Sharp identified sets for discrete variable IV models

Andrew Chesher and Konrad Smolinski*
CeMMAP & UCL

May 4th 2010

Abstract. Instrumental variable models for discrete outcomes are set, not point, identifying. The paper characterises identified sets of structural functions when endogenous variables are discrete. Identified sets are unions of large numbers of convex sets and may not be convex nor even connected. Each of the component sets is a projection of a convex set that resides in a much higher dimensional space onto the space in which a structural function resides. The paper develops a symbolic expression for this projection and gives a constructive demonstration that it is indeed the identified set. We provide a Mathematica™ notebook which computes the set symbolically. We derive properties of the set, suggest how the set can be used in practical econometric analysis when outcomes and endogenous variables are discrete and propose a method for estimating identified sets under parametric or shape restrictions..

Keywords: Discrete endogenous variables, Discrete outcomes, Endogeneity, Fourier-Motzkin Elimination, Incomplete models, Instrumental variables, Set identification, Threshold Crossing Models.

JEL Codes: C10, C14, C50, C51.

1. Introduction

This paper gives new results on the identifying power of single equation instrumental variable (SEIV) models in which both the outcome of interest and potentially endogenous explanatory variables are discrete. These models generally set rather than point identify structural functions.[1] The paper derives the identified set for the general case in which there is an M-valued outcome and there are endogenous variables with K points of support.

The discrete outcome, discrete endogenous variable case studied here arises frequently in applied econometrics practice. Examples of settings in which the results of the paper are useful include situations in which a binary or ordered probit, or a logit or a count data analysis or some semiparametric or nonparametric alternative would be considered and explanatory variables are endogenous. We study nonparametric models but, as we show, characterizations of identified sets for nonparametric models are very useful in constructing identified sets in parametric cases.

In the instrumental variable model studied here an M-valued outcome, Y, is determined by a structural function characterised by $M-1$ threshold functions of possibly

*This is a revised and extended version of the paper "Set identifying models with discrete outcomes and endogenous variables" February 5th 2010. 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) and the intellectual support of our colleagues at CeMMAP with whom we have had many useful discussions of this work.

[1] See Chesher (2010).

endogenous variables X. Instrumental variables, Z, are excluded from these threshold functions. The instrumental variables and the stochastic term whose value relative to the threshold functions determines the value of Y are independently distributed. When endogenous variables have K points of support the structural function is characterised by $N = K(M-1)$ parameters: the values of the $M-1$ threshold functions at the K values of X. A conventional parametric model, for example an ordered probit model, places restrictions on these N objects.

The model studied here places no restrictions on the process generating values of the potentially endogenous variables X. It is in this sense that it is a *single equation* model.

By contrast the commonly employed control function approach to identification employs a more restrictive triangular model which places restrictions on the process generating the potentially endogenous variables.[2] That model generally fails to deliver point identification when endogenous variables are discrete so the SEIV model is a leading contender for application in practice.

The single equation approach taken in this paper has some other points to recommend it. For example, structural simultaneous equations models for discrete endogenous variables throw up coherency issues, first studied in Heckman (1978) and subsequently discussed in, for example, Lewbel (2007). These can be neglected in a single equation analysis. Economic models involving simultaneous determination of values of discrete outcomes can involve multiple equilibria. See for example Tamer (2003). Taking a single equation approach as here one is free to leave the equilibrium selection process unspecified.

After normalisation the structural function in the discrete-outcome, discrete-endogenous variable case is characterised a point in the unit N-cube. The set that is identified by a single equation instrumental variable model is a subset of this space. We show that the identified set is the union of many convex sets each of which is an intersection of linear half-spaces. The faces of these component convex polytopes are arranged either parallel to or at 45° angles to faces of the N-cube. The identified set may not be convex or even connected.

The convex components of the identified sets are projections of high-dimensional sets onto the space in which the structural function resides. Direct computation of these sets is challenging. Calculation for small scale problems can be done using the method of Fourier-Motzkin elimination. However for M or K larger than 4 the computations are prohibitively time consuming because of the very large number of inequalities produced during the process of projection. Almost all of these are redundant, but determining which are redundant is very computationally intensive. The key to solving this problem is to make use of the structure placed on the problem by the SEIV model.

We consider probability distributions for Y and X conditional on $Z = z$ for values of z in some set of instrumental values $\mathcal{Z}$. We develop a system of inequalities which must be satisfied by the N values that characterise a structural function for all structural functions that are elements of structures admitted by the SEIV model which generate these probability distributions. The identified set of structural func-

[2] See for example Blundell and Powell (2003, 2004), Chesher (2003), Imbens and Newey (2009).

tions must be a subset of the set defined by these inequalities. We show using a constructive proof that the set is precisely the identified set.

Calculation of the convex components of an identified set using the expressions we present here is very easy. The remaining, non-trivial, computational challenge is to deal with the very large number of convex components that arises when M or K is large. This problem disappears if sufficiently strong shape restrictions can be invoked. Parametric models are useful in providing these. An alternative is to employ shape constrained sieve approximations.

We show how recently developed results on set estimation and inference when sets are defined by intersection bounds can be used to operationalise the results given here.

The restrictions of the SEIV model are now set out and then the results given here are set in the context of earlier work.

1.1. The single equation instrumental variable model. In the SEIV model a scalar discrete outcome, Y, is determined by a structural function h as follows.

$$Y = h(X, U) \tag{1}$$

Here U is a scalar unobservable continuously distributed random variable and X is a list of explanatory variables. These explanatory variables may be endogenous in the sense that U and X may not be independently distributed. The focus is on identification of the structural function h.

In practice there may be variables appearing in h that are restricted to be exogenous (distributed independently of U) and the results of the paper are easily extended to accommodate these but for simplicity we proceed with the structural function specified as in equation (1).

The structural function h is restricted to be monotone in U for all values of X. It is normalized weakly increasing in what follows and the marginal distribution of U is normalized uniform on the closed unit interval $[0, 1]$. The support of X is denoted by $\mathcal{X}$.

The discrete outcome Y has M fixed points of support and without loss of generality these are taken to be the integers $1, \ldots, M$. Since h varies monotonically with U there is the following threshold crossing representation of the structural function: for $m \in \{1, \ldots, M\}$:

$$h(x, u) = m \text{ if and only if } h_{m-1}(x) < u \leq h_m(x)$$

with $h_0(x) = 0$ and $h_M(x) = 1$ for all $x \in \mathcal{X}$. In this paper we study the case in which X is discrete with a finite number, K, of points of support: $\mathcal{X} = \{x_1, \ldots, x_K\}$.

In this set-up a standard parametric probit model for $Y \in \{1, 2\}$ would have threshold functions as follows:

$$h_0(x) = 0 \qquad h_1(x) = \Phi(\alpha_0 + \alpha_1 x) \qquad h_2(x) = 1$$

where $\Phi(\cdot)$ is the standard normal distribution function. A standard logit model would have $h_1(x) = (1 + \exp(\alpha_0 + \alpha_1 x))^{-1}$.

If the model restricted X to be exogenous then it would identify the threshold functions at each point in the support of X because in that case $\Pr[Y \leq m | X = x] = h_m(x)$.

The SEIV model does not require X to be exogenous but admits instrumental variables, one or many, discrete or continuous, arranged in a vector Z which takes values in a set $\mathcal{Z}$. The instrumental variables Z and U are independently distributed and Z is excluded from the structural function.[3] The model set identifies the structural function.

1.2. Relation to earlier work. The SEIV model studied here is an example of the sort of nonseparable model studied in Chernozhukov and Hansen (2005), Blundell and Powell (2003, 2004), Chesher (2003) and Imbens and Newey (2009).

All but the first of these papers study complete models which specify triangular equation systems in which there are structural equations for endogenous explanatory variables as well as for the outcome of interest. When endogenous variables are continuous these models can point identify structural functions but when endogenous variables are discrete they do not. Dealing with the discrete endogenous variable case, Chesher (2005) introduces an additional restriction on the nature of the dependence amongst unobservables providing a set identifying triangular model with discrete endogenous variables. Jun, Pinkse and Xu (2009) provide some refinements. Point identification can be achieved under parametric restrictions such as those used in Heckman (1978).

Discreteness of endogenous variables is not a problem for SEIV models, indeed it brings simplifications - for example eliminating the "ill posed inverse problem" which arises when endogenous variables are continuous. This is shown clearly in Das (2005) where *additive error* nonparametric models with discrete endogenous variables and instrumental variable restrictions are considered. Because of the additive error restrictions this construction is not well suited to modelling discrete outcomes which sit more comfortably in the nonseparable error setting studied here.

Chernozhukov and Hansen (2005) study a nonadditive-error SEIV model like that considered here, focussing on the case in which the outcome is *continuous*. The identification results of that paper are built around the following equality which, when Y is continuous, holds for all $\tau \in (0,1)$ and all $z \in \mathcal{Z}$.

$$\Pr[Y = h(X,\tau)|Z = z] = \tau \tag{2}$$

Additional (completeness) conditions are provided under which the model point identifies the structural function.

