y = \begin{cases} 1 & , \quad 0 \leq U \leq \Phi(s^{-1}(c_1 - a_0 - a_1 X)) \\ 2 & , \quad \Phi(s^{-1}(c_1 - a_0 - a_1 X)) < U \leq \Phi(s^{-1}(c_2 - a_0 - a_1 X)) \\ \vdots & \vdots \\ M & , \quad \Phi(s^{-1}(c_{M-1} - a_0 - a_1 X)) < U \leq 1 \end{cases} \quad (12)$$

Here Φ denotes the standard normal distribution function, the constants c_1, \dots, c_{M-1} are threshold values defining cells within which a latent normal random variable is classified, and a_0, a_1 and s are constant parameters. Throughout X is binary with support

¹⁰ The element in \mathcal{Z} which delivers this minimal value may vary with l and s .

$\{-1, +1\}$, There is the independence restriction: $U \perp Z$, U is normalised $Unif(0, 1)$.

In one portfolio of illustrations (A) the model specifies the values of the threshold parameters c_1, \dots, c_{M-1} as known, and s as known and normalised to one. This leaves just two unrestricted parameters, a_0 and a_1 , and it is easy to display the identified sets graphically. In these illustrations M , the number of levels of the outcome, is varied from 2 to 130. In econometric practice with interval censored data it would be normal to know the values of the thresholds but the parameter s would typically be unknown. The methods employed here could be used with s unknown.

In another illustration (B) M is held fixed at 3 and the model specifies the thresholds, c_1 and c_2 , along with the slope coefficient, a_1 , as unknown parameters. In these illustrations, the values of a_0 and s are normalised to respectively 0 and 1 because their values cannot be identified once the thresholds are unrestricted. A similar exercise could be conducted with larger values of M but the graphical display of results would be problematic. In econometric practice using attitudinal responses it would be normal to have unknown thresholds as in this example.

In all cases the instrumental variable takes equally spaced values in an interval $[-\zeta, \zeta]$. In two of the illustrations the effect on the identified sets of varying the support of the instrument is studied.

4.2. Calculation procedures

The calculation of an identified set of parameter values for a particular distribution $F_{Y|XZ}^0$ and a set of instrumental values Z proceeds as follows.¹¹

A fine grid of values of the parameters (e.g. a_0 and a_1 in the illustrations in set A) is defined. A value, (a_0^*, a_1^*) , is selected from the grid and the value of γ , say γ^* , determined by (a_0^*, a_1^*) is calculated. Recall that γ is a list of values of the threshold functions defined by a model at the points of support of the discrete endogenous variable.

For the selected value γ^* the ordering of its elements, say l^* , is determined and the linear equalities and inequalities defining the convex set $\tilde{C}_{l^*}^0(Z)$ are calculated, minimising and maximising expressions over $z \in Z$ as set out in the definitions of the sets $C_0(Z)$ and $\mathcal{D}_0(Z)$ at the start of Section 3.4.

In all the illustrations, because X is binary, $\mathcal{H}_{l^*}^0(Z) = \tilde{C}_{l^*}^0(Z)$. If γ^* falls in this set then (a_0^*, a_1^*) is in the identified set of parameter values, otherwise it is not.

Passing across the grid the identified set is computed. Care is required because the set may not be connected and sometimes component connected subsets of the identified set can be small. To avoid missing component subsets, dense grids of values are used in the calculations presented here.

4.3. Illustration A1

The probability distributions used in this illustration are generated by triangular Gaussian structures with structural equations as follows.

$$Y^* = \alpha_1 X + W$$

$$X^* = 0.5Z + V$$

$$Y = \begin{cases} 1 & , & -\infty \leq Y^* \leq c_1 \\ 2 & , & c_1 < Y^* \leq c_2 \\ \vdots & \vdots & \vdots \\ M & , & c_{M-1} < Y^* \leq +\infty \end{cases}$$

¹¹ All the computation in the paper was carried out using R (Ihaka and Gentleman (1996)).

Table 3
Illustration A1: threshold values.

M	Threshold values (c_i)	Shading in Fig. 1
2	{0.0}	Red
4	{ $\pm 0.1, 0.0$ }	Blue
6	{ $\pm 0.3, \pm 0.1, 0.0$ }	Red
8	{ $\pm 0.7, \pm 0.3, \pm 0.1, 0.0$ }	Blue
10	{ $\pm 1.1, \pm 0.7, \pm 0.3, \pm 0.1, 0.0$ }	Red
12	{ $\pm 1.5, \pm 1.1, \pm 0.7, \pm 0.3, \pm 0.1, 0.0$ }	Green
14	{ $\pm 1.8, \pm 1.5, \pm 1.1, \pm 0.7, \pm 0.3, \pm 0.1, 0.0$ }	Black

$$X = \begin{cases} -1 & , & -\infty \leq X^* \leq 0 \\ +1 & , & 0 < X^* \leq +\infty. \end{cases}$$

The value of α_1 in this illustration is 1 and the distribution of (W, V) is specified to be Gaussian and independent of Z .

$$\begin{bmatrix} W \\ V \end{bmatrix} | Z \sim N_2 \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1.0 & 0.5 \\ 0.5 & 1.0 \end{bmatrix} \right).$$

These structures are closely related to a special case of the parametric Gaussian models of discrete outcomes studied in Heckman (1978).

Expressed in terms of a random variable U which is uniformly distributed on the unit interval the structural functions are as follows.

$$h(X, U) = \begin{cases} 1 & , & 0 \leq U \leq \Phi(c_1 + X) \\ 2 & , & \Phi(c_1 + X) < U \leq \Phi(c_2 + X) \\ \vdots & \vdots & \vdots \\ M & , & \Phi(c_{M-1} + X) < U \leq 1. \end{cases}$$

Two sets of instrumental values, Z_1 and Z_2 , are considered each containing just 10 values evenly spaced in $[-1, 1]$ and in $[-2, 2]$ respectively.

$$Z_1 = \{\pm 1.0, \pm 0.777, \pm 0.555, \pm 0.333, \pm 0.111\}$$

$$Z_2 = \{\pm 2.0, \pm 1.556, \pm 1.111, \pm 0.667, \pm 0.222\}.$$

In the structures used to generate the probability distributions applied in this and the next illustration, the effect of doubling the values in the support of Z is the same as the effect of doubling the coefficient on Z in the equation for the latent variable X^* . The effect on the identified sets of such changes in the support or strength of the instrument are discussed shortly.

In this illustration the number of classes in which Y is observed is increased from 2 through 14 with threshold values as set out in Table 3.

Identified sets for the two parameters, (a_0, a_1) , are drawn in Fig. 1. The upper pane in the figure is obtained using Z_1 and the lower pane is obtained using Z_2 . The sets are rhombuses arranged with edges parallel to 45° and 225° lines. Identified sets are superimposed one upon another.

Using the set of instrumental values Z_2 with broader support delivers substantially smaller identified sets. In this case (but not in Illustration A2) the structure of the identified sets is very similar using instrumental values Z_1 and Z_2 , and the following discussion applies to the upper and the lower pane in Fig. 1.

The largest rhombus drawn in Fig. 1 is the identified set with $M = 2$. Since the outcome is binary, this is the set $C^0(Z)$.

The identified set with $M = 4$ is the rhombus comprising the lowest blue chevron and what lies above it but excluding a very narrow strip on the edge of the two upper boundaries. This strip (coloured green) is the set $C^0(Z) \cap \overline{\mathcal{D}^0(Z)}$. Notice that this does not extend all the way along the upper edges of the set because for the case $M = 2$, $\tilde{C}^0(Z) = C^0(Z) \subseteq \mathcal{D}^0(Z)$.

