cemmap

centre for microdata methods and practice

IV models of ordered choice

Andrew Chesher
Konrad Smolinski

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP37/09

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

An ESRC Research Centre

IV Models of Ordered Choice

Andrew Chesher and Konrad Smolinski*
CeMMAP & UCL

December 4th 2009

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.

Keywords: Endogeneity, Incomplete models, Instrumental variables, Ordered choice, Ordered Probit, Set Identification, Threshold Crossing Models.

JEL Codes: C10, C14, C50, C51.

1 Introduction

This paper studies single equation instrumental variables 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,

*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).

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), Chesher (2003).[1]

Unfortunately this strategy does not generally work when endogenous variables are discrete.[2] And, as explained in Chesher(2009), the control function approach exploits strong restrictions concerning the process generating the endogenous variables, restrictions which may not be found plausible in many econometric settings. By contrast here we work with a model which is far less restrictive in this regard, imposing conditions *only* on the structural function generating the discrete response.

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 U. The observed vector of explanatory variables, X, and U may not be independently distributed. However the model requires that U be distributed independently of instruments, Z. We call the model a *Single Equation Instrumental Variable* (SEIV) model. The SEIV model places no restrictions at all on the process generating the endogenous variable, X, and in this respect is incomplete.

Thinking about Manski's (2003) "Law of Decreasing Credibility" 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. Dropping the restrictions of the control function model leads to ambiguity.

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

Since the structural functions of a SEIV model are monotone in scalar U there is a threshold crossing representation in which U is normalised marginally uniformly distributed on the unit interval.

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

In the discrete endogenous variable case a nonparametrically specified structural function, h, is characterised by $N = K \times (M - 1)$ parameters,

[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.

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 $F^0_{YX|Z}$, 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 at first sight 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 (2008) 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 (2008), 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.

Finally 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 observed rises 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 such as 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 give 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 and systems of inequalities obeyed by values of these functions that lie in the identified set are set out in Section 3.3 Proofs of propositions are given in an Annex.

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. 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:*

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

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 normalised to have a marginal uniform distribution on $[0,1]$.

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. *U and instrumental variables Z which take values in some set $\mathcal{Z}$ 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. Neither restriction imposes any conditions on the process generating X. Now consider the identifying power of this model.

Let $F^0_{YX|Z}$ 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 $F_{UX|Z} = F_{UX|Z}^0$.

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

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 $\mathcal{S}^0(\mathcal{Z})$ 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.

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

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 $F_{UX|Z}$ that can accommodate a particular structural function may require searching across an infinite dimensional space of functions.

However Chesher (2008) shows that when Y is *binary* the identified set is determined by a system of inequalities in which the distribution function $F_{UX|Z}$ 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 $F_{UX|Z}$ 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

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

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

In the discrete X case an identified set of structural functions is a set of values of γ, comprising a subset of the unit N-cube.

When determining whether a structural function characterised by a value of γ 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.

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

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

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

$$F_{U|XZ}\left(\gamma_{mi}|x_i, z\right) = F^0_{Y|XZ}(m|x_i, z) \tag{2}$$

must hold. The conditional distribution $F^0_{X|Z}$ is identified so (2) is effectively the observational equivalence condition (1).

The independence restriction together with the uniform distribution normalisation of the marginal distribution of U requires that for each m, i and z the following condition holds:

$$E^0_{X|Z=z}[F_{U|XZ}\left(\gamma_{mi}|X, z\right)] \equiv \sum_{j=1}^{K} F_{U|XZ}\left(\gamma_{mi}|x_j, z\right) \mathrm{Pr}_0[X = x_j|Z = z] = \gamma_{mi} \tag{3}$$

so values of $\beta_{mij}(z)$ with $i \neq j$ are also relevant. Here $E^0_{X|Z=z}$ indicates expectation taken with respect to the distribution $F^0_{X|Z}$ 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 $u \in \gamma$ are relevant when determining whether γ is in the identified set. Other values of u are not relevant because they play no role in the fulfillment of the observational equivalence condition (2) or the independence condition (3).

If γ_{mi} and $\gamma_{m'i'}$ are adjacent[3] 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 distributions.[4]

[3]If there is no $\gamma_{st} \in \gamma$ 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:

$$E^0_{X|Z=z}[F_{U|XZ}\left(u|X, z\right)] = u$$

Let $\Pr_0$ denote probabilities calculated using a particular distribution function $F^0_{YX|Z}$. Define conditional probabilities for X given Z:

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

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

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

be conditional probabilities of X given Z

Define conditional probabilities and cumulative probabilities of the outcome:

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

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

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

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

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

Consider a *structure* characterised by

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

Such a structure lies in the set identified by the SEIV model associated with probabilities $\alpha^0(z)$ and $\delta^0(z)$ 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 $m \in \{1, \ldots, M\}$ and $i \in \{1, \ldots, K\}$

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

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

$$\sum_{j=1}^{K} \delta_j^0(z) \beta_{mij}(z) = \gamma_{mi}.$$

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

---

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

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 $L \equiv (K(M-1))!/((M-1)!)^K$ 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 $l \in \{1, \ldots, L\}$ define sets $\mathcal{S}_l^0(z)$ and $\mathcal{H}_l^0(z)$ as follows.

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

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

The set $\mathcal{S}_l^0(z)$ is the set of structures admitted by the SEIV model that have γ in order l and deliver the distribution $F^0_{YX|Z}$ 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 algorithm[6], 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 $\mathcal{H}_l^0(z)$:

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

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.

If there is a value l such that $\mathcal{H}_l(\mathcal{Z})$ contains values of γ in which no pair of elements have a common value and for more than one value of l there are sets $\mathcal{H}_l(\mathcal{Z})$ which are non-empty then the identified set is not connected.

[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).

Table 1: Number of admissible orderings of gamma with (upper) and without (lower) monotonicity wth respect to X

M	Monotonicity with respect to X	K: 2	3	4	5
2	Yes	1	1	1	1
	No	2	6	24	120
3	Yes	2	5	14	42
	No	6	90	$2,520$	$113,400$
4	Yes	5	42	462	6006
	No	20	$1,680$	$369,600$	$168,168,000$
5	Yes	14	462	$24,024$	$1,662,804$
	No	70	$34,650$	$6,306,300$	$305,540,235,000$

This is so because each set $\mathcal{H}_l(\mathcal{Z})$ lies in one of the $N!$ orthoschemes[7] of the unit N-cube and the orthoschemes have intersections only at their faces where there is equality of two or more elements of γ. 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 subsets of the identified set. With M and K as low as 4 the value of L is $369,600$ and as M or K increase the value of L quickly becomes astronomical.

Additional restrictions can bring some simplification. For example suppose the threshold functions are restricted to be monotone in a scalar explanatory variable X, with a common direction of dependence, say all nondecreasing.

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 $\{1, 2, \ldots, (M-1)K\}$ such that the array increases both across columns and across rows. With $K = 2$ this is the Catalan number $\frac{1}{M+1}\binom{2(M-1)}{(M-1)}$ and the restriction of monotonicity with respect to$X$ brings an $(M-1)$-fold reduction in the number of admissible orderings.

Table 1 shows the value of L for values of M and K up to 5 together with the number of admissible orderings once monotonicity with respect to

[7]The orthoschemes of the unit cube are the regions within which points obeying a particular weak ordering of coordinate values lie. For example in a 3-cube within which lie (x, y, z) there are 6 orthoschemes defined by the inequalities $x \leq y \leq z$, $y \leq x \leq z$, etc. See Coxeter (1973).