The condition (2) does not hold when Y is discrete. Instead, as shown in Chernozhukov and Hansen (2001), there are the following inequalities which hold for all $\tau \in (0,1)$ and $z \in Z$.

$$\Pr[Y < h(X,\tau)|Z = z] < \tau \leq \Pr[Y \leq h(X,\tau)|Z = z]$$

These imply that the inequalities:

$$\max_{z\in\mathcal{Z}} \Pr[Y < h(X,\tau)|Z = z] < \tau \leq \min_{z\in\mathcal{Z}} \Pr[Y \leq h(X,\tau)|Z = z] \tag{3}$$

[3] At no point in the development is Z required to be a random variable. It could for example be a variable whose values are set by an experimenter. The key requirement is that the conditional distribution of U given $Z = z$ be invariant with respect to changes in z within the set $\mathcal{Z}$.

hold for all $\tau \in (0,1)$ as shown in Chesher (2007, 2010). The result is that the SEIV model generally fails to point identify the structural function when the *outcome* Y is discrete. However the model can be informative about the structural function as long as $\mathcal{Z}$ is not a singleton.

To see this suppose that for some value m and two values in $\mathcal{Z}$, z_1 and z_2, $\Pr[Y \leq m|Z = z_1] \neq \Pr[Y \leq m|Z = z_2]$. The restrictions of the model imply that in this case $h_m(x)$ is *not constant* for variations in x in admissible structures which generate the probability distribution under consideration. This is so because if $h_m(x)$ were constant, equal say to h_m^*, then *for all* $z \in \mathcal{Z}$, $\Pr[Y \leq m|Z = z] = h_m^*$ so *any* variation in $\Pr[Y \leq m|Z = z]$ with z rules out the possibility that $h_m(x)$ is constant for variations in x.[4] At least the set of structural functions identified by the SEIV model excludes structures with constant threshold functions if the outcome and instruments are not independently distributed.

The set identifying power of the SEIV model when the outcome is discrete was first studied in Chesher (2007). Let $\mathcal{H}(\mathcal{Z})$ denote the identified set of structural functions associated with some probability distribution $F_{YX|Z}$ for Y and X given $Z = z \in \mathcal{Z}$.[5] Chesher (2007, 2010) develops a set, denoted here by $\mathcal{C}(\mathcal{Z})$, based on the inequalities (3). It is shown that, when Y is binary and X is continuous, $\mathcal{H}(\mathcal{Z}) = \mathcal{C}(\mathcal{Z})$ and $\mathcal{C}(\mathcal{Z})$ provides tight set identification. In other cases it is an outer set in the sense that it can be that $\mathcal{H}(\mathcal{Z}) \subset \mathcal{C}(\mathcal{Z})$.

Chesher (2009) studies the binary outcome case, proving $\mathcal{H}(\mathcal{Z}) = \mathcal{C}(\mathcal{Z})$ when endogenous variables are discrete, considering the impact of parametric restrictions and shape restrictions, and giving some results on estimation under shape restrictions employing results on inference using intersection bounds given in Chernozhukov, Lee and Rosen (2009).

In Chesher and Smolinski (2009) a refinement[6] to $\mathcal{C}(\mathcal{Z})$, denoted $\tilde{\mathcal{D}}(\mathcal{Z})$, is developed. This delivers the identified set when there is a single binary endogenous variable no matter how many points of support the outcome Y has. The results are used in an investigation of the nature of the reduction in extent of the identified set as the number of points of support of Y increases in an endogenous parametric ordered probit example.

This paper studies the general finite M-outcome, K-point of support discrete

[4] There is the following.

$$\begin{aligned}\Pr[Y \leq m|Z = z] &= \sum_k \Pr[U \leq h_m^*|X = x_k, Z = z] \Pr[X = x_k|Z = z] \\ &= \Pr[U \leq h_m^*|Z = z] \\ &= h_m^*\end{aligned}$$

Since the model excludes Z from the structural function h and requires U and Z to be independent the only way in which Z can affect the distribution of Y is through its effect on X and then only if h is sensitive to variations in X.

[5] It would be clearer to give a distinctive symbol to the probability distribution under consideration, e.g. $F^0_{YX|Z}$ and label the various sets accordingly thus: $\mathcal{H}^0(\mathcal{Z})$, $\mathcal{C}^0(\mathcal{Z})$ and so forth. We do not do this here because the notation quickly becomes cumbersome. However it is important to keep in mind that each of the sets under discussion is associated with a particular probability distribution.

[6] By a refinement we mean that $\tilde{\mathcal{D}}(\mathcal{Z}) \subseteq \mathcal{C}(\mathcal{Z})$ with the possibility that $\tilde{\mathcal{D}}(\mathcal{Z}) \subset \mathcal{C}(\mathcal{Z})$

endogenous variable case and develops a further refinement[7] to $\mathcal{C}(\mathcal{Z})$, denoted $\mathcal{E}(\mathcal{Z})$ and shows that $\mathcal{E}(\mathcal{Z})$ is precisely the identified set, $\mathcal{H}(\mathcal{Z})$.

1.3. Plan of the paper. Section 2 defines the set identified by the SEIV model and reviews its characteristics.

Section 3 develops the new set, $\mathcal{E}(\mathcal{Z})$, shows that it contains the identified set, and then gives a constructive proof that $\mathcal{E}(\mathcal{Z})$ is the identified set. This is done by proposing an algorithm for construction of a proper distribution for the unobservable U and endogenous X conditional on values of instrumental variables which for any value of γ in $\mathcal{E}(\mathcal{Z})$ delivers the probabilities used to construct the set while respecting the restriction that U and Z be independently distributed.

Section 3.1 develops some properties of the identified set. Section 3.2 presents the algorithm for constructing the distribution of U and X given Z which is used to demonstrate that $\mathcal{E}(\mathcal{Z})$ is the identified set. Section 3.3 gives an alternative derivation of the set $\mathcal{E}(\mathcal{Z})$ which is useful in linking the results of this paper with earlier results.

Section 3.4 sets out properties of the set $\mathcal{E}(\mathcal{Z})$ when the outcome is binary. Section 3.5 shows how the set $\mathcal{E}(\mathcal{Z})$ is related to the set defined in Chesher (2010). Section 3.6 gives alternative expressions for the inequalities defining the new set which help clarify its relationship to the set defined by the inequalities (3).

Section 4 gives some illustrative calculations, describes a `Mathematica` notebook which does symbolic calculation of the convex components of the identified set and discusses estimation and inference. Section 5 concludes.

2. The identified set

In this Section the set identified by the SEIV model is defined and notation is introduced.

We consider situations in which X, which may be a scalar or a vector, is discrete and takes values in the set $\mathcal{X} = \{x_k\}_{k=1}^K$. In this case the structural function h is characterized by $N \equiv K(M-1)$ parameters as follows,

$$\gamma_{mk} \equiv h_m(x_k), \qquad m \in \{1, \ldots, M-1\}, \quad k \in \{1, \ldots, K\}$$

which are arranged in a vector $\boldsymbol{\gamma}$, as follows.[8]

$$\boldsymbol{\gamma} \equiv [\gamma_{11}, \ldots, \gamma_{1K}, \gamma_{21}, \ldots, \gamma_{2K}, \ldots, \gamma_{M-1,1}, \ldots, \gamma_{M-1,K}]$$

Identification of the vector $\boldsymbol{\gamma}$ is studied in this paper. Each element of $\boldsymbol{\gamma}$ lies in the unit interval so each value of $\boldsymbol{\gamma}$ is a point in the unit N-cube. The identified set is a

[7] Here, by a refinement we mean that $\mathcal{E}(\mathcal{Z}) \subseteq \tilde{\mathcal{D}}(\mathcal{Z}) \subseteq \mathcal{C}(\mathcal{Z})$.

[8] If Γ is a matrix with (m,k) element equal to γ_{mk} then $\gamma \equiv \text{vec}(\Gamma')$. Considering γ_r, the rth element of γ, there are the following relationships.

$$r = (m-1)K + k$$

$$k = r \text{ modulo } K \qquad m = (r-k)/K + 1$$

subset of the unit N-cube. There are the restrictions $\gamma_{lk} < \gamma_{mk}$ for $l < m$ and all k. Henceforth "for all k" means for $k \in \{1, \ldots, K\}$.

Consider a particular probability distribution for Y and X given $Z = z \in \mathcal{Z}$. The identified set of values of $\boldsymbol{\gamma}$ associated with this distribution contains all and only values of $\boldsymbol{\gamma}$ for which there exist admissible conditional probability distributions of U and X given $Z = z$ for all values of z in $\mathcal{Z}$ such that the resulting structures deliver the probability distribution under consideration. Notation for that probability distribution is now introduced.

For values $z \in \mathcal{Z}$, for $m \in \{1, \ldots, M\}$ and $k \in \{1, \ldots, K\}$, there are the following point probabilities:

$$\rho_{mk}(z) \equiv \Pr[Y = m \wedge X = x_k | Z = z]$$

and cumulative probabilities:

$$\bar{\rho}_{mk}(z) \equiv \Pr[Y \leq m \wedge X = x_k | Z = z]$$

and probabilities marginal with respect to Y:

$$\delta_k(z) \equiv \Pr[X = x_k | Z = z] = \bar{\rho}_{Mk}(z).$$

For all k define: $\gamma_{0k} = 0$, $\gamma_{Mk} = 1$, $\rho_{0k}(z) = \bar{\rho}_{0k}(z) = 0$. In what follows "for all m" means for $m \in \{0, \ldots, M\}$.