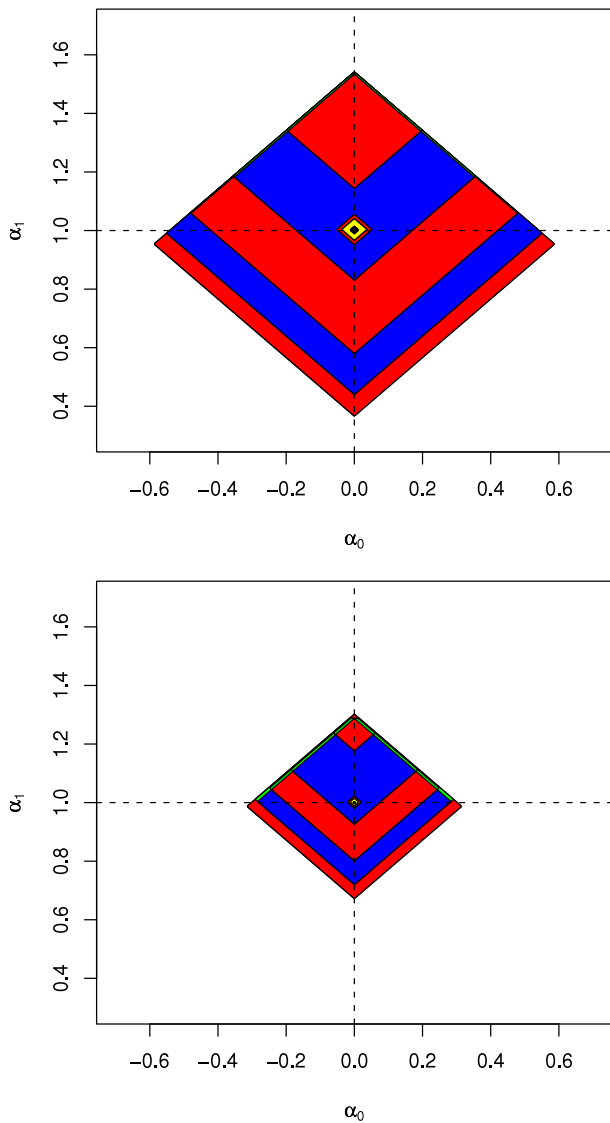


Fig. 1. Illustration A1. Outer sets and identified sets in a binary endogenous variable SEIV model with a parametric ordered probit structural function with threshold functions of the form $\Phi(c_i - a_0 - a_1x)$ as the number of categories of the outcome varies from 2 to 10. The upper (lower) pane represents identified sets when the support of the instrumental variable, Z , is Z_1 (Z_2). The dark blue strip at the upper margin of the rhombuses is not part of the identified sets. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

The identified set with $M = 6$ (respectively 8) is the rhombus comprising the second lowest red (respectively blue) chevron and all that lies above it apart from the narrow dark blue shaded strip on the edge of the two upper boundaries.

The identified set with $M = 10$ is disconnected and comprises the two small red shaded rhombuses in the upper part of the picture. The identified set when $M = 12$ is the small yellow shaded rhombus in the centre of the picture and the identified set when $M = 14$ is the tiny black shaded rhombus at the intersection of the horizontal and vertical dashed lines. Further increases in numbers of classes deliver sets which are barely distinguishable from points at the scale chosen for Fig. 1.

As the number of classes rises the extent of the identified sets falls rapidly but the passage towards point identification is complex and even when the sets are quite small they can be disconnected.

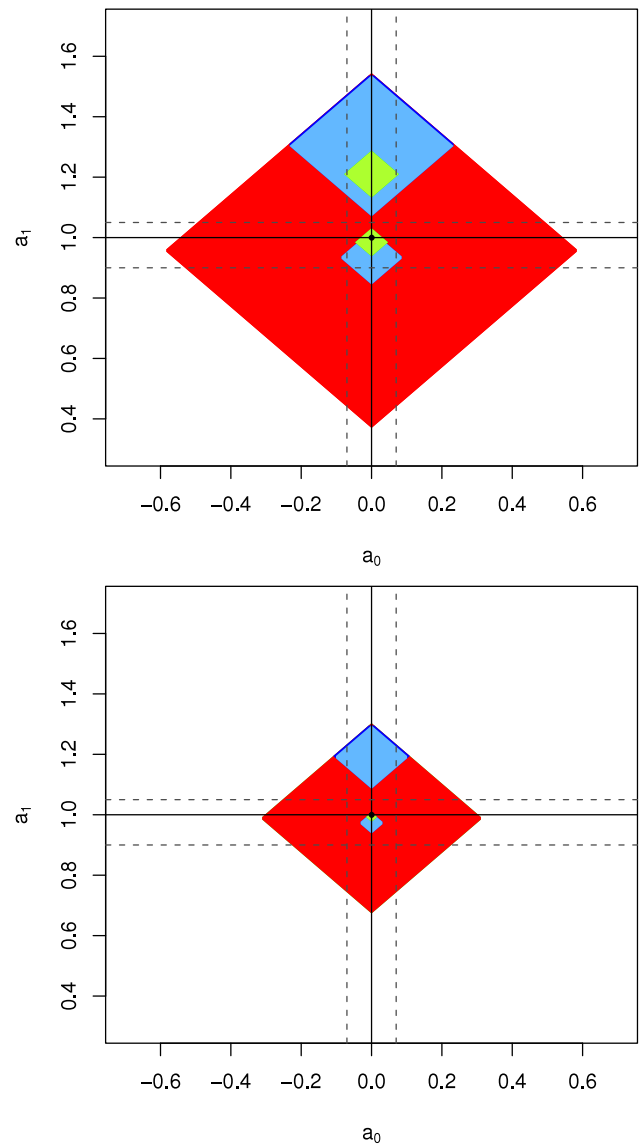


Fig. 2. Illustration A2. Outer sets and identified sets delivered by a binary endogenous variable SEIV model with a parametric ordered probit structural function, intercept a_0 and slope a_1 . The upper (lower) pane represents identified sets when the support of the instrumental variable, Z , is Z_1 (Z_2). Number of categories of the outcome, M : 2 (red), 4 (blue) and 6 (green). The dark blue strip at the upper margin is not in the identified sets. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

4.4. Illustration A2

In this illustration the class of structures generating probability distributions is as in Illustration A1 and, as there, $\alpha_1 = 1$. Also as in Illustration A1 two sets of instrumental values are employed, Z_1 and Z_2 .

In this illustration the number of classes is varied through the following sequence.

$$M \in \{2, 4, 6, 8, 10, 12, 14, 16, 18, 25, 50, 75\}.$$

Threshold values (c_i) are chosen to do a good job of covering the main probability mass of the distribution of Y marginal with respect to X and Z . They are chosen as quantiles of an $N(0, (2.4)^2)$ distribution associated with equally spaced probabilities in $[0, 1]$, e.g. $\{1/2\}$ for $M = 2$, $\{1/3, 2/3\}$ for $M = 3$, and so on. The identified sets are drawn in Figs. 2–5. In each case the upper (lower) pane gives identified sets obtained using the instrumental values Z_1 (Z_2).