X is imposed.[8] The monotonicity restriction can bring large reductions in numbers of admissible orderings but when M or K are at all large there remain huge numbers of admissible orderings of γ.

3.3 Characterisation of the identified set

Chesher (2008) shows that all structural functions in the set identified by the SEIV model associated with a conditional distribution function $F^0_{YX|Z}$ and a set of instrumental values $\mathcal{Z}$ satisfy the following inequalities for all $u \in (0,1)$ and $z \in \mathcal{Z}$.

$$\text{Pr}_0[Y < h(X,u)|Z = z] < u \leq \text{Pr}_0[Y \leq h(X,u)|Z = z]$$

In terms of threshold functions these inequalities are as follows.

$$\sum_{m=1}^{M} \text{Pr}_0[(Y = m) \wedge (h_m(x) < u)|Z = z] < u \leq \sum_{m=1}^{M} \text{Pr}_0[(Y = m) \wedge (h_{m-1}(x) < u)|Z = z]$$

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

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

These inequalities have implications for γ as set out in the following Proposition which is proved in the Annex.

Proposition 1. *For any z, if the inequalities (4) hold for all $u \in (0,1)$ then for all $l \in \{1, \dots, M\}$ and $s \in \{1, \dots, K\}$ the following inequalities hold.*

$$\sum_{i=1}^{K} \sum_{m=1}^{M-1} \delta_i^0(z) \alpha_{mi}^0(z) 1(\gamma_{mi} \leq \gamma_{ls}) \leq \gamma_{ls} \leq \sum_{i=1}^{K} \sum_{m=1}^{M} \delta_i^0(z) \alpha_{mi}^0(z) 1(\gamma_{m-1,i} < \gamma_{ls}) \tag{5}$$

For any ordering l of γ let $\mathcal{C}_l^0(z)$ denote the set of values of γ that satisfy the inequalities (5) of Proposition 1. Since these inequalities define an intersection of halfspaces each set $\mathcal{C}_l^0(z)$ is convex or empty, as is its intersection

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

[8] The row and column ascending matrices encountered here are special cases of Young Tableaux. The `NumberOfTableaux` command in the Combinatorica package (Pemmaraju and Skienka, 2003) of Mathematica (Wolfram Research, Inc., 2008) was used to compute those entries in Table 1 in which montonicity with respect to X is imposed.

Define $\mathcal{C}^0(\mathcal{Z})$ as the set of values of γ of any ordering that satisfy the inequalities of Proposition 1 for all $z \in \mathcal{Z}$ when calculations are done using a distribution $F^0_{YX|Z}$. This is the union of the sets $\mathcal{C}^0_l(\mathcal{Z})$:

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

and, like the identified set, $\mathcal{H}^0(\mathcal{Z})$, the set of values γ defined by the inequalities of Proposition 1, $\mathcal{C}^0(\mathcal{Z})$, is a union of convex sets. It may not itself be convex nor need it be connected.

Chesher (2008, 2009) shows that, when Y is binary, $\mathcal{C}^0(\mathcal{Z})$ is precisely the identified set, $\mathcal{H}^0(\mathcal{Z})$. When Y is not binary this may not be so.

This can be seen by considering Proposition 2, proved in the Annex. Proposition 2, which follows directly from conditions I1-I3, places restrictions on values of γ that lie in the identified set. It will be demonstrated in Section 4 that there can be values of γ which satisfy the inequalities of Proposition 1 and fail to satisfy the inequalities of Proposition 2.

Proposition 2. *If γ lies in the identified set associated with probabilities $\bar{\alpha}^0(z)$ and $\delta^0(z)$ for instrumental values, z, varying in $\mathcal{Z}$, then for all $(m, n) \in \{1, \ldots, M\}$ with $n > m$ and all $i \in \{1, \ldots, K\}$ there are the following inequalities, (i) for each $z \in \mathcal{Z}$:*

$$\gamma_{ni} - \gamma_{mi} \geq \delta^0_i(z) \left(\bar{\alpha}^0_{ni}(z) - \bar{\alpha}^0_{mi}(z)\right) \tag{6}$$

and (ii):

$$\gamma_{ni} - \gamma_{mi} \geq \max_{z \in \mathcal{Z}} \left(\delta^0_i(z) \left(\bar{\alpha}^0_{ni}(z) - \bar{\alpha}^0_{mi}(z)\right)\right). \tag{7}$$

Let $\mathcal{D}^0(\mathcal{Z})$ denote the set of values of γ that satisfy the system of inequalities (7) of Proposition 2. Since $\mathcal{D}^0(\mathcal{Z})$ is an intersection of halfspaces it is a convex set.

Values of γ that lie in the set identified by the SEIV model obey the inequalities of Proposition 1 and Proposition 2 so the identified set lies in the intersection of the sets defined by the inequalities of the two Propositions as stated in Proposition 3.

Proposition 3. *The identified set, $\mathcal{H}^0(\mathcal{Z})$, is a subset of $\tilde{\mathcal{C}}^0(\mathcal{Z}) \equiv \mathcal{C}^0(\mathcal{Z}) \cap \mathcal{D}^0(\mathcal{Z})$.*

Like $\mathcal{C}^0(\mathcal{Z})$ the set $\tilde{\mathcal{C}}^0(\mathcal{Z})$ is a union of convex sets as can be seen by expressing it as follows.

$$\tilde{\mathcal{C}}^0(\mathcal{Z}) = \bigcup_{l=1}^{L} \left(\mathcal{C}^0_l(\mathcal{Z}) \cap \mathcal{D}^0(\mathcal{Z})\right)$$

When Y is *binary* the inequalities (6) of Proposition 2 reduce to the following.

$$\delta_i^0(z)\alpha_{1i}^0(z) \leq \gamma_{1i} \leq 1 + \delta_i^0(z)(1 - \alpha_{1i}^0(z)) \qquad i \in \{1, \ldots, K\} \tag{8}$$

The inequality (5) of Proposition 1 requires that

$$\sum_{j=1}^{i} \delta_j^0(z)\alpha_{1j}^0(z) \leq \gamma_{1i} \leq 1 + \sum_{j=i}^{K} \delta_j^0(z)(1 - \alpha_{1j}^0(z)) \qquad i \in \{1, \ldots, K\} \tag{9}$$

and it is clear that (8) is satisfied if (9) is satisfied. Therefore when Y is binary $\mathcal{C}^0(\mathcal{Z}) \subseteq \mathcal{D}^0(\mathcal{Z})$ so $\tilde{\mathcal{C}}^0(\mathcal{Z}) \equiv \mathcal{C}^0(\mathcal{Z})$ confirming the result of Chesher (2008) for the binary endogenous variable case: for binary Y, $\mathcal{C}^0(\mathcal{Z})$ *is* the identified set $\mathcal{H}^0(\mathcal{Z})$.

If the *explanatory* variable, X, is binary then $\tilde{\mathcal{C}}^0(\mathcal{Z})$ is the identified set, as stated in Proposition 4, which is proved in the Annex.