Associated with a particular value of $\boldsymbol{\gamma}$ and each value $z \in \mathcal{Z}$, define a piecewise uniform conditional distribution for U and X given Z, such that for all m, k and k':

$$\bar{\eta}_{mkk'}(z) \equiv \Pr[U \leq \gamma_{mk} \wedge X = x_{k'} | Z = z]$$

and let $\bar{\boldsymbol{\eta}}(z)$ denote the complete list of $(M-1)K^2$ such terms.[9]

A list of values of $(\boldsymbol{\gamma}, \bar{\boldsymbol{\eta}}(z))$ produced as z varies in $\mathcal{Z}$ characterizes a *structure* which is *admissible* if it satisfies the following *independence* and *properness* conditions.

[**1**]. **Independence.** For all $z \in \mathcal{Z}$ and for all m and k the following equalities hold.[10]

$$\sum_{k'=1}^{K} \bar{\eta}_{mkk'}(z) = \gamma_{mk} \tag{4}$$

[**2**]. **Properness**. For all $z \in \mathcal{Z}$ and for all j, k, l, m and k', $\bar{\eta}_{ljk'}(z) \leq \bar{\eta}_{mkk'}(z)$ if and only if $\gamma_{lj} \leq \gamma_{mk}$. For all $z \in \mathcal{Z}$ and for all k, k', $\bar{\eta}_{0kk'}(z) = 0$. For all $z \in \mathcal{Z}$ and for all k' $\sum_{k=1}^{K} \bar{\eta}_{Mkk'}(z) = 1$.

If in addition the following *observational equivalence* condition is satisfied then the structure generates the probability distribution under consideration.

[9] Between each pair of adjacent knots, γ_{mk}, each conditional density function for Y given X and Z is uniform. The construction is justified in Chesher (2009). The conditional density functions have a histogram-like appearance.

[10] The left hand sides are $\Pr[U \leq \gamma_{mk} | Z = z]$ which the independence restriction requires to be free of z. The values γ_{mk} on the right hand sides arises because of the uniform distribution normalisation of the marginal distribution of U. See Chesher (2010).

[3]. Observational equivalence. For all $z \in \mathcal{Z}$ and for all m and k the following equalities hold.

$$\bar{\eta}_{mkk}(z) = \bar{\rho}_{mk}(z) \tag{5}$$

All and *only* structures that obey conditions [1], [2] and [3] are in the set of structures identified by the model for the probabilities considered. Let $\mathcal{S}(\mathcal{Z})$ denote that set.

The identified set of structural *functions*, $\mathcal{H}(\mathcal{Z})$, is the set of values of $\boldsymbol{\gamma}$ for which there are values of $\bar{\boldsymbol{\eta}}(z)$ for $z \in \mathcal{Z}$ such that the resulting structure is in the identified set, $\mathcal{S}(\mathcal{Z})$. The identified set for $\boldsymbol{\gamma}$, $\mathcal{H}(\mathcal{Z})$, is the *projection* of $\mathcal{S}(\mathcal{Z})$ onto the unit N-cube within which all values of $\boldsymbol{\gamma}$ lie.

The geometry of these sets is considered in Chesher and Smolinski (2009). A brief account is given here. Because of the properness condition [2] the order in which the elements of $\boldsymbol{\gamma}$ lie is an important consideration.

There are

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

admissible arrangements of the elements of $\boldsymbol{\gamma}$.[11]

In each arrangement, $t \in \{1, \ldots, T\}$, the set of admissible observationally equivalent structures defined by [1], [2] and [3], denoted by $\mathcal{S}_t(\mathcal{Z})$, is either empty or a convex polytope because it is an intersection of bounded linear half spaces. The identified set of structures is the union of the sets obtained under each admissible arrangement.

$$\mathcal{S}(\mathcal{Z}) = \bigcup_{t=1}^{T} \mathcal{S}_t(\mathcal{Z})$$

In each arrangement, t, the identified set of structural functions obtained by projecting away $\bar{\boldsymbol{\eta}}(z)$ for $z \in \mathcal{Z}$, denoted $\mathcal{H}_t(\mathcal{Z})$, is also either empty or a convex polytope. The complete identified set of structural functions is the union of these convex sets. The result may not itself be convex, nor even connected.

$$\mathcal{H}(\mathcal{Z}) = \bigcup_{t=1}^{T} \mathcal{H}_t(\mathcal{Z})$$

In problems in which M or K are at all large computation of an identified set of structural functions is difficult. A head on attack would consider each admissible

[11] Arrangements in which there is a pair of indices m and m' with $m > m'$ such that for some k, $\gamma_{mk} \leq \gamma_{m'k}$ are inadmissible. The formula for T arises as follows. There are $\binom{N}{M-1}$ ways of placing $\gamma_{11}, \gamma_{21}, \ldots, \gamma_{M-1,1}$ in the $N = (M-1)K$ places available and only one order in which those values can lie. There are then $\binom{N-(M-1)}{M-1}$ ways of placing $\gamma_{12}, \gamma_{22}, \ldots, \gamma_{M-1,2}$ in the remaining $N-(M-1)$ places. Continuing in this way it is clear that there are

$$\prod_{k=1}^{K} \binom{N-(k-1)(M-1)}{M-1}$$

admissible arrangements of γ which on simplification yields the formula (6) for T.

arrangement in turn and use the method of Fourier-Motzkin elimination[12] to project away the $(M-1)K^2$ elements in $\bar{\boldsymbol{\eta}}(z)$ for each $z \in \mathcal{Z}$ but this is computationally infeasible when M and K are large.

In the next Section we develop easy-to-compute sets which are shown to be precisely the identified set of structural functions, $\mathcal{H}(\mathcal{Z})$.

3. SHARP SET IDENTIFICATION OF THE STRUCTURAL FUNCTION

Conditions [1] - [3] place restrictions on values of $(\boldsymbol{\gamma}, \bar{\boldsymbol{\eta}}(z))$ for z varying in $\mathcal{Z}$. They define the identified set of *structures*: $\mathcal{S}(\mathcal{Z})$.

In this Section we develop implications of these restrictions for admissible values of $\boldsymbol{\gamma}$, that is for values of $\boldsymbol{\gamma}$ that lie in the identified set of *structural functions*: $\mathcal{H}(\mathcal{Z})$. The result is a list of inequalities that define a set denoted $\mathcal{E}(\mathcal{Z})$. We show that this is the identified set $\mathcal{H}(\mathcal{Z})$.

The ordering of the elements of $\boldsymbol{\gamma}$ is important. The set $\mathcal{E}(\mathcal{Z})$ is a union of convex sets, $\mathcal{E}_t(\mathcal{Z})$, one associated with each admissible arrangement, t, of $\boldsymbol{\gamma}$.

$$\mathcal{E}(\mathcal{Z}) = \bigcup_{t=1}^{T} \mathcal{E}_t(\mathcal{Z})$$

Each set $\mathcal{E}_t(\mathcal{Z})$ is defined as an intersection of linear half spaces.

We proceed to develop a definition of a set $\mathcal{E}_t(\mathcal{Z})$ obtained *under a particular arrangement*, t, of the elements of $\boldsymbol{\gamma}$. First it is necessary to develop notation and functions for dealing with arrangements.

Let $\gamma_{[n]}$ be the nth largest value in an arrangement. Recall there are $N \equiv (M-1)K$ elements in $\boldsymbol{\gamma}$. We adopt the notation used in the literature on order statistics to denote the ordered values of $\boldsymbol{\gamma}$:

$$\gamma_{[1]} \leq \gamma_{[2]} \cdots \leq \gamma_{[N-1]} \leq \gamma_{[N]}$$

and we define $\gamma_{[0]} \equiv 0$ and $\gamma_{[N+1]} \equiv 1$.[13]

Define functions $m(n)$ and $k(n)$ such that $\gamma_{m(n)k(n)} = \gamma_{[n]}$. Define $m(0) = 0$. With $M = 3$ and $K = 3$, for which $N = 6$ and

$$\boldsymbol{\gamma} = [\gamma_{11}, \gamma_{12}, \gamma_{13}, \gamma_{21}, \gamma_{22}, \gamma_{23}]$$

and for the arrangement

$$[\gamma_{11}, \gamma_{12}, \gamma_{21}, \gamma_{13}, \gamma_{23}, \gamma_{22}] \tag{7}$$

[12] See Zeigler (2007).