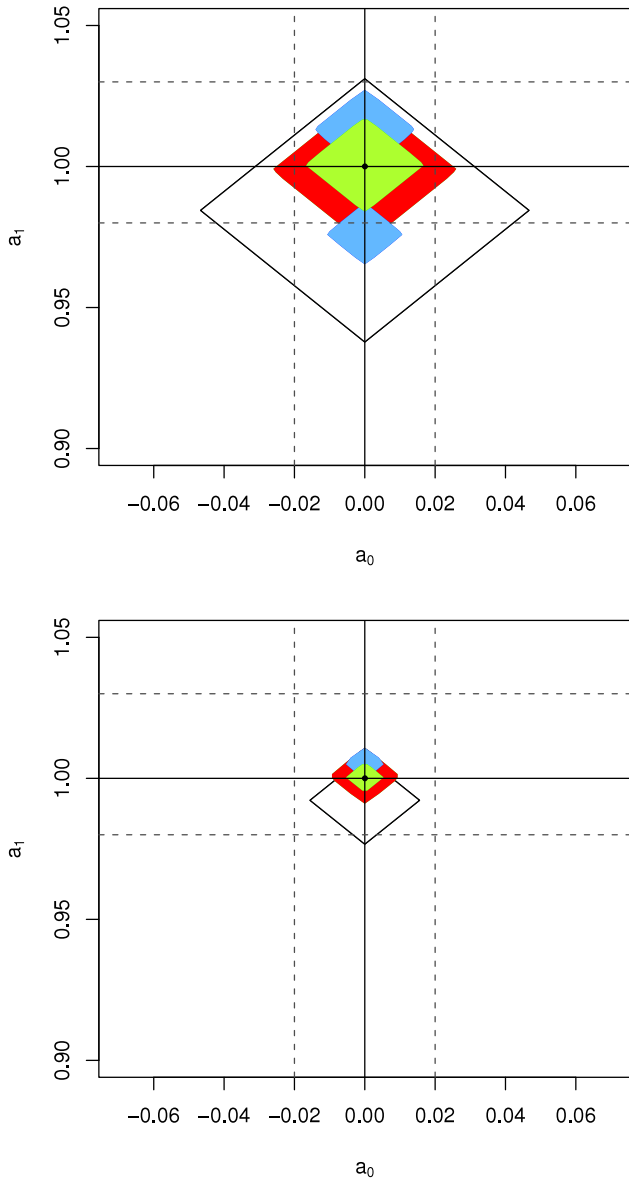


Fig. 3. Illustration A2. Identified sets delivered by a binary endogenous variable SEIV model with a parametric ordered probit structural function, intercept a_0 and slope a_1 . The upper (lower) pane represents identified sets when the support of the instrumental variable, Z , is Z_1 (Z_2). Number of categories of the outcome, M : 8 (red), 10 (blue) and 12 (green). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

First, consider the *upper* panes in each of Figs. 2–5 which are produced using the more restricted set of instrumental values Z_1 . Fig. 2 shows identified sets for $M = 2$ (red), $M = 4$ (blue) and $M = 6$ (green). Notice that in the latter two cases the identified sets are disconnected comprising two rhombuses. On the upper edges of the upper rhombus in the case $M = 4$ is a narrow dark blue strip marking the intersection $e^0(Z) \cap \mathcal{D}^0(Z)$ which does not lie in the identified set. This intersection is empty in the other cases that are shown in this figure and in Figs. 3–5.

Fig. 3 shows identified sets for $M = 8$ (red), $M = 10$ (blue) and $M = 12$ (green). The identified set for $M = 10$ is disconnected. Notice that the scale is greatly expanded in this figure, the identified sets are rapidly decreasing in size as the number of classes observed for Y increases. The outline unshaded rhombus in Fig. 3 is the identified set for $M = 6$ copied across from Fig. 2. Boxes formed by the dashed lines in Fig. 2 show the region focussed on in Fig. 3.

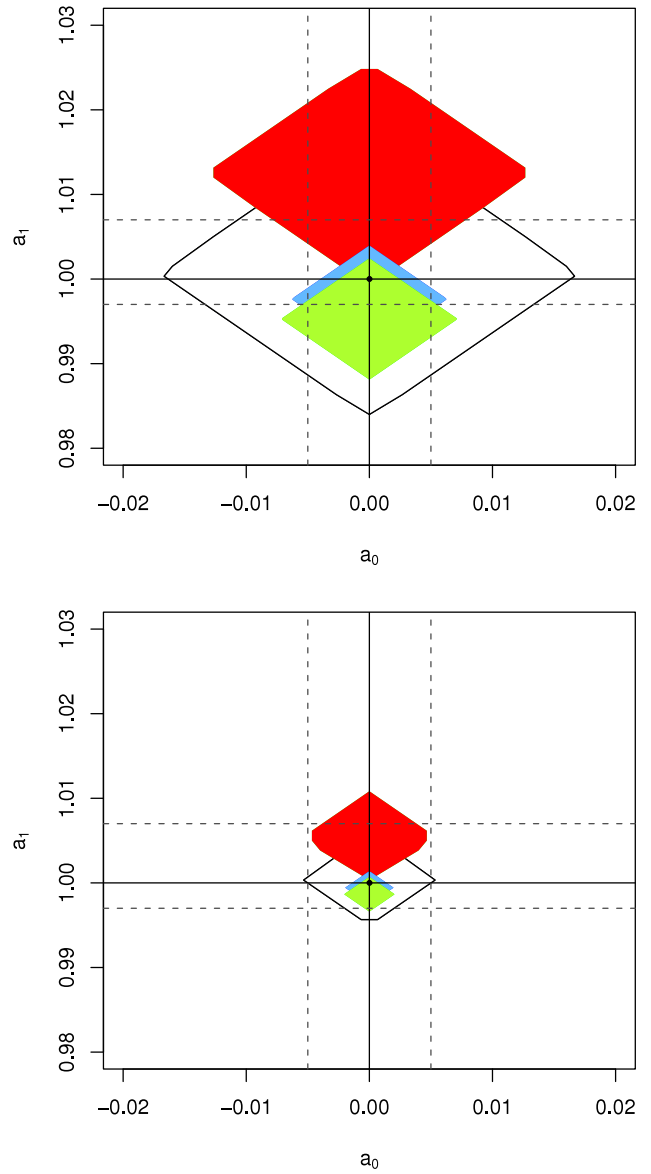


Fig. 4. Illustration A2. Identified sets delivered by a binary endogenous variable SEIV model with a parametric ordered probit structural function, intercept a_0 and slope a_1 . The upper (lower) pane represents identified sets when the support of the instrumental variable, Z , is Z_1 (Z_2). Number of categories of the outcome, M : 14 (red), 16 (blue) and 18 (green). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Fig. 4 shows identified sets for $M = 14$ (red), $M = 16$ (blue) and $M = 18$ (green). Again the scale is greatly expanded relative to the previous figure. The outline unshaded rhombus is the identified set for $M = 12$ copied across from Fig. 3.

Fig. 5 shows identified sets for $M = 25$ (red), $M = 50$ (blue) and $M = 75$ (green). Yet again the scale is greatly expanded relative to the previous figure. The lower part of the identified set for $M = 18$ is drawn in outline. All the identified sets are connected and of very small extent. The situation is now very close to point identification. The identified set at $M = 100$ is not distinguishable from a point at the chosen scale.

Now consider the *lower* panes in Figs. 2–5 which are obtained using the set of instrumental values Z_2 . The scales of the graphs in the upper and lower panes are identical. In each case the doubled range of support results in substantially smaller identified sets. These identified sets have many features in common with the sets obtained with more limited instrument support. They are