Proposition 4. *When X is binary $\mathcal{H}^0(\mathcal{Z}) = \tilde{\mathcal{C}}^0(\mathcal{Z})$ no matter how many points of support Y has.*

The inequalities defining $\tilde{\mathcal{C}}^0(\mathcal{Z})$ of Proposition 4 involve probabilities about which data is informative and the value γ that characterises a structural function. The values of the elements of $\beta(\mathcal{Z})$ that define the conditional distribution functions of U given X and Z do not appear in these inequalities. So Proposition 4 points the way to fast computation of the identified set. In Section 4 it provides the basis for computations that illustrate identified sets in a parametrically restricted ordered probit model with a binary endogenous variable and from $M = 2$ to $M = 130$ points of support for the ordered outcome Y.

4 Discreteness and identified sets in a parametric ordered probit model

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.

$$Y = \begin{cases} 1 \quad , & 0 \leq U \leq \Phi(s^{-1}(c_1 - a_0 - a_1 X)) \\ 2 \quad , & \Phi(s^{-1}(c_1 - a_0 - a_1 X)) < U \leq \Phi(s^{-1}(c_2 - a_0 - a_1 X)) \\ \vdots \quad \vdots & \vdots \quad \vdots \quad \vdots \\ M \quad , & \Phi(s^{-1}(c_{M-1} - a_0 - a_1 X)) < U \leq 1 \end{cases} \tag{10}$$

Here Φ denotes the standard normal distribution function, the constants $c_1, \ldots, c_{M-1}$ 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 $\{-1, +1\}$, There is the independence restriction: $U \perp\!\!\!\perp Z$, U is normalised $Unif(0,1)$.

In one portfolio of illustrations (A) the model specifies the values of the threshold parameters $c_1, \ldots, c_{M-1}$ as known, and s as known and normalised to one. This leaves just two unknown 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 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.

In all cases the instrumental variable takes equally spaced values in the interval $[-1, 1]$.

There are a number of reasons for choosing this particular parametric model and set up for this exercise.

1. Many researchers doing applied work will base their analysis on parametric models and the ordered probit model is a leading case considered in practice.

2. When studying the impact of the discreteness of the outcome on identified sets it is convenient to have objects like the parameters a_0 and a_1 which remain stable with a common meaning as the discreteness of the outcome is varied.

3. The number of unknown objects in a fully nonparametric analysis is $N = K(M-1)$ and the identified set can be highly complex comprising the union of an enormous number of sets associated with each possible ordering of the N values delivered by the structural function - see Table 1. The parametric model severely restricts the number of feasible orderings and, as explained below, it is not necessary to search across many possible orderings when determining the extent of the identified set.

4.2 Calculation procedures

The calculation of an identified set of parameter values for a particular distribution $F^0_{YX|Z}$ and set of instrumental values $\mathcal{Z}$ proceeds as follows.

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, say (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.

With a value γ^* to hand the ordering of its elements, say l^*, is determined and the linear equalities and inequalities defining the convex set $\mathcal{H}^0_{l^*}(\mathcal{Z})$ can be calculated. In all the illustrations, because X is binary, $\mathcal{H}^0_{l^*}(\mathcal{Z}) = \tilde{\mathcal{C}}^0_{l^*}(\mathcal{Z})$. If γ^* falls in this set then (a_0^*, a_1^*) is in the identified set, 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.

$$\begin{aligned} Y^* &= \alpha_1 X + W \\ X^* &= 0.5Z + V \end{aligned}$$

$$Y = \begin{cases} 1 &, \quad -\infty \leq Y^* \leq c_1 \\ 2 &, \quad c_1 < Y^* \leq c_2 \\ \vdots & \vdots \qquad \vdots \quad \vdots \quad \vdots \\ M &, \quad c_{M-1} < Y^* \leq +\infty \end{cases} \qquad X = \begin{cases} -1 &, \quad -\infty \leq X^* \leq 0 \\ +1 &, \quad 0 < X^* \leq +\infty \end{cases}$$

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.

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

There are 10 values in $\mathcal{Z}$ as follows.

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

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 2.

Table 2: Illustration A1: Threshold values

Number of Classes: M	Threshold Values (c_i)	Shading in Figure 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

Identified sets for the two parameters, (a_0, a_1), are drawn in Figure 1. The sets are rhombuses arranged with edges parallel to 45° and 225° lines. Identified sets are superimposed one upon another.

The largest rhombus drawn in Figure 1 is the identified set with $M = 2$. Because the outcome is binary this is the set $\mathcal{C}^0(\mathcal{Z})$.

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

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 green 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 Figure 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.

4.4 Illustration A2

In this illustration the class of structures generating probability distributions is as in Illustration A1 and, as there, $\alpha_1 = 1$. But there are now 5 values in Z as follows

$$\mathcal{Z} = \{\pm 1.0, \pm 0.5, 0.0\}$$

and 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 are chosen to "cover" the main probability mass of the distribution of Y marginal with respect to X and Z. They are chosen as quantiles of a $N(0, (2.4)^2)$ distribution associated with equally spaced probabilities in $[0, 1]$, e.g. $\{1/2\}$ for $M = 2$, $\{1/3, 2/3\}$ for $M = 3$. The identified sets are drawn in Figure 2-5.

Figure 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 $\mathcal{C}^0(\mathcal{Z}) \cap \overline{\mathcal{D}^0(Z)}$ which does not lie in the identified set. This intersection is *empty* in the other cases shown in this Figure and in Figures 3 - 5.

Figure 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 Figure 3 is the identified set for $M = 6$ copied across from Figure 2. Boxes formed by the dashed lines in Figure 2 show the region focussed on in Figure 3.

Figure 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 Figure 3.

Figure 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.

The two panes of Figure 6 plot logarithm (base e) of the lengths of identified intervals for a_0 and a_1 against the logarithm of the number of classes in which Y is observed. Figure 7 plots the logarithm of he area of the identified set for a_0 and a_1 against the logarithm of the number of classes. 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 classes[9]. OLS estimates indicate that the lengths of the sets for a_0 and a_1 both fall at a rate proportional to

[9] Where sets are disconnected the lengths of the identified sets for individual parameters are the 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.

$M^{-2.1}$ and that the area of the identified set for a_0 and a_1 falls at a rate proportional to $M^{-3.6}$.

The fine details of this approach and the geometry of the identified sets depends 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 and, as in that illustration there are 10 values in $\mathcal{Z}$, as follows.

$$\mathcal{Z} = \{\pm 1, \pm 0.777, \pm 0.555, \pm 0.333, \pm 0.111\}$$

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 set up 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

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

and $\alpha_1 = 1$.

The identified set resides in a 3-dimensional square prism of infinite extent: $\mathbb{R} \times (0,1)^2$. Figures 8, 9 and 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 connected and resides in the upper orthoscheme of 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 $\mathcal{C}^0(\mathcal{Z}) \cap \overline{\mathcal{D}^0(Z)}$. 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 $\tilde{\mathcal{C}}^0(\mathcal{Z}) = \mathcal{C}^0(\mathcal{Z}) \cap \mathcal{D}^0(\mathcal{Z})$. 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 models for ordered discrete outcomes generally set identify structural functions or, 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 a SEIV model in the case when endogenous variables are discrete. We have shown that 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, Lee and Rosen (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.

References

Blundell Richard W., and James L. Powell (2003): "Endogeneity in Nonparametric and Semiparametric Regression Models," in Dewatripont, M., L.P. Hansen, and S.J. Turnovsky, eds., *Advances in Economics and Econometrics: Theory and Applications, Eighth World Congress, Vol. II.* Cambridge: Cambridge University Press.