[13] A more precise notation would carry an identifier of the arrangement t under consideration and when stating formal results we do employ such a notation, for example denoting the ordered elements of an arrangement t by $\gamma^t_{[1]}, \ldots, \gamma^t_{[N]}$. During the exposition, while it is clear that a particular arrangement is under consideration, we simplify notation and do not make dependence on the arrangement under consideration explicit.

the functions $m(\cdot)$ and $k(\cdot)$ are as shown below. We will work with this example throughout this Section.

n	$m(n)$	$k(n)$
1	1	1
2	1	2
3	2	1
4	1	3
5	2	3
6	2	2

Figure 1 shows a configuration of threshold functions that is consonant with this arrangement. In this case $M = 3$ so there are two threshold functions, $h_1(x)$ and $h_2(x)$.

Values of X are measured along the horizontal axis in Figure 1 and three points of support, x_1, x_2 and x_3 are marked. Values of threshold functions are measured along the vertical axis which is the unit interval $[0, 1]$. This axis also measures values of the unobservable variable U.

At any value of x, values of U falling on or below the lowest threshold function deliver the value 1 for Y, values of U falling between the two threshold functions or on the highest threshold function deliver the value 2 for Y and values of U falling above the highest threshold deliver the value 3 for Y. Notice that the upper threshold function is not monotone in x reflecting the inequality $\gamma_{23} < \gamma_{22}$.

Now define an *inverse* function $n(m, k)$ such that $\gamma_{mk} = \gamma_{[n(m,k)]}$ and note that $n(m(n), k(n)) = n$. For all k define $n(0, k) = 0$.

For the arrangement (7) considered above the function $n(\cdot, \cdot)$ delivers values as shown below.

	$k = 1$	$k = 2$	$k = 3$
$m = 1$	1	2	4
$m = 2$	3	6	5

The functions $m(\cdot)$, $k(\cdot)$ and $n(\cdot, \cdot)$ are specific to the particular arrangement under consideration and we could make this dependence explicit by writing $m_t(\cdot)$, $k_t(\cdot)$ and $n_t(\cdot, \cdot)$ but for the most part this is not done in order to avoid excessively complex notation.

We use an abbreviated notation as follows: $\bar{\rho}_{[n]}$ denotes $\bar{\rho}_{m(n)k(n)}$, thus:

$$\bar{\rho}_{[n]} \equiv \bar{\rho}_{m(n)k(n)} = \Pr[Y \leq m(n) \wedge X = x_{k(n)} | Z = z]$$

and $\bar{\eta}_{[n]k'}$ denotes $\bar{\eta}_{m(n)k(n)k'}$ thus.

$$\bar{\eta}_{[n]k'} \equiv \bar{\eta}_{m(n)k(n)k'} = \Pr[U \leq \gamma_{m(n)k(n)} \wedge X = x_{k'} | Z = z]$$

There are associated non-cumulative probabilities as follows.[14]

$$\rho_{[n]} = \Pr[Y = m(n) \wedge X = x_{k(n)} | Z = z]$$

[14] Here too we could make dependence on the arrangement under consideration explicit in the notation, e.g. writing $\rho^t_{[n]}$ and $\eta^t_{[i]k}$, but do not do so until we come to formal statements of results.

$$\eta_{[n]k'} = \Pr[U \in (\gamma_{[n-1]}, \gamma_{[n]}] \wedge X = x_{k'} | Z = z]$$

It is important to understand that $\bar{\eta}_{[n]k'} = \sum_{j=1}^{n} \eta_{[j]k'}$ but in general $\bar{\rho}_{[n]} \neq \sum_{j=1}^{n} \rho_{[j]}$ rather:

$$\bar{\rho}_{[n]} = \sum_{j=1}^{m(n)} \rho_{jk(n)} = \sum_{j=1}^{m(n)} \rho_{[n(j,k(n))]}.$$

All these probabilities depend on the instrumental value z under consideration but this dependence is not made explicit in the notation for the moment. Define $\rho_{[0]} = \bar{\rho}_{[0]} = 0$, $\eta_{[n]k} = \bar{\eta}_{[n]k} = 0$.

It is helpful to extend the definitions to cover probability masses associated with the Mth point of support of Y so for all k define

$$n(M, k) = N + k$$

and

$$\begin{aligned} m(N+k) &= M \\ k(N+k) &= k \end{aligned}$$

which lead to

$$\rho_{[N+k]} = \Pr[Y = M \wedge X = x_k | Z = z]$$

with associated cumulative probabilities

$$\bar{\rho}_{[N+k]} = \Pr[Y \leq M \wedge X = x_k | Z = z] = \Pr[X = x_k | Z = z] = \delta_k.$$

Table 1 exhibits the probability masses $\eta_{[n]k}$ for a general case with an M-valued outcome and endogenous variables with K points of support and $N = K(M-1)$. All values are non-negative, values in column k sum to δ_k and the sum of all $K(N+1)$ probability masses in the table is 1. We will make extensive reference to tables like this in what follows.

We consider a particular value $z \in \mathcal{Z}$ and construct a set $\mathcal{E}_t(z)$, defining $\mathcal{E}_t(\mathcal{Z})$ as the intersection of such sets for z varying in $\mathcal{Z}$.

$$\mathcal{E}_t(\mathcal{Z}) \equiv \bigcap_{z \in \mathcal{Z}} \mathcal{E}_t(z)$$

To avoid cumbersome notation we do not make the dependence of probabilities on the chosen value z explicit in the notation for the moment.

Each set $\mathcal{E}_t(z)$ is obtained by considering restrictions that Conditions [1] - [3] place on the elements of $\boldsymbol{\gamma}$ when they are in arrangement t. The restrictions arise because for all values of $\boldsymbol{\gamma}$ (see the final column of the Table) that lie in the identified set there exist values of the probability masses $\eta_{[n]k}$ such that:

1. the sum of probability masses lying in rows 1 through n is equal to $\gamma_{[n]}$, equivalently, the sum of probability masses in row n is equal to $\gamma_{[n]} - \gamma_{[n-1]}$,

$$\sum_{i=1}^{n} \sum_{k=1}^{K} \eta_{[i]k} = \gamma_{[n]}$$

	Value of x				Ordered
n	x_1	x_2	$\cdots$	x_K	values of $\boldsymbol{\gamma}$
0	0	0	$\cdots$	0	$\gamma_{[0]}$
1	$\eta_{[1]1}$	$\eta_{[1]2}$	$\cdots$	$\eta_{[1]K}$	$\gamma_{[1]}$
2	$\eta_{[2]1}$	$\eta_{[2]2}$	$\cdots$	$\eta_{[2]K}$	$\gamma_{[2]}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$
n	$\eta_{[n]1}$	$\eta_{[n]2}$	$\cdots$	$\eta_{[n]K}$	$\gamma_{[n]}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$
N	$\eta_{[N]1}$	$\eta_{[N]2}$	$\cdots$	$\eta_{[N]K}$	$\gamma_{[N]}$
$N+1$	$\eta_{[N+1]1}$	$\eta_{[N+1]2}$	$\cdots$	$\eta_{[N+1]K}$	$\gamma_{[N+1]}$

Table 1: Piece-wise uniform joint distribution of U and X conditional on a value of Z arranged by ordered values of γ (rows) and points of support of the endogenous variable X (columns).

$$\sum_{k=1}^{K} \eta_{[i]k} = \gamma_{[n]} - \gamma_{[n-1]} \equiv \Delta\gamma_{[n]}$$

2. all probability masses are non-negative,

3. probability masses sum over appropriate blocks of cells in the table to deliver the observed probabilities $\rho_{[1]}, \ldots, \rho_{[n]}$.

Table 2 exhibits the blocks of cells over which probability masses must be aggregated for the arrangement shown in equation (7). Table 3 shows the values that must be achieved when summing within blocks if the observational equivalence condition is to be satisfied. For example in the arrangement considered there is $m(4) = 1$, $k(4) = 3$ and

$$\rho_{[4]} = \Pr[Y = 1 \wedge X = x_3 | Z = z] = \sum_{i=1}^{4} \eta_{[i]3} = \Pr[U \leq \gamma_{[4]} \wedge X = x_3 | Z = z]$$

must hold if the observational equivalence restriction is to be satisfied. Another example: $m(7) = 3$, $k(7) = 1$ and so observational equivalence requires that the following equalities hold.

$$\rho_{[7]} = \Pr[Y = 3 \wedge X = x_1 | Z = z] = \sum_{i=4}^{7} \eta_{[i]1} = \Pr[\gamma_{[3]} < U \leq \gamma_{[7]} \wedge X = x_1 | Z = z].$$