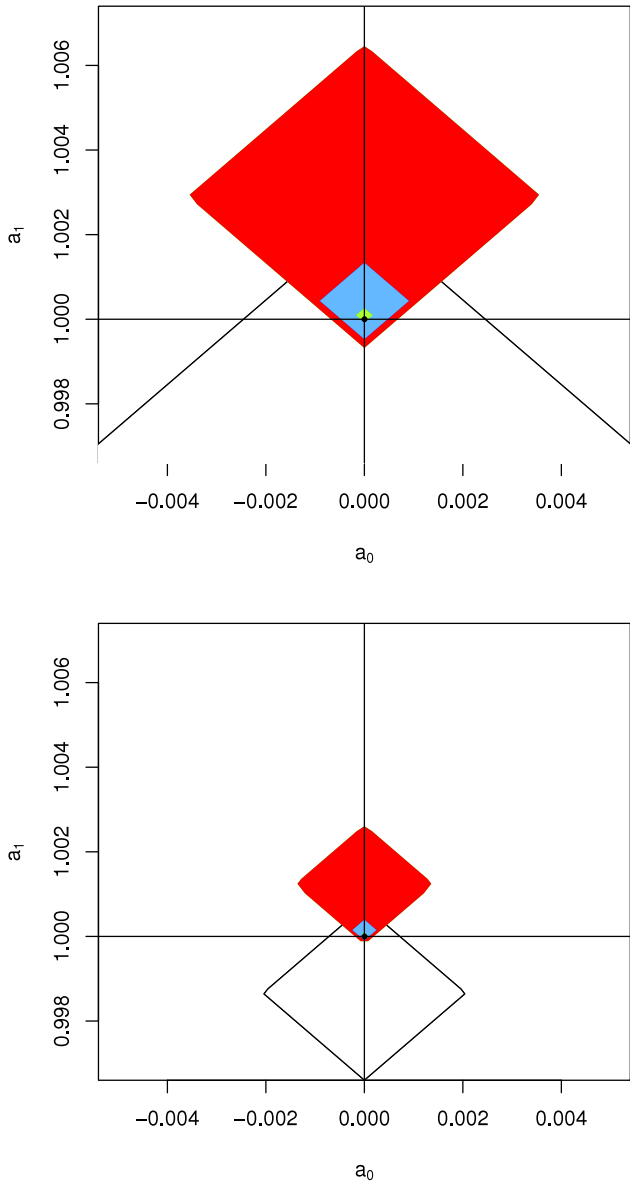


Fig. 5. Illustration A2. Identified sets delivered by a binary endogenous variable SEIV model with a parametric ordered probit structural function, intercept a_0 and slope a_1 . The upper (lower) pane represents identified sets when the support of the instrumental variable, Z , is Z_1 (Z_2). Number of categories of the outcome, M : 25 (red), 50 (blue) and 75 (green). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

disconnected in some cases but when the instrument has greater support this phenomenon disappears sooner as the discreteness of the outcome is reduced.

In Chesher (2010) it is shown that the values of threshold functions $h_m(x^*)$ are point identified at any value x^* for which there exists an instrumental value z^* such that $\Pr_0[X = x^*|Z = z^*] = 1$. In the light of this result these probabilities are a good indicator of the “strength” of an instrument. Table 4 shows the values of these probabilities for the structures that generate the probability distributions used in this illustration and for the two sets of instrumental values Z_1 and Z_2 . The largest probability is 0.84 using Z_2 compared with 0.69 using Z_1 .

In the structures used to generate the probability distributions used in this illustration

$$\Pr_0[X = +1|Z = +\infty] = 1 = \Pr_0[X = -1|Z = -\infty].$$

Table 4
Predictive probabilities for X at values of z in two sets of instrumental values.

z	Z_1		Z_2	
	z	$\Pr[X = 1 Z = z]$	z	$\Pr[X = 1 Z = z]$
-1.00	.31		-2.00	.16
-0.78	.35		-1.56	.22
-0.56	.39		-1.11	.29
-0.33	.43		-0.67	.37
-0.11	.48		-0.22	.46
+0.11	.52		+0.22	.54
+0.33	.57		+0.67	.63
+0.56	.61		+1.11	.71
+0.78	.65		+1.56	.78
+1.00	.69		+2.00	.84

If Z included the values $+\infty$ and $-\infty$ then, given the linear specification in the parametric model for the structural function, there would be point identification of the slope parameter a_1 . But the structure generating the probability distributions employed here is quite special. In econometric practice point identification “at infinity” is unlikely to be achievable.

The upper and lower panes in Fig. 6 plot logarithm (base e) of the lengths of identified intervals for respectively a_0 and a_1 against the logarithm of the number of classes in which Y is observed. Results are shown for the weak and strong sets of instrumental values, respectively Z_1 and Z_2 . The lengths of the identified intervals are always smaller using Z_2 which is the set of instrumental values with wider support.

Fig. 7 plots the logarithm of the area of the identified set for a_0 and a_1 against the logarithm of the number of classes. The areas of the identified sets are always smaller using Z_2 . In each case the points are quite tightly scattered around a negatively sloped linear relationships suggesting approach to point identification at a rate proportional to a power of the number of classes¹². OLS estimates indicate that the lengths of the sets for a_0 and a_1 both fall at a rate proportional to $M^{-2.1}$, and that the area of the identified set for a_0 and a_1 falls at a rate proportional to $M^{-3.6}$.

The details of this approach towards point identification and the geometry of the identified sets depend on fine details of the specification of the structures generating the probability distributions such as the precise spacing of the thresholds.

4.5. Illustration B1

The class of structures generating probability distributions is as in Illustration A1. In this illustration we use the instrumental values in Z_1

In this illustration there are $M = 3$ classes throughout. The values of a_0 and s are normalised to respectively zero and one. The unknown parameters are the thresholds c_1 and c_2 and the slope coefficient a_1 . This is the sort of setup one finds when modelling attitudinal data where threshold values are unknown parameters of considerable interest.

In the structure generating the probability distributions the values of the thresholds are as follows

$$(c_1, c_2) = (-0.667, +0.667)$$

$$\text{and } \alpha_1 = 1.$$

The identified set resides in a three-dimensional square prism of infinite extent: $\mathbb{R} \times (0, 1)^2$. Figs. 8–10 show slices taken through this at a sequence of values of a_1 showing at each chosen value of a_1 the associated identified set for (c_1, c_2) . In all cases this is

¹² Where sets are disconnected the lengths of the identified sets for individual parameters are calculated as the sum of the lengths of disjoint intervals, and the area of the sets for a pair of parameters is calculated as the sum of the areas of the connected component sets.

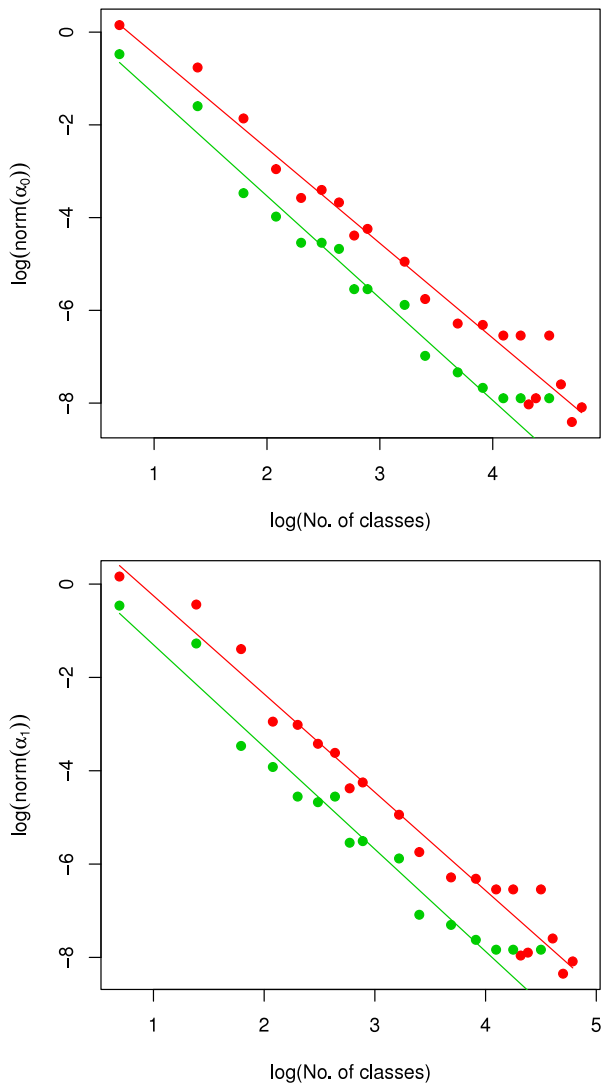


Fig. 6. Illustration A2. Shrinkage of identified sets as the number of outcome categories increases: (upper pane) logarithm of the length of the identified interval for a_0 plotted against the logarithm of the number of categories of the outcome, Y , (lower pane) logarithm of the length of the identified interval for a_1 plotted against the logarithm of the number of categories of the outcome, Y . In green (red) the identified sets are computed with the support of Z equal to Z_1 (Z_2). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

connected and resides above the 45° line in the unit square because the restriction $c_2 > c_1$ has been imposed.