Blundell, Richard W., and James .L. Powell (2004): "Endogeneity in Semiparametric Binary Response Models," *Review of Economic Studies*, 71, 655-679.

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

Chernozhukov, Victor, Sokbae Lee and Adam Rosen (2009): "Intersection Bounds: Estimation and Inference," CeMMAP Working Paper 19/09.

CHESHER, ANDREW D., (2003): "Identification in Nonseparable Models," *Econometrica*, 71, 1405-1441.
CHESHER, ANDREW D., (2005): "Nonparametric Identification under Discrete Variation," *Econometrica*, 73, 1525-1550.
CHESHER, ANDREW D., (2008): "Instrumental Variable Models for Discrete Outcomes," CeMMAP Working Paper 30/08, forthcoming in *Econometrica.*
CHESHER, ANDREW D., (2009): "Single Equation Endogenous Binary Response Models," CeMMAP Working Paper 23/09.
COXETER, HAROLD S.M., (1973): *Regular Polytopes.* New York: Dover.
GREENE, WILLIAM (2007): *LIMDEP 9.0 Reference Guide.* Econometric Software, Inc., New York.
HECKMAN, JAMES J., (1978): "Dummy Endogenous Variables in a Simultaneous Equations System," *Econometrica*, 46, 931-959.
IHAKA, ROSS, AND ROBERT GENTLEMAN (1996): "R: A Language for Data Analysis and Graphics," *Journal of Computational and Graphical Statistics*, 5, 299-314.
MANSKI, CHARLES F., (2003): *Partial Identification of Probability Distributions.* New York: Springer-Verlag.
PEMMARAJU, SRIRAM AND STEVEN SKIENKA (2003); *Combinatorial Discrete Mathematics.* New York: Cambridge University Press.
RIVERS, DOUGLAS AND QUANG VUONG (1988): "Limited Information Estimators and Exogeneity Tests for Simultaneous Probit Models," *Journal of Econometrics* 39, 347-366.
SMITH, RICHARD J., AND RICHARD 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." College Station, TX: StataCorp LP.
WOLFRAM RESEARCH INC. (2008): *Mathematica Edition: Version 7.0.* Champaign IL: Wolfram Research Inc.
ZIEGLER, GUNTER M., (2007): *Lectures on Polytopes.* (Updated seventh printing of the first edition*).* New York: Springer Science.

Annex: Proofs of Propositions

Proof of Proposition 1. 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 side of (4), reproduced here.

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

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:

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

which is the right hand side of (5).

Now consider some arrangement of the elements of γ 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 (4), reproduced here.

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

This inequality must hold for all u in $(\gamma_{ls}, \gamma_{pr}]$ and so must hold as in (4) 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:

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

which is the left hand side of (5). $\square$

Proof of Proposition 2. Since γ is in the identified set for each $z \in \mathcal{Z}$ there exists a distribution function characterised by $\beta(z)$ satisfying conditions I1-I3. Conditions I1 and I2 imply that:

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

and the result (i) follows on subtracting and noting that the properness condition I3 ensures that for, each i and j, $\beta_{nij}(z) \geq \beta_{mij}(z)$ because $n > m$.

The result (ii) follows directly on intersecting the intervals obtained at each value $z \in \mathcal{Z}$. $\square$

Proof of Proposition 4. Consider candidate structural functions, that is, values of γ_{m1} and γ_{m2}, $m \in \{1, \ldots, M-1\}$. 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, to satisfy Condition I1:

$$\beta_{m11}(z) = \bar{\alpha}^0_{m1}(z) \qquad \beta_{m22}(z) = \bar{\alpha}^0_{m2}(z) \tag{11}$$

and to satisfy Condition I2:

$$\begin{aligned} \delta^0_1(z)\beta_{m11}(z) + \delta^0_2(z)\beta_{m12}(z) &= \gamma_{m1} \\ \delta^0_1(z)\beta_{m21}(z) + \delta^0_2(z)\beta_{m22}(z) &= \gamma_{m2} \end{aligned}$$

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

$$\beta_{m12}(z) = \frac{\gamma_{m1} - \delta^0_1(z)\bar{\alpha}^0_{m1}(z)}{\delta^0_2(z)} \qquad \beta_{m21}(z) = \frac{\gamma_{m2} - \delta^0_2(z)\bar{\alpha}^0_{m2}(z)}{\delta^0_1(z)} \tag{12}$$

It is now shown that for every $\gamma \in \tilde{\mathcal{C}}^0(\mathcal{Z})$ the value of $\beta(\mathcal{Z})$ defined by (11) and (12) as z varies across $\mathcal{Z}$ satisfies the properness condition I3. It follows that $\tilde{C}^0(\mathcal{Z}) \subseteq \mathcal{H}^0(Z)$ and Proposition 3 states that $\mathcal{H}^0(Z) \subseteq \tilde{\mathcal{C}}^0(\mathcal{Z})$, so it must be that $\mathcal{H}^0(Z) = \tilde{\mathcal{C}}^0(\mathcal{Z})$ in this binary endogenous variable case.

To proceed, consider the distribution function characterised by $\beta_{mj1}(z)$ for $m \in \{1, \ldots, M-1\}$ 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 $\gamma_{si} < \gamma_{tj}$:

$$\beta_{si1}(z) \leq \beta_{tj1}(z)$$

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 (11) ensures that this holds because $\beta_{s11} = \bar{\alpha}^0_{s1}(z) \leq \bar{\alpha}^0_{s+1,1}(z) = \beta_{s+1,11}$.

A2 $i = 1$, $j = 2$. Properness requires that $\beta_{s11} \leq \beta_{t21}$ which, on using (11) and (12), requires that:

$$\bar{\alpha}^0_{s1}(z) \leq \frac{\gamma_{t2} - \delta^0_2(z)\bar{\alpha}^0_{t2}(z)}{\delta^0_1(z)}$$

which is written as follows.

$$\delta^0_1(z)\bar{\alpha}^0_{s1}(z) + \delta^0_2(z)\bar{\alpha}^0_{t2}(z) \leq \gamma_{t2} \tag{13}$$

If $\gamma \in \mathcal{C}^0(z)$ then the inequality (5) holds and, on its left hand side, replacing γ_{ls} by γ_{t2} there is:

$$\sum_{i=1}^{K} \sum_{m=1}^{M-1} \delta_i^0(z)\alpha_{mi}^0(z)1(\gamma_{mi} \leq \gamma_{t2}) \leq \gamma_{t2} \tag{14}$$

and since $\gamma_{s1} < \gamma_{t2}$ and the values are adjacent the left hand side of (14) 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 (13) holds.