In general there are M blocks in each column of the table. In each row exactly one block terminates. The block of cells in which the mass $\rho_{[n]}$ must lie is in the column of the table associated with $x_{k(n)}$ and in the rows that end at $n(m(n), k(n)) = n$ and start at $n(m(n) - 1, k(n)) + 1$. So the observational equivalence restriction requires that the conditions:

$$\rho_{[n]} = \sum_{i=n(m(n)-1,k(n))+1}^{n} \eta_{[i]k(n)}$$

n	Value of x			Differences of ordered values of $\boldsymbol{\gamma}$	Ascending list of elements in $\boldsymbol{\gamma}$
	x_1	x_2	x_3		
0					0
1	$\eta_{[1]1}$	$\eta_{[1]2}$	$\eta_{[1]3}$	$\gamma_{[1]} - \gamma_{[0]}$	γ_{11}
2	$\eta_{[2]1}$	$\eta_{[2]2}$	$\eta_{[2]3}$	$\gamma_{[2]} - \gamma_{[1]}$	γ_{12}
3	$\eta_{[3]1}$	$\eta_{[3]2}$	$\eta_{[3]3}$	$\gamma_{[3]} - \gamma_{[2]}$	γ_{21}
4	$\eta_{[4]1}$	$\eta_{[4]2}$	$\eta_{[4]3}$	$\gamma_{[4]} - \gamma_{[3]}$	γ_{13}
5	$\eta_{[5]1}$	$\eta_{[5]2}$	$\eta_{[5]3}$	$\gamma_{[5]} - \gamma_{[4]}$	γ_{23}
6	$\eta_{[6]1}$	$\eta_{[6]2}$	$\eta_{[6]3}$	$\gamma_{[6]} - \gamma_{[5]}$	γ_{22}
7	$\eta_{[7]1}$	$\eta_{[7]2}$	$\eta_{[7]3}$	$\gamma_{[7]} - \gamma_{[6]}$	1

Table 2: Conditional mass function values arranged by ordered values of γ and values of the conditioning variable X showing blocks of cells whose mass must be aggregated when considering the observational equivalence condition.

n	Value of x			Differences of ordered values of $\boldsymbol{\gamma}$	List of elements of $\boldsymbol{\gamma}$ in ascending order
	x_1	x_2	x_3		
0					0
1	$\rho_{[1]}$			$\gamma_{[1]} - \gamma_{[0]}$	γ_{11}
2		$\rho_{[2]}$		$\gamma_{[2]} - \gamma_{[1]}$	γ_{12}
3	$\rho_{[3]}$			$\gamma_{[3]} - \gamma_{[2]}$	γ_{21}
4			$\rho_{[4]}$	$\gamma_{[4]} - \gamma_{[3]}$	γ_{13}
5			$\rho_{[5]}$	$\gamma_{[5]} - \gamma_{[4]}$	γ_{23}
6		$\rho_{[6]}$		$\gamma_{[6]} - \gamma_{[5]}$	γ_{22}
7	$\rho_{[7]}$	$\rho_{[8]}$	$\rho_{[9]}$	$\gamma_{[7]} - \gamma_{[6]}$	1

Table 3: Sums of probability masses in blocks of cells must aggregate to the indicated probabilities if the observational equivalence condition is to be satisfied.

hold for $n = 1, \ldots, N + K$.

A particular value of $\boldsymbol{\gamma}$ in arrangement t can only support an allocation of probability mass satisfying Conditions (1) - (3) if the elements $\gamma_{[1]}, \ldots, \gamma_{[N]}$ are spaced sufficiently far apart to permit the allocation of probability mass in the required amounts in the blocks of cells that arise in the arrangement. For example, in the arrangement (7), considering Table 3, $\gamma_{[1]}$ must be at least equal to $\rho_{[1]}$, $\gamma_{[2]}$ must be at least equal to $\rho_{[1]} + \rho_{[2]}$ and so forth.

There are additional restrictions. For example $\gamma_{[5]} - \gamma_{[3]}$ must be at least equal to $\rho_{[5]}$ and $\gamma_{[7]} - \gamma_{[4]} = 1 - \gamma_{[4]}$ must be at least $\rho_{[5]} + \rho_{[8]} + \rho_{[9]}$. We now develop a complete characterisation of these inequalities which determine the spacing between elements of $\boldsymbol{\gamma}$ under a particular arrangement such that the allocation of probability mass to blocks of cells that is required to deliver observational equivalence is feasible.

To proceed we introduce the idea of the *active indexes in a row*. The active indexes in row n are K distinct elements of the list $\{1, 2, \ldots, N + K\}$. These are the indexes, i, of probabilities $\rho_{[i]}$ to which cells in row n contribute. The active index

for column k of row n is defined as follows.

$$a_{nk} \equiv \min \{i : (n \leq i \leq N + K) \wedge (k(i) = k)\}$$

The active index list for row n is defined thus: $a_n \equiv \{a_{n1}, \ldots, a_{nK}\}$. Clearly $a_{mk} \leq a_{nk}$ for all $m \leq n$ and k. For all k define $a_{0k} = 0$. Each active index list a_n has n as a member and it is always the smallest member.

For the arrangement given in (7) there are, on considering Table 3, the following active index lists.

n	a_n
0	$\{0, 0, 0\}$
1	$\{1, 2, 4\}$
2	$\{3, 2, 4\}$
3	$\{3, 6, 4\}$
4	$\{7, 6, 4\}$
5	$\{7, 6, 5\}$
6	$\{7, 6, 9\}$
7	$\{7, 8, 9\}$

We now introduce the idea of *last-discharged indexes for a row.* The last discharged index for column k in row n is the index, i, of the probability $\rho_{[i]}$ falling in column k whose block of cells was most recently completed at row n. The last-discharged index for column k in row n is defined for all k and $n \in \{1, \ldots, N\}$ as follows.

$$d_{nk} \equiv \max \{i : (0 \leq i \leq n) \wedge (k(i) = k)\}$$

For all k define $d_{0k} \equiv 0$ and $d_{N+1,k} \equiv N + k$. Clearly $d_{mk} \leq d_{nk}$ for all $m \leq n$ and k.

The row n last-discharged index list is defined as $d_n \equiv \{d_{n1}, \ldots, d_{nK}\}$. Each list d_n has n as a member and, except in row $N + 1$, it is the largest member. For the arrangement (7) there are, on considering Table 3, the following last-discharged index lists.

n	d_n
0	$\{0, 0, 0\}$
1	$\{1, 0, 0\}$
2	$\{1, 2, 0\}$
3	$\{3, 2, 0\}$
4	$\{3, 2, 4\}$
5	$\{3, 2, 5\}$
6	$\{3, 6, 5\}$
7	$\{7, 8, 9\}$

For a pair of indexes, $(r, s) \in \{0, 1, \ldots, N + 1\}$ with $r < s$ there is a minimal probability mass required to fall between $\gamma_{[r]}$ and $\gamma_{[s]}$ if observational equivalence is to be achieved.

This minimal mass is calculated as follows. In a column, k, there is a probability mass equal to $\bar{\rho}_{[d_{sk}]}$ required to lie below $\gamma_{[s]}$ because d_{sk} is the discharged index

associated with row s and column k. From this must be removed any probability mass associated with the active index in column k of row r, a_{rk}, and any probability mass associated with active indexes for rows prior to r in column k (and so discharged by row r). This mass is given by $\bar{\rho}_{[a_{rk}]}$. This can exceed $\bar{\rho}_{[d_{sk}]}$ so the minimal probability mass required to fall in the interval $(\gamma_{[r]}, \gamma_{[s]}]$ associated with $X = x_k$ is $\max(0, \bar{\rho}_{[d_{sk}]} - \bar{\rho}_{[a_{rk}]})$ and the total (across all values of X) minimal probability mass required to fall in the interval $(\gamma_{[r]}, \gamma_{[s]}]$ is $\sum_{k=1}^{K} \max(0, \bar{\rho}_{[d_{sk}]} - \bar{\rho}_{[a_{rk}]})$.

By way of example consider the arrangement (7) used before and the cases considered earlier.

1. $\gamma_{[1]} - \gamma_{[0]}$. The last-discharged indexes in row 1 are $\{1, 0, 0\}$ and the active indexes in row 0 are $\{0, 0, 0\}$. Only column 1 delivers a positive value and, noting that $\rho_{[0]} = 0$ there is:
$$\gamma_{[1]} - \gamma_{[0]} = \gamma_{[1]} \geq \bar{\rho}_{[1]} = \rho_{[1]}.$$

2. $\gamma_{[2]} - \gamma_{[0]}$. The last-discharged indexes in row 2 are $\{1, 2, 0\}$ and the active indexes in row 0 are $\{0, 0, 0\}$. Columns 1 and 2 deliver a positive value and, there is:
$$\gamma_{[2]} - \gamma_{[0]} = \gamma_{[2]} \geq \bar{\rho}_{[1]} + \bar{\rho}_{[2]} = \rho_{[1]} + \rho_{[2]}.$$

3. $\gamma_{[5]} - \gamma_{[3]}$. The last-discharged indexes in row 5 are $\{3, 2, 5\}$ and the active indexes in row 3 are $\{3, 6, 4\}$. Notice that $d_{52} = 2 < a_{32} = 6$ so column 2 produces no positive contribution. Only column 3 delivers a positive value and there is:
$$\gamma_{[5]} - \gamma_{[3]} \geq \bar{\rho}_{[5]} - \bar{\rho}_{[4]} = \rho_{[5]}.$$

4. $\gamma_{[7]} - \gamma_{[4]}$. The last-discharged indexes in row 7 are $\{7, 8, 9\}$ and the active indexes in row 4 are $\{7, 6, 4\}$. Columns 2 and 3 produce positive values and there is:
$$\gamma_{[7]} - \gamma_{[4]} = 1 - \gamma_{[4]} \geq \left(\bar{\rho}_{[8]} - \bar{\rho}_{[6]}\right) + \left(\bar{\rho}_{[9]} - \bar{\rho}_{[4]}\right) = \rho_{[8]} + \rho_{[5]} + \rho_{[9]}.$$

From the argument so far it follows that for every pair of indexes

$$(r, s) \in \{0, 1, \ldots, N+1\}$$

with $r < s$ the following inequality must hold if the value of $\boldsymbol{\gamma}$ in the arrangement under consideration is to allow the allocations of non-negative probability mass required to satisfy the observational equivalence restriction.

$$\gamma_{[s]} - \gamma_{[r]} \geq \sum_{k=1}^{K} \max(0, \bar{\rho}_{[d_{sk}]} - \bar{\rho}_{[a_{rk}]}) \tag{8}$$

This system of $(N+1)(N+2)/2$ inequalities defines a set of values of $\boldsymbol{\gamma}$ denoted by $\mathcal{E}_t(z)$ associated with arrangement t and instrumental value z.