In each case the rectangular regions (shaded red and green) indicate combinations of (c_1, c_2) which at the chosen value of a_1 lie in the set $\mathcal{C}^0(\mathcal{Z})$. The green shaded regions indicate combinations of (c_1, c_2) that at the chosen value of a_1 are in the intersection $\mathcal{C}^0(\mathcal{Z}) \cap \mathcal{D}^0(\mathcal{Z})$. These combinations of (a_1, c_1, c_2) do not lie in the identified set. The red shaded regions indicate combinations of (c_1, c_2) that at the chosen value of a_1 are in the intersection $\hat{\mathcal{C}}^0(\mathcal{Z}) = \mathcal{C}^0(\mathcal{Z}) \cap \mathcal{D}^0(\mathcal{Z})$. These combinations of (a_1, c_1, c_2) are in the identified set.

The extent of the regions in the $c_1 \times c_2$ plane grows as a_1 falls towards the value 1.0 and then shrinks as a_1 falls further.

5. Concluding remarks

Single equation instrumental variable (SEIV) models for ordered discrete outcomes generally set identify structural functions or,

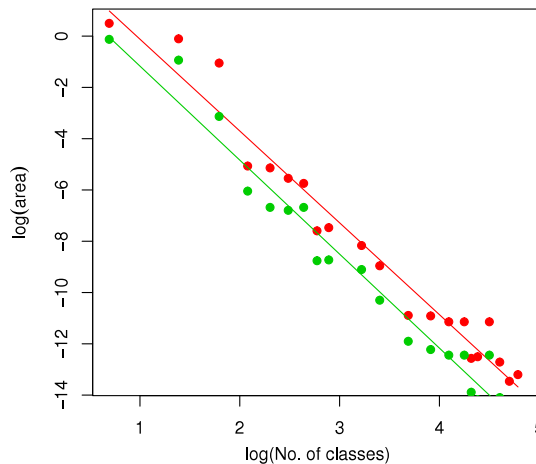


Fig. 7. Illustration A2. Shrinkage of identified sets as the number of outcome categories increases. Logarithm of the area of the identified set plotted against the logarithm of the number of categories of the outcome, Y . In green (red) the identified sets computed with the support of Z equal to Z_1 (Z_2). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

if there are parametric restrictions, parameter values. Complete models, for example the triangular control function model, can be point identifying, but in applied econometric practice there may be no good reason to choose one point identifying model over another.

For any particular distribution of observable variables the sets delivered by the SEIV model give information about the variety of structural functions or parameter values that would be delivered by one or another of the point identifying models which are restricted versions of the SEIV model.

For the nonparametric case we have developed a system of equalities and inequalities that bound the identified sets of structural functions delivered by an SEIV model in the case when endogenous variables are discrete. When either the outcome or the endogenous variable is binary the inequalities sharply define the identified set. The inequalities involve only probabilities about which data is informative and the identified sets can be estimated and inferences drawn using the methods set out in Chernozhukov et al. (2009). Some illustrative calculations for the binary outcome case are given in Chesher (2009).

Calculations in a parametric model suggest that the degree of ambiguity attendant on using the SEIV model reduces rapidly as the discreteness of the outcome is reduced. Research to determine the extent to which this is true in less restricted settings is one of a number of topics of current research.

Acknowledgements

We thank Martin Cripps and Adam Rosen for helpful discussions. Some of the results in this paper were presented at the conference Identification and Decisions held May 8–9, 2009 at Northwestern University. We 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).

Appendix A. Proof of Proposition 5

Consider candidate structural functions, that is, values of γ_{m1} and γ_{m2} , $m \in \{1, \dots, M - 1\}$. Define $\beta(\mathcal{Z})$ so that conditions I1 and I2 are satisfied for all $z \in \mathcal{Z}$. There is only one way to do this: for each m and each $z \in \mathcal{Z}$, to satisfy Condition I1:

$$\beta_{m11}(z) = \bar{\alpha}_{m1}^0(z) \quad \beta_{m22}(z) = \bar{\alpha}_{m2}^0(z) \tag{A1.1}$$

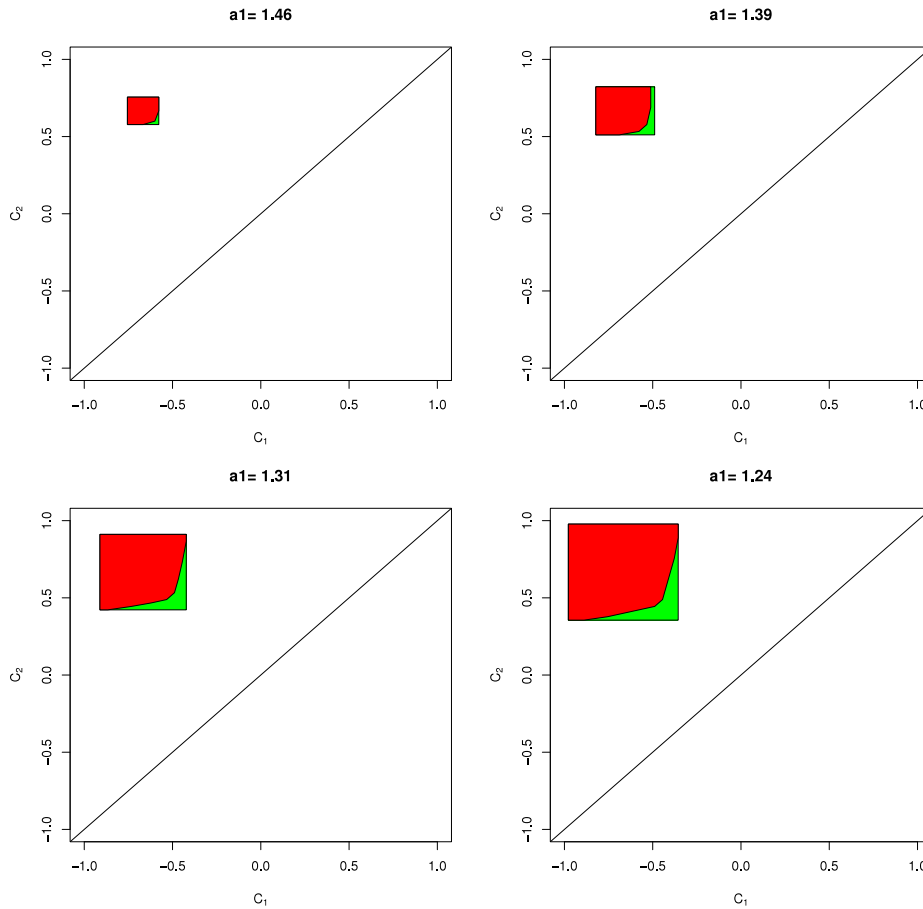


Fig. 8. Illustration B1. Three class ordered probit model with unknown threshold parameters c_1 and c_2 and slope coefficient a_1 . Cross-section of the identified set (red) and outer set (red and green) for c_1 , c_2 and a_1 at selected values of a_1 . (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

and to satisfy Condition I2:

$$\delta_1^0(z)\beta_{m11}(z) + \delta_2^0(z)\beta_{m12}(z) = \gamma_{m1}$$

$$\delta_1^0(z)\beta_{m21}(z) + \delta_2^0(z)\beta_{m22}(z) = \gamma_{m2}$$

and, on combining these results, for $m \in \{1, \dots, M\}$ there are the following expressions

$$\beta_{m12}(z) = \frac{\gamma_{m1} - \delta_1^0(z)\bar{\alpha}_{m1}^0(z)}{\delta_2^0(z)}$$

$$\beta_{m21}(z) = \frac{\gamma_{m2} - \delta_2^0(z)\bar{\alpha}_{m2}^0(z)}{\delta_1^0(z)} \tag{A1.2}$$

It is now shown that for every $\boldsymbol{\gamma} \in \tilde{\mathcal{C}}^0(\mathcal{Z})$ the value of $\beta(\mathcal{Z})$ defined by (A1.1) and (A1.2) as z varies across \mathcal{Z} satisfies the properness condition I3. It follows that $\tilde{\mathcal{C}}^0(\mathcal{Z}) \subseteq \mathcal{H}^0(\mathcal{Z})$ and Proposition 4 states that $\mathcal{H}^0(\mathcal{Z}) \subseteq \tilde{\mathcal{C}}^0(\mathcal{Z})$, so it must be that $\mathcal{H}^0(\mathcal{Z}) = \tilde{\mathcal{C}}^0(\mathcal{Z})$ in this binary endogenous variable case.