A3 $i = 2$, $j = 1$. Properness requires that $\beta_{s21} \leq \beta_{t11}$ which, on using (11) and (12), 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.

$$\gamma_{s2} \leq \delta_1^0(z)\bar{\alpha}_{t1}^0(z) + \delta_2^0(z)\bar{\alpha}_{s2}^0(z) \tag{15}$$

If $\gamma \in \mathcal{C}^0(z)$ then the inequality (5) holds and, on its right hand side, replacing γ_{ls} by γ_{s2} there is:

$$\gamma_{s2} \leq \sum_{i=1}^{K} \sum_{m=1}^{M} \delta_i^0(z)\alpha_{mi}^0(z)1(\gamma_{m-1,i} < \gamma_{s2}) \tag{16}$$

and since $\gamma_{s2} < \gamma_{t1}$ and the values are adjacent the right hand side of (16) is as follows:

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

and so (15) 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 (12), 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.

$$\delta_2^0(z)\alpha_{s+1,2}^0(z) \leq \gamma_{s+1,2} - \gamma_{s2} \tag{17}$$

If $\gamma \in \mathcal{D}^0(z)$ then the inequality (6) of Proposition 2 holds and replacing γ_{ni} and γ_{mi} by respectively $\gamma_{s+1,2}$ and γ_{s2} gives the following:

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

and so (17) holds.

It has been shown that for any $z \in \mathcal{Z}$ and for all $\gamma \in \tilde{\mathcal{C}}^0(z) = \mathcal{C}^0(z) \cap \mathcal{D}^0(z)$ there are conditional distribution functions characterised by $\beta(z)$ defined as in (11) and (12) such that conditions I1, I2 and I3 hold.

Let $\beta(\mathcal{Z})$ be the conditional distribution functions generated using the definitions (11) and (12) as z varies within $\mathcal{Z}$. Since $\tilde{\mathcal{C}}^0(\mathcal{Z}) = \bigcap_{z \in \mathcal{Z}} \tilde{\mathcal{C}}^0(z)$, values $\gamma \in \tilde{\mathcal{C}}^0(\mathcal{Z})$ lie in every set $\tilde{\mathcal{C}}^0(z)$ and so for each such value of γ there are conditional distribution functions in $\beta(\mathcal{Z})$ such that conditions I1, I2 and I3 are satisfied. It follows that $\tilde{\mathcal{C}}^0(\mathcal{Z}) \subseteq \mathcal{H}^0(Z)$ and since $\mathcal{H}^0(Z) \subseteq \tilde{\mathcal{C}}^0(\mathcal{Z})$, it follows that $\mathcal{H}^0(Z) = \tilde{\mathcal{C}}^0(\mathcal{Z})$. $\square$

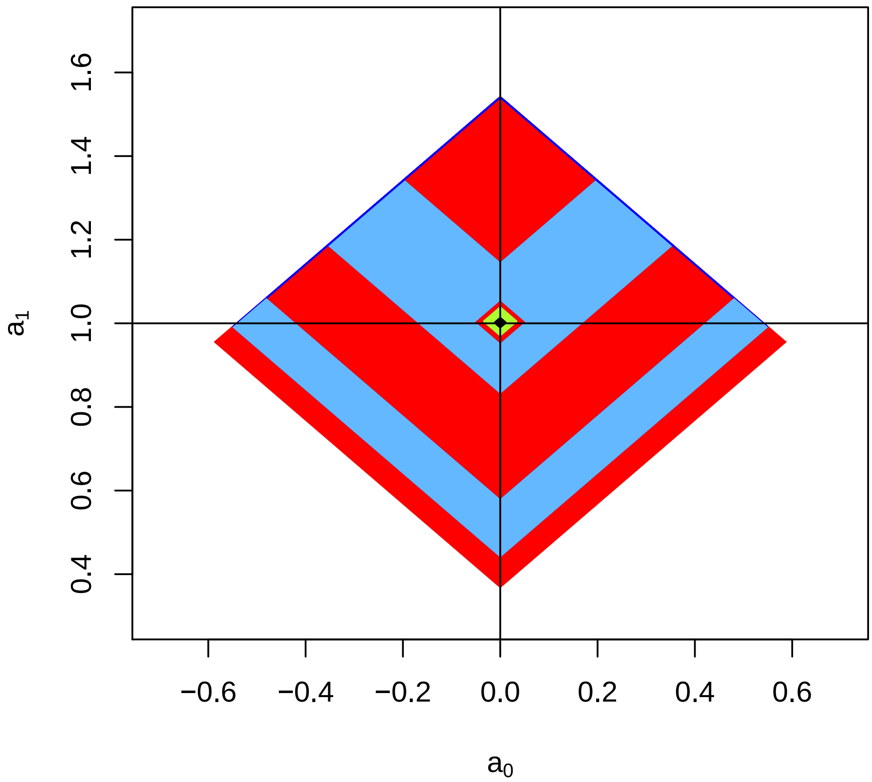

Figure 1: **Illustration A1.** Outer sets and identified sets in a binary endogenous variable SEIV model with a parametric ordered probit structural function with threshold functions of the form $\Phi(c_i - a_0 - a_1 x)$ as the number of categories of the outcome varies from 2 to 10. The dark blue strip at the upper margin of the rhombi is not part of the identified sets.

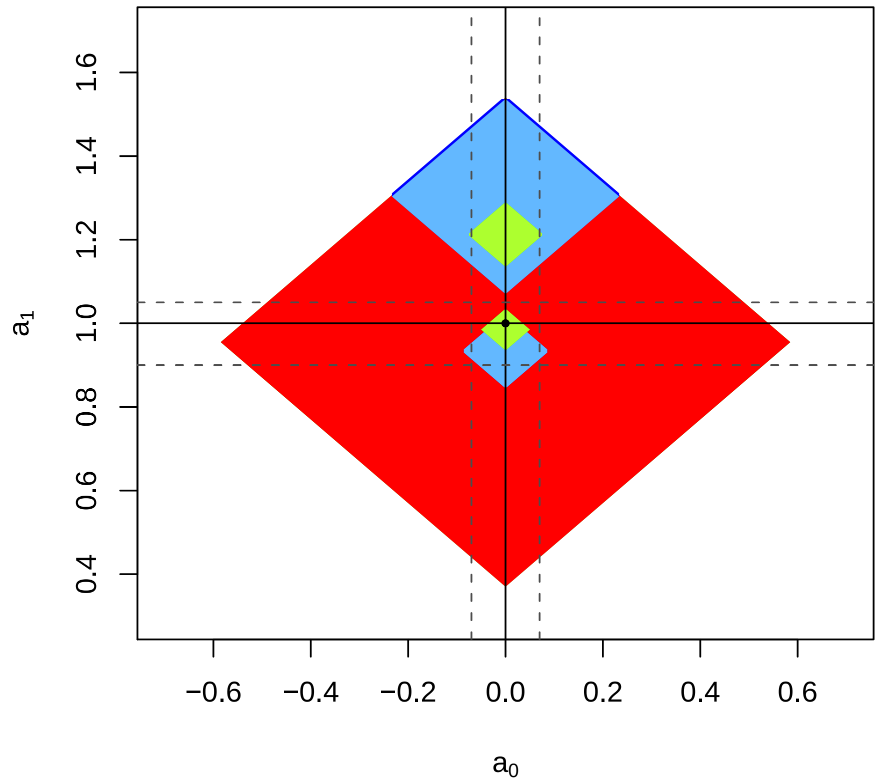

Figure 2: **Illustration A2.** Outer sets and identified sets delivered by a binary endogenous variable SEIV model with a parametric ordered probit structural function, intercept a_0 and slope a_1. Number of categories of the oucome, M: 2(red), 4(blue) and 6(green). The dark blue strip at the upper margin is not in the identified sets.