All values of $\boldsymbol{\gamma}$ in $\mathcal{H}_t(z)$ must satisfy these inequalities, so $\mathcal{H}_t(z) \subseteq \mathcal{E}_t(z)$.

We can now give a formal statement regarding the convex components of the identified set of structural functions. At this point we make explicit in the notation the dependence of objects on the arrangement under consideration, t, and on the instrumental value, z.

THEOREM 1: *Consider an arrangement t of*

$$\boldsymbol{\gamma} \equiv [\gamma_{11}, \ldots, \gamma_{1K}, \gamma_{21}, \ldots, \gamma_{2K}, \ldots, \gamma_{M-1,1}, \ldots, \gamma_{M-1,K}]$$

with ith largest element $\gamma^t_{[n]}$ such that

$$0 \equiv \gamma^t_{[0]} \leq \gamma^t_{[1]} \leq \cdots \leq \gamma^t_{[N]} \leq \gamma^t_{[N+1]} \equiv 1$$

where $N \equiv K(M-1)$. The correspondence between elements of the ordered and unordered lists is given by arrangement-specific functions $m_t(\cdot)$, $k_t(\cdot)$ and an inverse function $n_t(\cdot,\cdot)$ which are such that for $n \in \{1,\ldots,N\}$, $k \in \{1,\ldots,K\}$ and $m \in \{1,\ldots,M-1\}$:

$$\begin{aligned} \gamma^t_{[n]} &= \gamma_{m_t(n)k_t(n)} \\ \gamma_{mk} &= \gamma^t_{[n_t(m,k)]}. \end{aligned}$$

For all arrangements define $m_t(0) \equiv 0$ and for all k: $m_t(N+k) \equiv M$, $k_t(N+k) \equiv k, n_t(M,k) \equiv N+k$, $n_t(0,k) \equiv 0$. For $n \in \{1,\ldots,N+K\}$ define:

$$\bar{\rho}^t_{[n]}(z) \equiv \Pr[Y \leq m_t(n) \wedge X = x_{k_t(n)} | Z = z]$$

with $\bar{\rho}^t_{[0]}(z) \equiv 0$. For all k and *$n \in \{1,\ldots,N\}$ define*

$$a^t_{nk}(z) \equiv \min\{i : (n \leq i \leq N+k) \wedge (k_t(i) = k)\}$$

$$d^t_{nk}(z) \equiv \max\{i : (0 \leq i \leq n) \wedge (k_t(i) = k)\}$$

and for all k define $d_{0k} \equiv 0$ and $d_{N+1,k} \equiv N+k$. Define a set of values of $\boldsymbol{\gamma}$, $\mathcal{E}_t(z)$, determined by the intersection of the following $(N+1)(N+2)/2$ linear half spaces:

$$\gamma^t_{[s]} - \gamma^t_{[r]} \geq \sum_{k=1}^{K} \max(0, \bar{\rho}^t_{[d_{sk}(z)]}(z) - \bar{\rho}^t_{[a_{rk}(z)]}(z)) \tag{9}$$

with $(r,s) \in \{0,1,\ldots,N+1\}$ and $s > r$.

Then:

1. *The set $\mathcal{H}_t(z)$ is a subset of $\mathcal{E}_t(z)$.*
2. *For every $\boldsymbol{\gamma} \in \mathcal{E}_t(z)$ there exists a distribution of U and X given $Z = z$ which is piecewise uniform for variations in U that:*
   (a) *is proper,*

(b) *satisfies the independence condition: for all* $i \in \{1, \ldots, N\}$

$$\sum_{j=1}^{i} \sum_{k=1}^{K} \eta_{[j]k}^{t}(z) = \gamma_{[i]}^{t}$$

(c) *delivers the probabilities* $\bar{\rho}_{[i]}^{t}(z)$ *for all* $i \in \{1, \ldots, N\}$ *and so satisfies the observational equivalence property.*

Result 1 of the Theorem has already been demonstrated to be true because we showed that all $\boldsymbol{\gamma} \in H_t(z)$ satisfy the inequalities that define the polytope $\mathcal{E}_t(z)$. It remains to show how to construct the distribution referred to in Result 2. That is the subject of Section 3.2. First two corollaries are stated and proved and simple upper and lower bounds on the elements of $\boldsymbol{\gamma}$ are derived.

COROLLARY 1: *For all* t *and* z, $\mathcal{E}_t(z) = \mathcal{H}_t(z)$.

PROOF: Result 1 of the Theorem states that $\mathcal{E}_t(z) \subseteq \mathcal{H}_t(z)$ and Result 2 implies that $\mathcal{H}_t(z) \subseteq \mathcal{E}_t(z)$, from which it follows that $\mathcal{E}_t(z) = \mathcal{H}_t(z)$.

COROLLARY 2: *The set of values of the structural function identified by the SEIV model is as follows.*

$$\mathcal{H}(\mathcal{Z}) = \bigcup_{t=1}^{T} \left(\bigcap_{z \in \mathcal{Z}} \mathcal{E}_t(z) \right)$$

PROOF: This follows directly from Corollary 1 on noting the composition of the identified set, $\mathcal{H}(\mathcal{Z})$.

$$\mathcal{H}(\mathcal{Z}) = \bigcup_{t=1}^{T} \left(\bigcap_{z \in \mathcal{Z}} \mathcal{H}_t(z) \right)$$

For a particular arrangement many of the inequalities defining a set $\mathcal{E}_t(z)$ will be redundant. The inequality given by setting $s = N+1$ and $r = 0$ is always redundant[15] so that there are at most $(N+1)(N+2)/2 - 1$ inequalities defining the polytope $\mathcal{E}_t(z)$ and often far fewer. This is investigated further in Section 4.

The identified set is determined by a large number of elementary inequalities which either place upper or lower bounds on elements of $\boldsymbol{\gamma}$ or lower bounds on differences of pairs of elements of $\boldsymbol{\gamma}$. The convex polytope within which identified values of $\boldsymbol{\gamma}$ lie in any particular arrangement is a facetted N-orthotope lying in the unit N-cube with all facets taken at angles of 45° to the faces of the unit N-cube.

3.1. Upper and lower bounds. The inequalities (9) deliver simple upper and lower bounds on elements of $\boldsymbol{\gamma}$ specific to an arrangement t and an instrumental value.

Suppressing dependence on the arrangement, t, and the instrumental value, z, and setting $r = 0$ in (8) and noting that $\gamma_{[0]} = 0$ and for all k, $a_{0k} = 0$ and $\bar{\rho}_{[0]} = 0$,

[15] In this case the inequality (8) is $1 \geq 1$.

there is for all $s \in \{1, \ldots, N+1\}$ the lower bound:

$$\gamma_{[s]} \geq \sum_{k=1}^{K} \bar{\rho}_{[d_{sk}]} \tag{10}$$

which can be expressed in terms of non-cumulative probabilities as follows.

$$\gamma_{[s]} \geq \sum_{i=1}^{s} \rho_{[i]}.$$

Setting $s = N+1$ in (8) and noting that $\gamma_{[N+1]} = 1$ there is for all $r \in \{1, \ldots, N+1\}$:

$$1 - \gamma_{[r]} \geq \sum_{k=1}^{K} \left(\delta_k - \bar{\rho}_{[a_{rk}]} \right)$$

and so the following upper bound.

$$\gamma_{[r]} \leq \sum_{k=1}^{K} \bar{\rho}_{[a_{rk}]} \tag{11}$$

These are the bounds given in Chesher (2007, 2010).

3.2. Construction of a joint distribution of U and X. We propose an algorithm for constructing a joint distribution for U and X given $Z = z$ for *any* value of $\boldsymbol{\gamma}$ that lies in a set $\mathcal{E}_t(z)$ constructed using a sequence of probabilities $\rho_{[1]}, \ldots, \rho_{[N]}$. We then prove Result 2 of Theorem 1 by showing that the distribution has the required properties, namely that it is proper, that it satisfies observational equivalence, delivering the probabilities $\rho_{[1]}, \ldots, \rho_{[N]}$ that determine $\mathcal{E}_t(z)$, and that it satisfies the independence restriction as expressed in (4).