To proceed, consider the distribution function characterised by $\beta_{mj1}(z)$ for $m \in \{1, \dots, M-1\}$ and $j \in \{1, 2\}$ and any $z \in \mathcal{Z}$. Here conditioning is on $X = x_1$ and $Z = z$. The argument when conditioning is on $X = x_2$ goes on similar lines and can be worked through by exchange of indices in what follows.

Condition I3 is satisfied if for every adjacent pair of values $\gamma_{s1} < \gamma_{tj}$:

$$\beta_{s11}(z) \leq \beta_{tj1}(z)$$

and there are four possibilities to consider as follows.

- A1. $i = 1, j = 1$. In this case $t = s + 1$ because $\gamma_{s1} < \gamma_{t1}$ are adjacent. Properness requires that $\beta_{s11} \leq \beta_{s+1,11}$ but (A1.1) ensures that this holds because $\beta_{s11} = \bar{\alpha}_{s1}^0(z) \leq \bar{\alpha}_{s+1,1}^0(z) = \beta_{s+1,11}$.
- A2. $i = 1, j = 2$. Properness requires that $\beta_{s11} \leq \beta_{t21}$ which, on using (A1.1) and (A1.2), requires that

$$\bar{\alpha}_{s1}^0(z) \leq \frac{\gamma_{t2} - \delta_2^0(z)\bar{\alpha}_{t2}^0(z)}{\delta_1^0(z)}$$

which is written as follows.

$$\delta_1^0(z)\bar{\alpha}_{s1}^0(z) + \delta_2^0(z)\bar{\alpha}_{t2}^0(z) \leq \gamma_{t2}. \tag{A1.3}$$

If $\boldsymbol{\gamma} \in \mathcal{C}^0(\mathcal{Z})$ then there is, on replacing γ_{ts} by γ_{t2} in the lower bounding inequality in the definition of $\mathcal{C}^0(\mathcal{Z})$ and considering a particular value of $z \in \mathcal{Z}$

$$\sum_{i=1}^K \sum_{m=1}^{M-1} \delta_i^0(z)\alpha_{mi}^0(z)1(\gamma_{mi} \leq \gamma_{t2}) \leq \gamma_{t2} \tag{A1.4}$$

and since $\gamma_{s1} < \gamma_{t2}$ and the values are adjacent the left-hand side of (A1.4) as follows:

$$\delta_1^0(z) \sum_{m=1}^s \alpha_{m1}^0(z) + \delta_2^0(z) \sum_{m=1}^t \alpha_{m2}^0(z)$$

$$= \delta_1^0(z)\bar{\alpha}_{s1}^0(z) + \delta_2^0(z)\bar{\alpha}_{t2}^0(z)$$

and so (A1.3) holds.

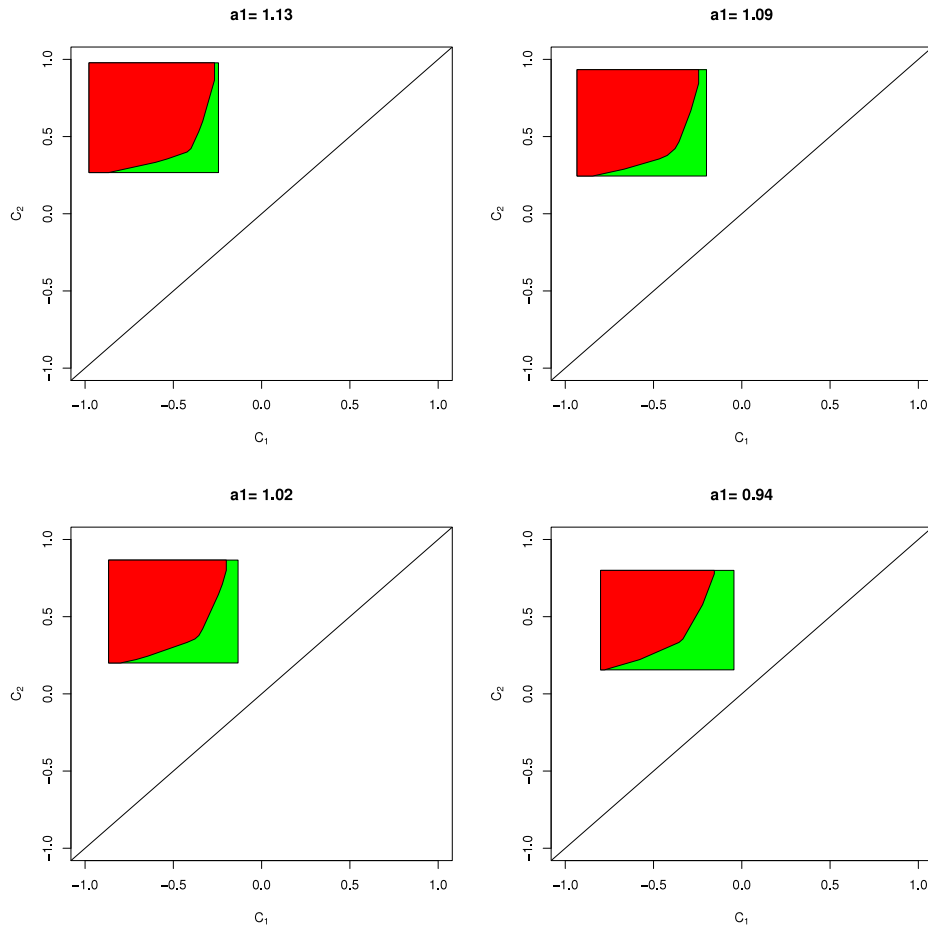


Fig. 9. Illustration B1. Three class ordered probit model with unknown threshold parameters c_1 and c_2 and slope coefficient a_1 . Cross-section of the identified set (red) and outer set (red and green) for c_1, c_2 and a_1 at selected values of a_1 . (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

A3. $i = 2, j = 1$. Properness requires that $\beta_{s21} \leq \beta_{t11}$ which, on using (A1.1) and (A1.2), requires that

$$\frac{\gamma_{s2} - \delta_2^0(z)\bar{\alpha}_{s2}^0(z)}{\delta_1^0(z)} \leq \bar{\alpha}_{t1}^0(z)$$

which is written as follows.

$$\gamma_{s2} \leq \delta_1^0(z)\bar{\alpha}_{t1}^0(z) + \delta_2^0(z)\bar{\alpha}_{s2}^0(z). \tag{A1.5}$$

If $\boldsymbol{\gamma} \in \mathcal{C}^0(\mathcal{Z})$ then there is, on replacing γ_{ts} by γ_{s2} in the upper bounding inequality in the definition of $\mathcal{C}^0(\mathcal{Z})$ and considering a particular value of $z \in \mathcal{Z}$:

$$\gamma_{s2} \leq \sum_{i=1}^K \sum_{m=1}^M \delta_i^0(z)\alpha_{mi}^0(z)1(\gamma_{m-1,i} < \gamma_{s2}) \tag{A1.6}$$

and since $\gamma_{s2} < \gamma_{t1}$ and the values are adjacent the right-hand side of (A1.6) is as follows:

$$\begin{aligned} & \delta_1^0(z) \sum_{m=1}^t \alpha_{m1}^0(z) + \delta_2^0(z) \sum_{m=1}^2 \alpha_{m2}^0(z) \\ &= \delta_1^0(z)\bar{\alpha}_{t1}^0(z) + \delta_2^0(z)\bar{\alpha}_{s2}^0(z) \end{aligned}$$

and so (A1.5) holds.