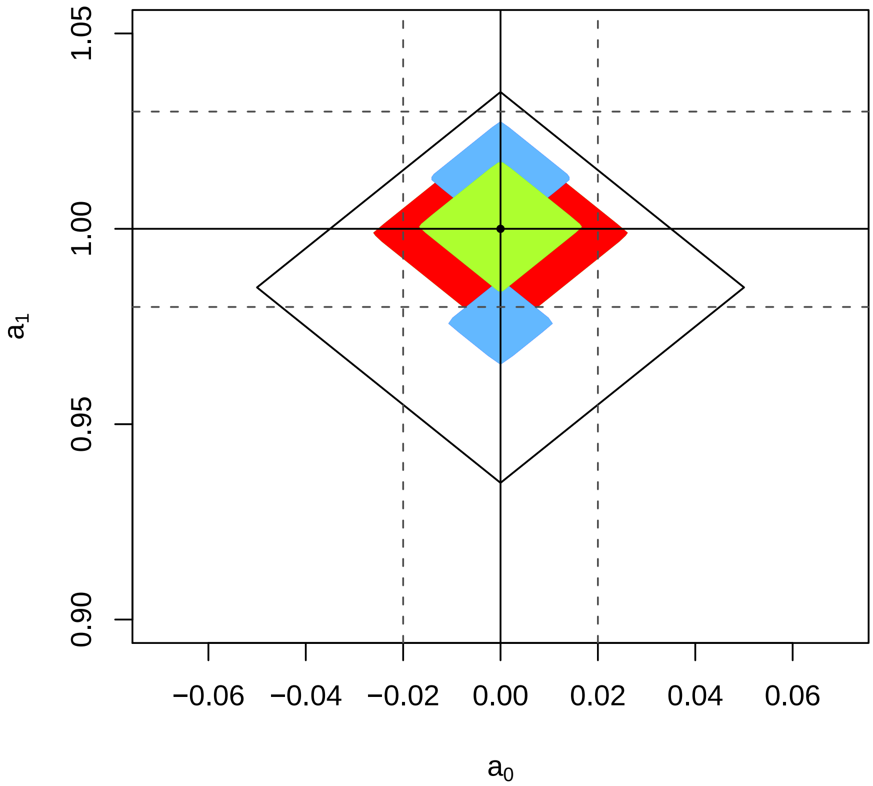

Figure 3: **Illustration A2.** Identified sets delivered by a binary endogenous variable SEIV model with a parametric ordered probit structural function, intercept a_0 and slope a_1. Number of categories of the outcome, M: 8(red), 10(blue) and 12(green).

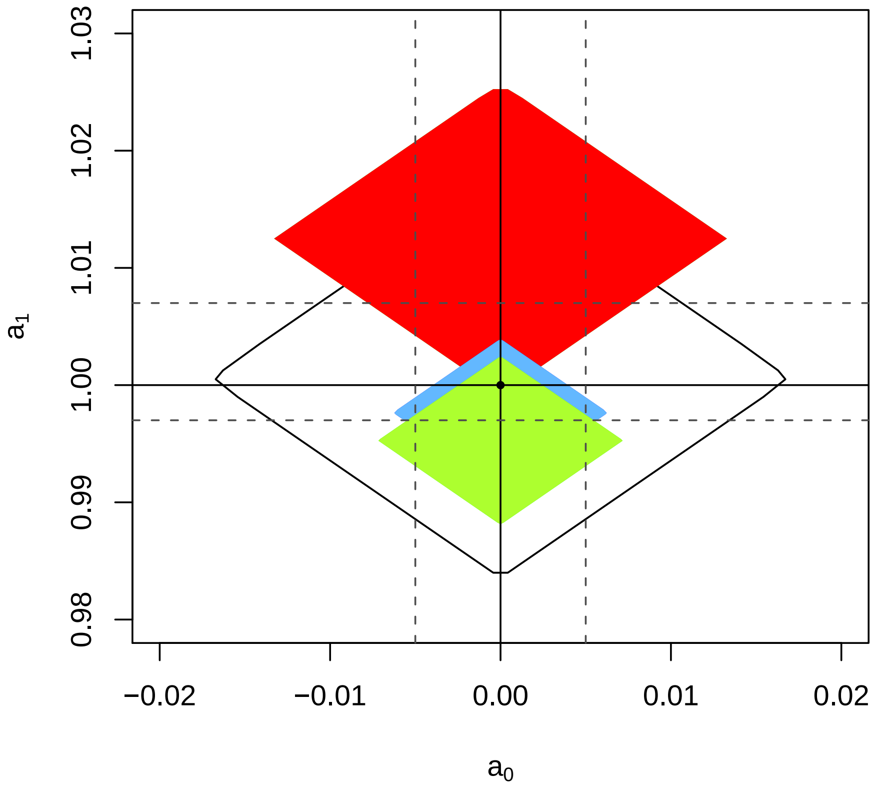

Figure 4: **Illustration A2.** Identified sets delivered by a binary endogenous variable SEIV model with a parametric ordered probit structural function, intercept a_0 and slope a_1. Number of categories of the outcome, M:14(red), 16(blue) and 18(green).

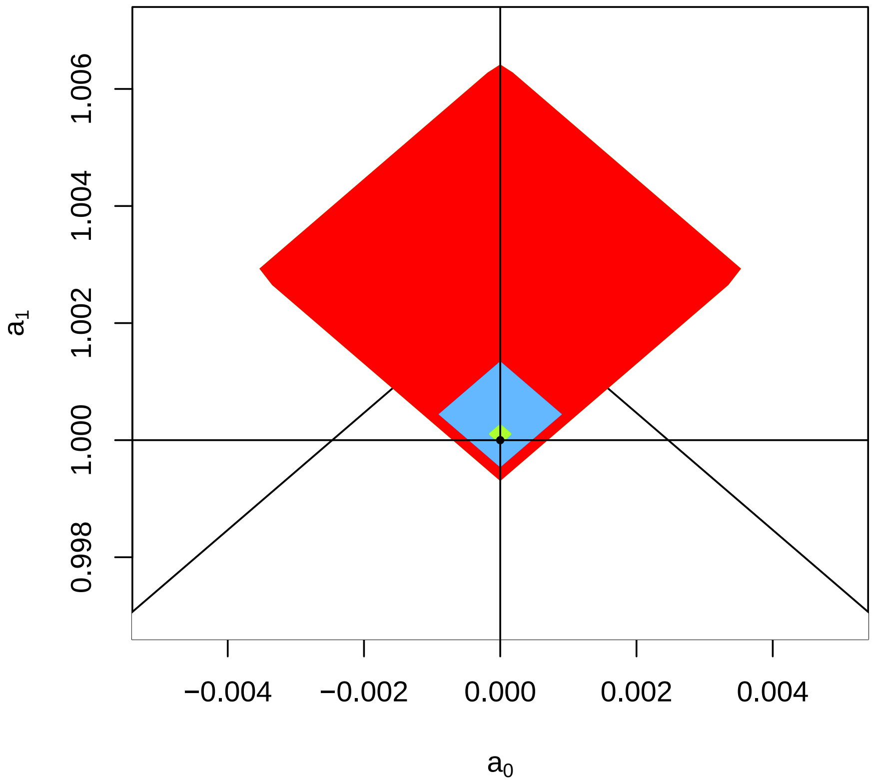

Figure 5: **Illustration A2.** Identified sets delivered by a binary endogenous variable SEIV model with a parametric ordered probit structural function, intercept a_0 and slope a_1. Number of categories of the outcome, M: 25(red), 50(blue) and 75(green).