While setting up notation and giving the details of the workings of the algorithm dependence of objects such as $\rho_{[n]}$, $\gamma_{]n]}$, $\eta_{[n]k}$, a_{nk} and d_{nk} on the arrangement under consideration and the instrumental value is suppressed in the notation.[16] In what follows sums from a to b, $\sum_{i=a}^{b}(\cdot)_i$, with $b < a$ are by convention equal to zero.

We have introduced the active index lists a_n and we now make use of *ordered* active index lists $a_n^o \equiv \{a_{n[1]}, \ldots, a_{n[K]}\}$ where:

$$\min\{a_{n1}, \ldots, a_{nK}\} \equiv a_{n[1]} < a_{n[2]} < \cdots < a_{n[K]} \equiv \max\{a_{n1}, \ldots, a_{nK}\}.$$

Note that for all $n \in \{1, \ldots, N+1\}$, $a_{n[1]} = n$. The ordered active index list for the

[16] Of course the IV restriction ensures that $\gamma_{[n]}$ does not vary with the instrumental value z.

arrangement (7) is as follows.

n	a_n^o
0	$\{0,0,0\}$
1	$\{1,2,4\}$
2	$\{2,3,4\}$
3	$\{3,4,6\}$
4	$\{4,6,7\}$
5	$\{5,6,7\}$
6	$\{6,7,9\}$
7	$\{7,8,9\}$

For i passing through the sequence: $1,2,\ldots,N+1$ and, at each value of i, for j passing through the ascending sequence: $1,2,\ldots,K$ the algorithm produces values $\eta_{[i]k}$ calculated recursively as follows:

$$\eta(i,j) = \min(G(i,j), \max(0, R(i,j))). \tag{12}$$

with component objects defined as follows.

$$\eta(i,j) \equiv \eta_{[i]k(a_{i[j]})}$$

$$G(i,j) \equiv \Delta\gamma_{[i]} - \sum_{j'=1}^{j-1} \eta(i,j')$$

$$R(i,j) \equiv \rho_{[a_{i[j]}]} - \sum_{i'=n(m(a_{i[j]})-1,k(a_{i[j]}))+1}^{i-1} \eta_{[i']k(a_{i[j]})}$$

It will shortly be shown that for every value of $\boldsymbol{\gamma}$ that lies in the set $\mathcal{E}_t(z)$ this algorithm fills the cells of a table with probability masses which are (i) non-negative while (ii) delivering total masses within groups of cells that respect the observational equivalence restriction and (iii) allocating a total mass of exactly $\Delta\gamma_{[i]}$ in row i for each $i \in \{1,\ldots,N+1\}$.

In each row $G(i,1)$ has the value $\Delta\gamma_{[i]}$ and for $j \in \{2,\ldots,K+1\}$ there is the following recursion.

$$G(i,j) = G(i,j-1) - \eta(i,j-1) \tag{13}$$

Note that

$$G(i,K+1) = \Delta\gamma_{[i]} - \sum_{j=1}^{K} \eta(i,j) = \Delta\gamma_{[i]} - \sum_{k=1}^{K} \eta_{[i]k}$$

so when $G(i,K+1) = 0$ for all i the algorithm delivers probabilities that satisfy the independence restriction.

In each row i the first cell to be addressed is the one with the smallest active index in that row. This is the cell that completes in row i the block of cells in column $k(i)$ which must contain probability mass $\rho_{[i]}$ if the observational equivalence condition is to be satisfied. $R(i,1)$ is the mass to be allocated to that cell to bring the total to $\rho_{[i]}$. The next cell to be addressed is the one in the column $k(a_{i[2]})$ corresponding

to the next active index to be discharged. Up to an amount $R(i,2)$ is allocated in this cell. This is the mass which, if allocated to that cell, would bring the mass in the block of cells in which the cell appears up to $\rho_{[a_{i[2]}]}$. The process proceeds with j incrementing until $j = K$ with probability mass being allocated to cells until all the mass $\Delta\gamma_{[i]}$ has been allocated after which (as will be shown) zero values appear in the cells of row i.

We now show that when $\boldsymbol{\gamma}$ lies in the set $\mathcal{E}_t(z)$ the algorithm delivers probability masses in each cell such that the *observational equivalence* condition is satisfied. Then we prove that the independence and properness conditions are satisfied.

If $\boldsymbol{\gamma}$ lies in the set $\mathcal{E}_t(z)$ then *by construction* there is sufficient probability mass available between every pair of values $\gamma_{[r]}$ and $\gamma_{[s]}$ to permit the allocation of probability masses $\rho_{[1]}, \ldots, \rho_{[N]}$ in their appropriate locations. The probability mass $\rho_{[i]}$ is equal to $\Pr[Y = m(i) \wedge X = x_{k(i)} | Z = z]$. The cells in which this mass must be allocated lie in the column associated with $x_{k(i)}$ and terminate in row i. They start in the row given by $n(m(i)-1, k(i)) + 1$. The observational equivalence condition is therefore: for all $i \in \{1, \ldots, N\}$

$$\sum_{t=n(m(i)-1,k(i))+1}^{i} \eta_{[t]k(i)} = \rho_{[i]}.$$

The proposed algorithm fill blocks of cells in index order, $\rho_{[1]}$ first, $\rho_{[2]}$ second and so on. At each step of the process the algorithm allocates as much probability mass as possible to the blocks of cells associated with probabilities $\rho_{[i]}$ which have the lowest values of i accessible at that point.[17] The algorithm delivers the required allocations of probability mass for values of $\boldsymbol{\gamma}$ that lie in $\mathcal{E}_t(z)$ because such values have elements that are sufficiently separated to permit the required allocation of probability masses.

It is shown in the Proposition below that *whether or not* the value of $\boldsymbol{\gamma}$ lies in $\mathcal{E}_t(z)$ the algorithm (i) allocates non-negative probability mass in every cell and (ii) never allocates more than an amount $\Delta\gamma_{[i]}$ in row i, for $i \in \{1, \ldots, N\}$. When the value of $\boldsymbol{\gamma}$ lies in $\mathcal{E}_t(z)$, a total mass of 1 is allocated by the algorithm because the observational equivalence condition is satisfied. Since, as shown below, an allocation exceeding $\Delta\gamma_{[i]}$ cannot occur for any i and $\sum_{i=1}^{N+1} \Delta\gamma_{[i]} = 1$, when the value of $\boldsymbol{\gamma}$ lies in $\mathcal{E}_t(z)$ the algorithm must place a probability mass *exactly* equal to $\Delta\gamma_{[i]}$ in each row $i \in \{1, \ldots, N\}$, thus satisfying the independence condition. The properness conditions is satisfied because (i) all probability masses allocated are non-negative and (ii) since the observational equivalence condition is satisfied a total mass of 1 is allocated.

Here is the Proposition setting out some properties of the algorithm. These obtain *whether or not* $\boldsymbol{\gamma} \in \mathcal{E}_t(z)$.

Proposition 1:

1. *For all i and j*:

   (a) $\eta(i,j) \leq G(i,j)$,

[17] This occurs because the algorithm uses *ordered* active indexes.

(b) $G(i,j) \geq 0$,

(c) $\eta(i,j) \geq 0$,

(d) $G(i,j)$ *is a non-increasing function of* j.

(e) *If for some* j, $G(i,j) > \eta(i,j)$ *then for all* $j' \leq j$, $G(i,j') > \eta(i,j')$ *and* $\eta(i,j') = \max(0, R(i,j'))$.

2. *The algorithm allocates a probability mass of at most* $\Delta\gamma_{[i]}$ *in row* i, *that is:*

$$\sum_{k=1}^{K} \eta_{[i]k} \leq \Delta\gamma_{[i]}.$$

The proof is in Annex 1.

3.3. An alternative derivation of the set $\mathcal{E}_t(z)$. In order to relate the inequalities that define the set $\mathcal{E}_t(z)$ to the inequalities given in Chesher (2010) and Chesher and Smolinski (2009) it is useful to give an alternative derivation and expression for the set $\mathcal{E}_t(z)$.

Associated with the lists of active and last-discharged indexes there are arrays of cumulative probabilities which are useful in the subsequent analysis. They also provide an alternative characterisation of the set $\mathcal{E}_t(z)$.

Consider the ith largest element $\gamma_{[i]}$. If this lies in the identified set then for each $i \in \{1, \ldots, N\}$ there exist non-negative values of the *cumulative* probabilities $\bar{\eta}_{[i]k}$ which sum to $\gamma_{[i]}$ across $k \in \{1, \ldots, K\}$. The set $\mathcal{E}_t(z)$ is derived by finding lower and upper bounds for each term $\bar{\eta}_{[i]k}$ in the sum, producing bounds on differences of elements of $\boldsymbol{\gamma}$ by combining the bounds.

Each cumulative probability $\bar{\eta}_{[i]k}$ is bounded *below* by the *maximum* of the terms $\bar{\rho}_{[j]}$ that appear in rows 1 through i of the column associated with x_k. That bound is $\lambda_{ik} \equiv \bar{\rho}_{[d_{ik}]}$ where d_{ik} is the last-discharged index in column k of row i.