A4. $i = 2, j = 2$. It must be that $t = s + 1$ because $\gamma_{s2} < \gamma_{t2}$ are adjacent. Properness requires that $\beta_{s21} \leq \beta_{s+1,21}$ which, on using (A1.2), requires that

$$\frac{\gamma_{s2} - \delta_2^0(z)\bar{\alpha}_{s2}^0(z)}{\delta_1^0(z)} \leq \frac{\gamma_{s+1,2} - \delta_2^0(z)\bar{\alpha}_{s+1,2}^0(z)}{\delta_1^0(z)}$$

which is written as follows.

$$\delta_2^0(z)\alpha_{s+1,2}^0(z) \leq \gamma_{s+1,2} - \gamma_{s2}. \tag{A1.7}$$

If $\boldsymbol{\gamma} \in \mathcal{D}^0(\mathcal{Z})$ then replacing γ_{mi} and γ_{ni} by respectively $\gamma_{s+1,2}$ and γ_{s2} in the definition of $\mathcal{D}^0(\mathcal{Z})$ and considering a particular value $z \in \mathcal{Z}$:

$$\gamma_{s+1,2} - \gamma_{s2} \geq \delta_2^0(z)(\bar{\alpha}_{s+1,2}^0(z) - \bar{\alpha}_{s2}^0(z)) = \delta_2^0(z)\alpha_{s+1,2}^0(z)$$

and so (A1.7) holds.

We can think of the set $\tilde{\mathcal{C}}^0(\mathcal{Z})$ as an intersection of sets, one obtained at each value of $z \in \mathcal{Z}$, thus.

$$\tilde{\mathcal{C}}^0(\mathcal{Z}) = \bigcap_{z \in \mathcal{Z}} \tilde{\mathcal{C}}^0(z).$$

It has been shown that for any $z \in \mathcal{Z}$ and for all $\boldsymbol{\gamma} \in \tilde{\mathcal{C}}^0(z)$ there are conditional distribution functions characterised by $\boldsymbol{\beta}(z)$ defined as in (A1.1) and (A1.2) such that conditions I1, I2 and I3 hold.

Let $\boldsymbol{\beta}(\mathcal{Z})$ be the conditional distribution functions generated using definitions (A1.1) and (A1.2) as z varies within \mathcal{Z} . Values $\boldsymbol{\gamma} \in \tilde{\mathcal{C}}^0(\mathcal{Z})$ lie in every set $\tilde{\mathcal{C}}^0(z)$, and so for each such value of $\boldsymbol{\gamma}$ there are conditional distribution functions in $\boldsymbol{\beta}(\mathcal{Z})$ such that conditions I1, I2 and I3 are satisfied. It follows that $\tilde{\mathcal{C}}^0(\mathcal{Z}) \subseteq \mathcal{H}^0(\mathcal{Z})$ and since $\mathcal{H}^0(\mathcal{Z}) \subseteq \tilde{\mathcal{C}}^0(\mathcal{Z})$, it follows that $\mathcal{H}^0(\mathcal{Z}) = \tilde{\mathcal{C}}^0(\mathcal{Z})$. This completes the proof of Proposition 5. \square

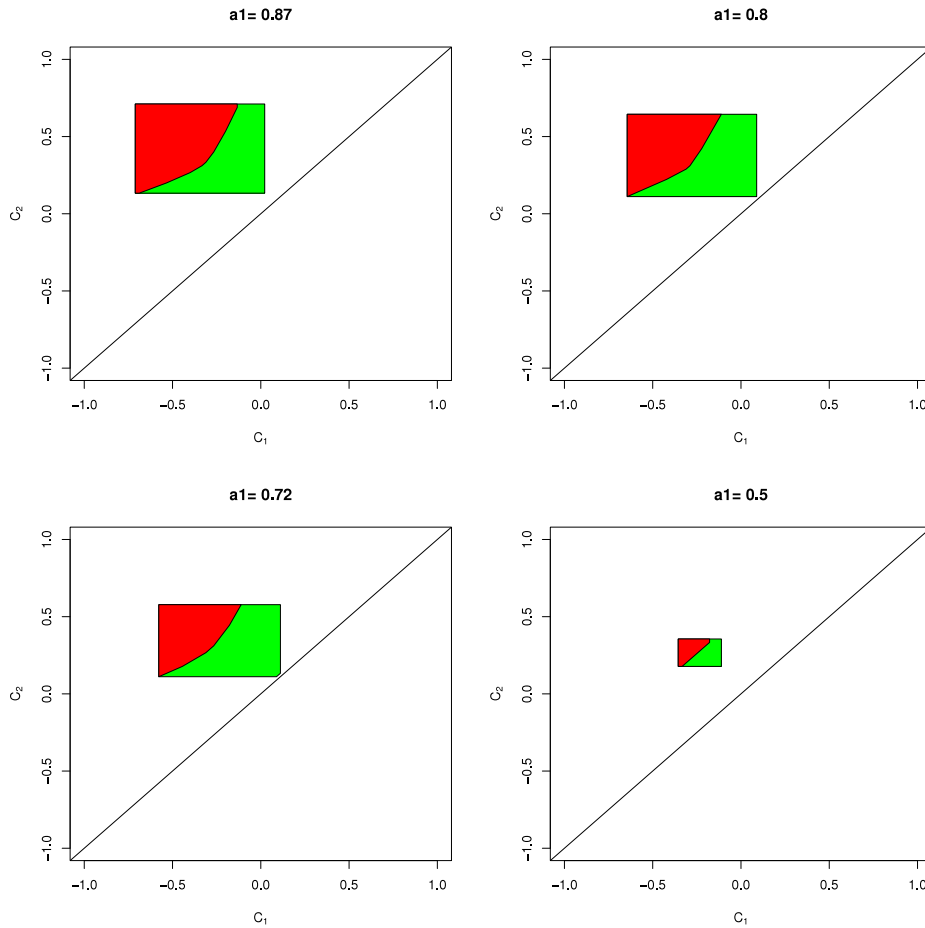


Fig. 10. Illustration B1. Three class ordered probit model with unknown threshold parameters c_1 and c_2 and slope coefficient a_1 . Cross-section of the identified set (red) and outer set (red and green) for c_1 , c_2 and a_1 at selected values of a_1 . (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Appendix B. Inequalities defining the identified set for the case $M = 3, K = 2$

When $M = 3$ and $K = 2$ there are six arrangements of the set of values of γ , of which three (B1–B3) can be obtained by the exchange of indexes labelling elements of support of X in arrangements A1–A3. Dependence of probabilities on z is suppressed in notation. The inequalities excluding (13) deliver set $\mathcal{C}^0(z)$. Intersection with $\mathcal{D}^0(z)$ introduces the additional inequality (13).

- (A1) : $\gamma_{11} < \gamma_{12} < \gamma_{21} < \gamma_{22}$
- (A2) : $\gamma_{11} < \gamma_{12} < \gamma_{22} < \gamma_{21}$
- (A3) : $\gamma_{11} < \gamma_{21} < \gamma_{12} < \gamma_{22}$
- (B1) : $\gamma_{12} < \gamma_{11} < \gamma_{21} < \gamma_{22}$
- (B2) : $\gamma_{12} < \gamma_{11} < \gamma_{22} < \gamma_{21}$
- (B3) : $\gamma_{12} < \gamma_{22} < \gamma_{11} < \gamma_{21}$

(A1)

$$\begin{aligned} \delta_1^0 \alpha_{11}^0 &< \gamma_{11} \leq \delta_1^0 \alpha_{11}^0 + \delta_2^0 \alpha_{12}^0 \\ \delta_1^0 \alpha_{11}^0 + \delta_2^0 \alpha_{12}^0 &< \gamma_{12} \leq \delta_1^0 (\alpha_{11}^0 + \alpha_{21}^0) + \delta_2^0 \alpha_{12}^0 \\ \delta_1^0 (\alpha_{11}^0 + \alpha_{21}^0) + \delta_2^0 \alpha_{12}^0 &< \gamma_{21} \leq \delta_1^0 (\alpha_{11}^0 + \alpha_{21}^0) \\ &\quad + \delta_2^0 (\alpha_{12}^0 + \alpha_{22}^0) \\ \delta_1^0 (\alpha_{11}^0 + \alpha_{21}^0) + \delta_2^0 (\alpha_{12}^0 + \alpha_{22}^0) &< \gamma_{22} \leq \delta_1^0 + \delta_2^0 (\alpha_{12}^0 + \alpha_{22}^0) \end{aligned}$$