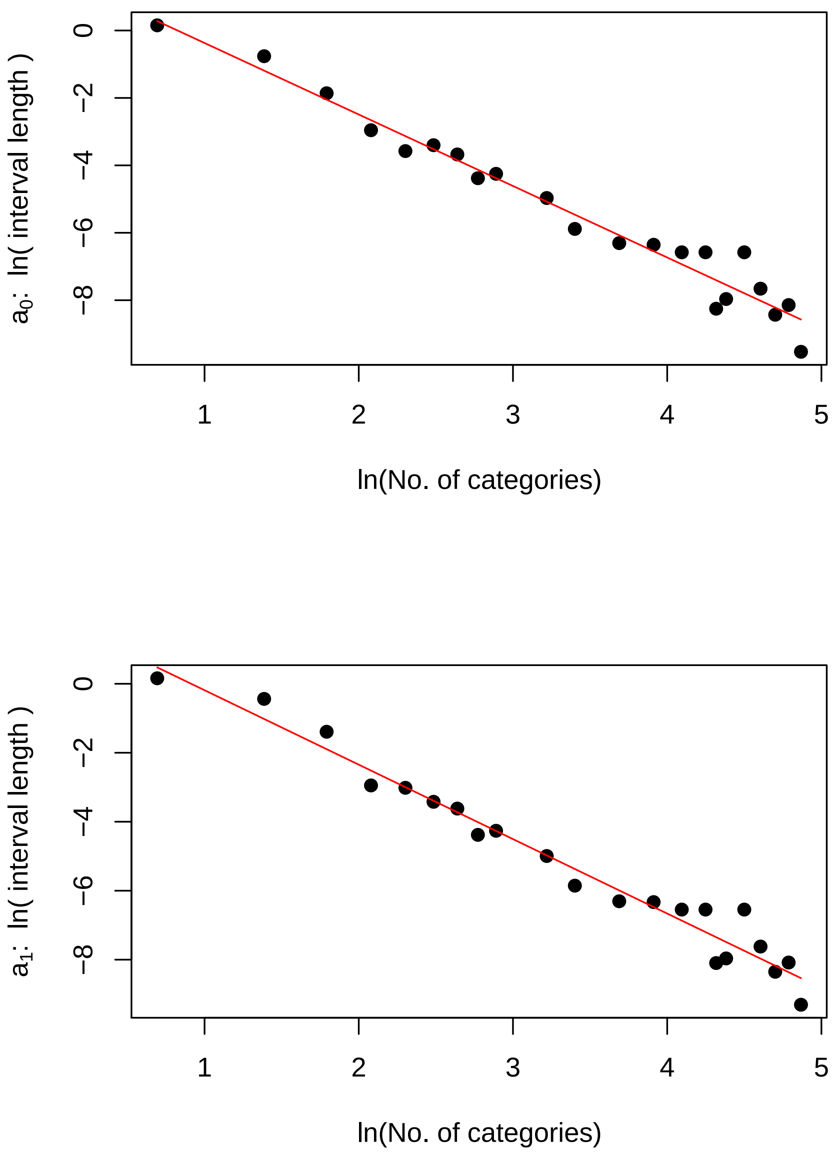

Figure 6: **Illustration A2.** Reduction of identified set as the number of outcome categories increases: (upper pane) logarithm of length of the identified interval for a_0 plotted against logarithm of number of categories of the outcome, Y, (lower pane) logarithm of length of the identified interval for a_1 plotted against logarithm of number of categories of the outcome, Y.

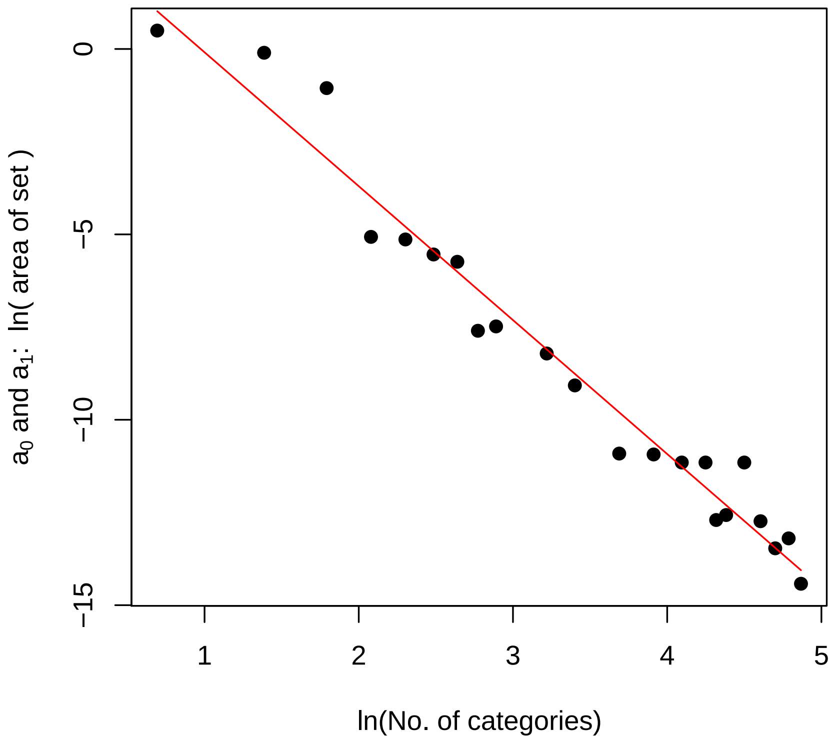

Figure 7: **Illustration A2.** Reduction of identified set as the number of outcome categories increases. Logarithm of area of the identified set plotted against logarithm of number of categories of the outcome, Y.