(A2)

$$\begin{aligned} \delta_1^0 \alpha_{11}^0 &< \gamma_{11} \leq \delta_1^0 \alpha_{11}^0 + \delta_2^0 \alpha_{12}^0 \\ \delta_1^0 \alpha_{11}^0 + \delta_2^0 \alpha_{12}^0 &< \gamma_{12} \leq \delta_1^0 (\alpha_{11}^0 + \alpha_{21}^0) + \delta_2^0 \alpha_{12}^0 \\ \delta_1^0 \alpha_{11}^0 + \delta_2^0 (\alpha_{12}^0 + \alpha_{22}^0) &< \gamma_{22} \leq \delta_1^0 (\alpha_{11}^0 + \alpha_{21}^0) \\ &\quad + \delta_2^0 (\alpha_{12}^0 + \alpha_{22}^0) \\ \delta_1^0 (\alpha_{11}^0 + \alpha_{21}^0) + \delta_2^0 (\alpha_{12}^0 + \alpha_{22}^0) &< \gamma_{21} \\ &\leq \delta_1^0 (\alpha_{11}^0 + \alpha_{21}^0) + \delta_2^0 \\ \delta_2^0 \alpha_{22}^0 &< \gamma_{22} - \gamma_{12} \end{aligned} \tag{13}$$

(A3)

$$\begin{aligned} \delta_1^0 \alpha_{11}^0 &< \gamma_{11} \leq \delta_1^0 \alpha_{11}^0 + \delta_2^0 \alpha_{12}^0 \\ \delta_1^0 (\alpha_{11}^0 + \alpha_{21}^0) &< \gamma_{21} \leq \delta_1^0 (\alpha_{11}^0 + \alpha_{21}^0) + \delta_2^0 \alpha_{12}^0 \\ \delta_1^0 (\alpha_{11}^0 + \alpha_{21}^0) + \delta_2^0 \alpha_{12}^0 &< \gamma_{12} \leq \delta_1^0 + \delta_2^0 \alpha_{12}^0 \\ \delta_1^0 (\alpha_{11}^0 + \alpha_{21}^0) + \delta_2^0 (\alpha_{11}^0 + \alpha_{21}^0) &< \gamma_{22} \leq \delta_1^0 + \delta_2^0 (\alpha_{11}^0 + \alpha_{21}^0) \end{aligned}$$

Appendix C. Falsifiability of the SEIV model

This Annex provides a simple example of probability distributions for Y and X given Z which cannot be supported by an SEIV model. Faced with these probability distributions the SEIV model delivers a system of inequalities that define an empty set. The existence of such cases ensures that the SEIV model is falsifiable. In the

example, Y , X and Z are all binary. It is easy to extend to more complex cases and there are many choices of values of the probabilities that deliver an empty identified set.

When Y is binary the structural function has a single threshold function taking values γ_{11} and γ_{12} at the two values of X . There are two orderings: $l = 1: \gamma_{11} \leq \gamma_{12}$ and $l = 2: \gamma_{12} \leq \gamma_{11}$. The identified sets associated with z_i , $i \in \{1, 2\}$ are, for each ordering:

$$\mathcal{H}_1(z_i) = \{\boldsymbol{\gamma} : \delta_1(z_i)\alpha_1(z_i) \leq \gamma_{11} \leq \delta_1(z_i)\alpha_1(z_i) + \delta_2(z_i)\alpha_2(z_i) \leq \gamma_{12} \leq \delta_1(z_i) + \delta_2(z_i)\alpha_2(z_i)\}$$

$$\mathcal{H}_2(z_i) = \{\boldsymbol{\gamma} : \delta_2(z_i)\alpha_2(z_i) \leq \gamma_{12} \leq \delta_1(z_i)\alpha_1(z_i) + \delta_2(z_i)\alpha_2(z_i) \leq \gamma_{11} \leq \delta_2(z_i) + \delta_1(z_i)\alpha_1(z_i)\}$$

and the identified set is as follows.

$$\mathcal{H}(Z) = (\mathcal{H}_1(z_1) \cap \mathcal{H}_1(z_2)) \cup (\mathcal{H}_2(z_1) \cap \mathcal{H}_2(z_2))$$

Here is an example with proper probabilities such that this set is empty.

$$\delta_1(z_1) = \delta_1(z_2) = \delta_2(z_1) = \delta_2(z_2) = 0.5 \quad (\text{A2.1})$$

$$\alpha_1(z_1) = \alpha_2(z_1) = 0.2 \quad \alpha_1(z_2) = \alpha_2(z_2) = 0.8. \quad (\text{A2.2})$$

The components of the identified sets are:

$$\mathcal{H}_1(z_1) = \{\boldsymbol{\gamma} : 0.1 \leq \gamma_{11} \leq 0.2 \leq \gamma_{12} \leq 0.6\}$$

$$\mathcal{H}_1(z_2) = \{\boldsymbol{\gamma} : 0.4 \leq \gamma_{11} \leq 0.8 \leq \gamma_{12} \leq 0.9\}$$

$$\mathcal{H}_2(z_1) = \{\boldsymbol{\gamma} : 0.1 \leq \gamma_{12} \leq 0.2 \leq \gamma_{11} \leq 0.6\}$$

$$\mathcal{H}_2(z_2) = \{\boldsymbol{\gamma} : 0.4 \leq \gamma_{12} \leq 0.8 \leq \gamma_{11} \leq 0.9\}$$

and the identified set, $\mathcal{H}(Z) = \phi$, the empty set, because

$$\mathcal{H}_1(z_1) \cap \mathcal{H}_1(z_2) = \mathcal{H}_2(z_1) \cap \mathcal{H}_2(z_2) = \phi.$$

There is no nonparametric SEIV model that supports the probability distribution that delivers (A2.1) and (A2.2).

References

- Blundell, R.W., Powell, J.L., 2003. In: Dewatripont, M., Hansen, L.P., Turnovsky, S.J. (Eds.), *Endogeneity in Nonparametric and Semiparametric Regression Models*. In: *Advances in Economics and Econometrics: Theory and Applications*, Eighth World Congress, vol. II. Cambridge University Press, Cambridge.
- Blundell, R.W., Powell, J.L., 2004. Endogeneity in semiparametric binary response models. *Review of Economic Studies* 71, 655–679.
- Chernozhukov, V., Hansen, C., 2005. An IV model of quantile treatment effects. *Econometrica* 73, 245–261.
- Chernozhukov, V., Lee, S., Rosen, A., 2009. Intersection Bounds: Estimation and Inference, CeMMAP Working Paper 19/09.
- 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.
- Chesher, A.D., 2009. Single Equation Endogenous Binary Response Models, CeMMAP Working Paper 23/09.
- Chesher, A.D., 2010. Instrumental variable models for discrete outcomes. *Econometrica* 78, 575–601.
- Frame, J.S., Robinson, G.de B., Thrall, R.M., 1954. The hook graphs of the symmetric groups. *Canadian Journal of Mathematics* 6, 316–324.
- Greene, W., 2007. LIMDEP 9.0 reference guide. Econometric Software, Inc., New York.
- Heckman, J.J., 1978. Dummy endogenous variables in a simultaneous equations system. *Econometrica* 46, 931–959.
- Ihaka, R., Gentleman, R., 1996. R: a language for data analysis and graphics. *Journal of Computational and Graphical Statistics* 5, 299–314.
- Manski, C.F., 2003. *Partial Identification of Probability Distributions*. Springer-Verlag, New York.
- Rivers, Douglas, Vuong, Quang, 1988. Limited information estimators and exogeneity tests for simultaneous probit models. *Journal of Econometrics* 39, 347–366.
- Smith, R.J., Blundell, R.W., 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*. College Station, TX: StataCorp LP.
- Ziegler, G.M., 2007. *Lectures on Polytopes*. Springer Science, New York, (Updated seventh printing of the first edition).