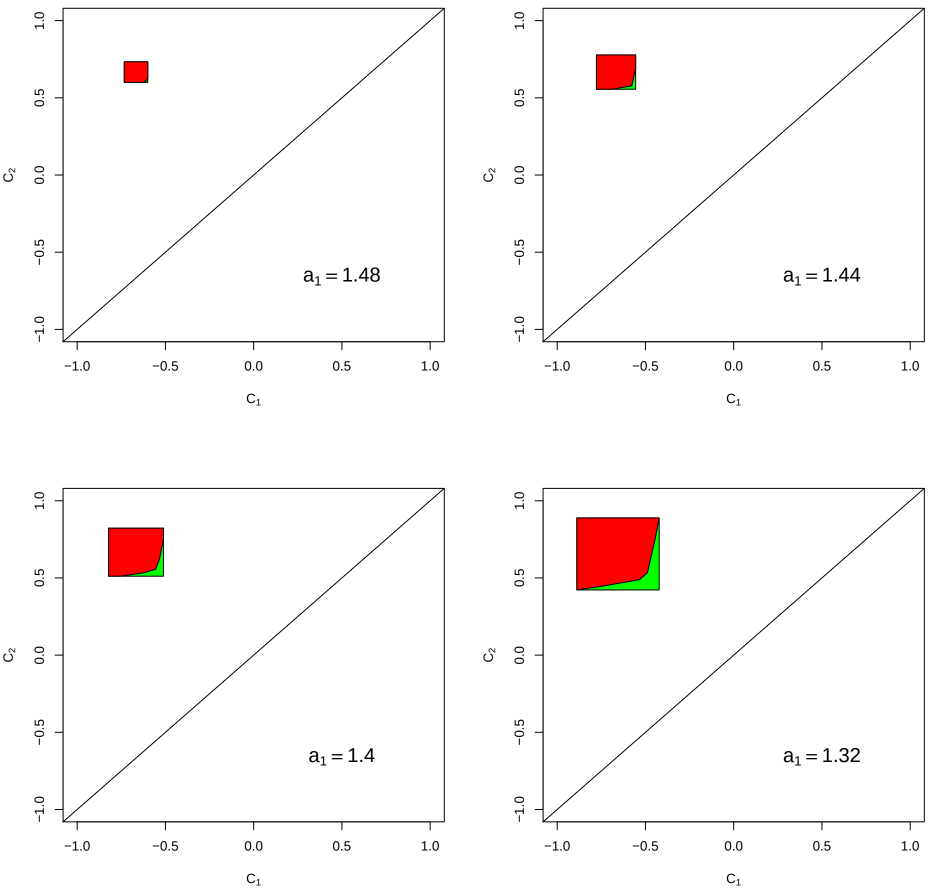

Figure 8: **Illustration B1.** Three class ordered probit model with unknown threshold parameters c_1 and c_2 and slope coefficient a_1. Cross-section of the identified set (red) and outer set (red and green) for c_1, c_2 and a_1 at selected values of a_1.

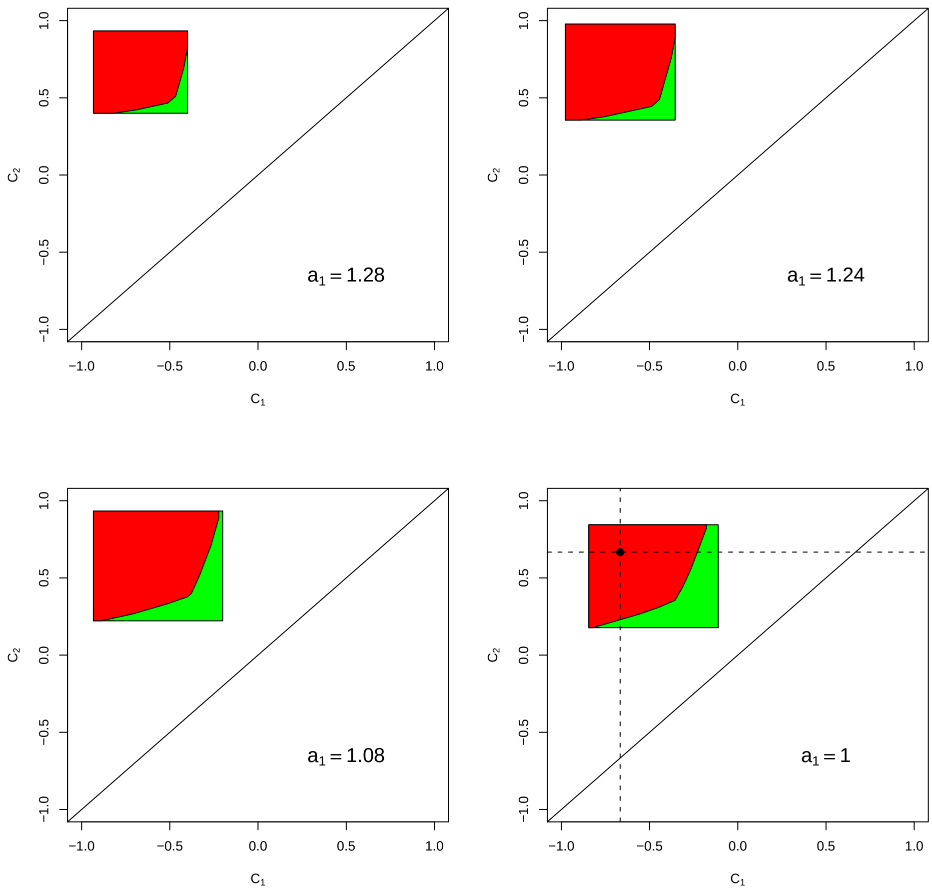

Figure 9: **Illustration B1.** Three class ordered probit model with unknown threshold parameters c_1 and c_2 and slope coefficient a_1. Cross-section of the identified set (red) and outer set (red and green) for c_1, c_2 and a_1 at selected values of a_1.

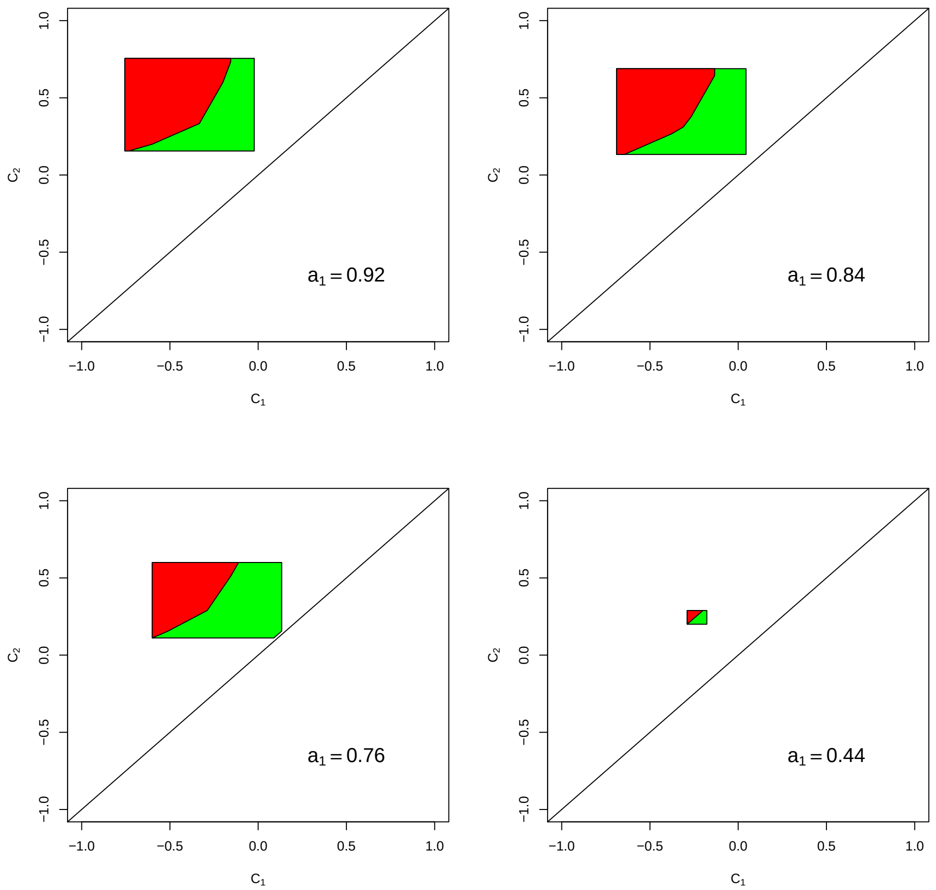

Figure 10: **Illustration B1.** Three class ordered probit model with unknown threshold parameters c_1 and c_2 and slope coefficient a_1. Cross-section of the identified set (red) and outer set (red and green) for c_1, c_2 and a_1 at selected values of a_1.