Each term $\bar{\eta}_{[i]k}$ is bounded *above* by the *minimum* of the terms $\bar{\rho}_{[j]}$ that appear in rows i through $N+1$ of column k. That bound is $\pi_{ik} \equiv \bar{\rho}_{[a_{ik}]}$ where a_{ik} is the active index in column k of row i.

Combining results there are the following bounds for all i and k:

$$\lambda_{ik} \leq \bar{\eta}_{[i]k} \leq \pi_{ik} \tag{14}$$

and on summing and noting that for $\boldsymbol{\gamma}$ in the identified set the independence condition holds so that $\gamma_{[i]} = \sum_{k=1}^{K} \bar{\eta}_{[i]k}$ there are the following lower and upper bounds.

$$\lambda_i \equiv \sum_{k=1}^{K} \lambda_{ik} \leq \gamma_{[i]} \leq \sum_{k=1}^{K} \pi_{ik} \equiv \pi_i \tag{15}$$

Making explicit dependence on the arrangement under consideration, t, and the instrumental value, z, defining

$$\lambda_{ik}^t(z) \equiv \bar{\rho}_{[d_{ik}^t]}^t(z) \qquad \pi_{ik}^t(z) \equiv \bar{\rho}_{[a_{ik}^t]}(z)$$

$$\lambda_i^t(z) \equiv \sum_{k=1}^{K} \lambda_{ik}^t(z) \qquad \pi_i^t(z) \equiv \sum_{k=1}^{K} \pi_{ik}^t(z)$$

and intersecting the bounds (20) across $z \in \mathcal{Z}$ gives the following inequalities which hold for each arrangement t and for all $i \in \{1, \ldots, N\}$.

$$\max_{z \in \mathcal{Z}} \left(\lambda_i^t(z)\right) \leq \gamma_{[i]} \leq \min_{z \in \mathcal{Z}} \left(\pi_i^t(z)\right) \tag{16}$$

The inequalities (14) can also be used to place bounds on *differences*, $\gamma_{[s]} - \gamma_{[r]}$, as follows. For all s and r in $\{0, \ldots, N+1\}$ and for all k there are bounds on $\bar{\eta}_{[s]k}$ and on $-\bar{\eta}_{[r]k}$ as follows:

$$\lambda_{sk} \leq \bar{\eta}_{[s]k} \leq \pi_{sk}$$

$$-\pi_{rk} \leq -\bar{\eta}_{[r]k} \leq -\lambda_{rk}$$

and on adding there are the following bounds.

$$\lambda_{sk} - \pi_{rk} \leq \bar{\eta}_{[s]k} - \bar{\eta}_{[r]k} \leq \pi_{sk} - \lambda_{rk} \tag{17}$$

Summing across k there are the following inequalities.

$$\lambda_s - \pi_r = \sum_{k=1}^{K} (\lambda_{sk} - \pi_{rk}) \leq \gamma_{[s]} - \gamma_{[r]} \leq \sum_{k=1}^{K} (\pi_{sk} - \lambda_{rk}) = \pi_s - \lambda_r \tag{18}$$

This is nothing more than a direct implication of (15) but the lower bound here *can be improved upon* by exploiting the properness condition [2]. Thus, consider values s and r such that $s > r$. If $\boldsymbol{\gamma}$ is in the identified set then for all $s > r$ the inequality $\bar{\eta}_{[s]k} - \bar{\eta}_{[r]k} \geq 0$ holds. The lower bound in (17) can therefore be tightened as follows.

$$\max(0, \lambda_{sk} - \pi_{rk}) \leq \bar{\eta}_{[s]k} - \bar{\eta}_{[r]k} \leq \pi_{sk} - \lambda_{rk} \tag{19}$$

Summing across k gives the following bounds which hold for all $s > r \in \{0, \ldots, N+1\}$.

$$\sum_{k=1}^{K} \max(0, \lambda_{sk} - \pi_{rk}) \leq \gamma_{[s]} - \gamma_{[r]} \leq \pi_s - \lambda_r \tag{20}$$

The set defined by these bounds is precisely the set $\mathcal{E}_t(z)$.

Making explicit the dependence of the terms in these bounds on the arrangement under consideration, t, and the instrumental value z, and intersecting the bounds (20) across $z \in \mathcal{Z}$ gives the following inequalities which hold for each arrangement t and for all $N + 1 \geq s \geq r \geq 0$.

$$\underline{\phi}_{sr}^t(\mathcal{Z}) \equiv \max_{z \in \mathcal{Z}} \left(\sum_{k=1}^{K} \max(0, \lambda_{sk}^t(z) - \pi_{rk}^t(z)) \right) \leq \gamma_{[s]}^t - \gamma_{[r]}^t \leq \min_{z \in \mathcal{Z}} \left(\pi_s^t(z) - \lambda_r^t(z)\right) \equiv \bar{\phi}_{sr}^t(\mathcal{Z}) \tag{21}$$

These bounds define $\mathcal{E}_t(\mathcal{Z})$, the component of the identified set in which $\boldsymbol{\gamma}$ is in arrangement t. The union of these sets, $\bigcup_{t=1}^{T} \mathcal{E}_t(\mathcal{Z})$, is the set $\mathcal{E}(\mathcal{Z})$, previously defined, which is equal to the identified set $\mathcal{H}(\mathcal{Z})$.

	Value of X						Ordered		
n	x_1	x_2	x_3	x_4	x_5	x_6	values of $\boldsymbol{\gamma}$	$\boldsymbol{\gamma}$	$k(n)$
0	0	0	0	0	0	0	$\gamma_{[0]}$	0	
1	$\bar{\rho}_{[1]}$	$\eta_{[1]2}$	$\eta_{[1]3}$	$\eta_{[1]4}$	$\eta_{[1]5}$	$\eta_{[1]6}$	$\gamma_{[1]}$	γ_{11}	1
2	$\eta_{[2]1}$	$\bar{\rho}_{[2]}$	$\eta_{[2]3}$	$\eta_{[2]4}$	$\eta_{[2]5}$	$\eta_{[2]6}$	$\gamma_{[2]}$	γ_{12}	2
3	$\eta_{[3]1}$	$\eta_{[3]2}$	$\bar{\rho}_{[3]}$	$\eta_{[3]4}$	$\eta_{[3]5}$	$\eta_{[3]6}$	$\gamma_{[3]}$	γ_{13}	3
4	$\eta_{[4]1}$	$\eta_{[4]2}$	$\eta_{[4]3}$	$\bar{\rho}_{[4]}$	$\eta_{[4]5}$	$\eta_{[4]6}$	$\gamma_{[4]}$	γ_{14}	4
5	$\eta_{[5]1}$	$\eta_{[5]2}$	$\eta_{[5]3}$	$\eta_{[5]4}$	$\bar{\rho}_{[5]}$	$\eta_{[5]6}$	$\gamma_{[5]}$	γ_{15}	5
6	$\eta_{[6]1}$	$\eta_{[6]2}$	$\eta_{[6]3}$	$\eta_{[6]4}$	$\eta_{[6]5}$	$\bar{\rho}_{[6]}$	$\gamma_{[6]}$	γ_{16}	6
7	δ_1	δ_2	δ_3	δ_4	δ_5	δ_6	$\gamma_{[7]}$	1	

Table 4: Conditional distribution-mass function values for a binary outcome example with observational equivalence restrictions imposed.

3.4. Binary outcomes. When Y is binary there is just one threshold function and the parameters of interest are $\gamma_{11}, \gamma_{12}, \ldots, \gamma_{1K}$. We now show that in this case the lower bound in (21) is zero when $a > 0$ so these bounds place no restrictions on $\boldsymbol{\gamma}$ additional to those defined by (16).

Without loss of generality we consider an arrangement t in which the elements of $\boldsymbol{\gamma}$ are arranged in the order of the index k. The situation for a case in which $K = 6$ is as pictured in Table 4. Notice that with the given arrangement of $\boldsymbol{\gamma}$ for every value of n, $k(n) = n$, so the values $\bar{\rho}_{[n]}$ lie on the diagonal in Table 4. Because Y is binary there is only one such entry in each column.

We now show that for all indices $s > r > 0$ the terms $\lambda_{sk} - \pi_{rk}$ are zero or negative for all k from which it follows that the lower bound in (21) is zero.

Consider some value k and the difference $\lambda_{sk} - \pi_{rk}$ with $s \geq r$. Referring to Table 4 it can be seen that values taken by λ_{sk} and π_{rk} are as follows.

$$\lambda_{sk} = \begin{cases} 0 & , \quad s < k \\ \bar{\rho}_{[k]} & , \quad s \geq k \end{cases} \qquad \pi_{rk} = \begin{cases} \bar{\rho}_{[k]} & , \quad r \leq k \\ \delta_k & , \quad r > k \end{cases}$$

The resulting values of $\lambda_{sk} - \pi_{rk}$ are therefore as shown below.

Values of $\lambda_{sk} - \pi_{rk}$

	$s < k$	$s = k$	$s > k$
$r < k$	$-\bar{\rho}_{[k]}$	$-\bar{\rho}_{[k]}$	0
$r = k$	$*$	0	0
$r > k$	$*$	$*$	$\bar{\rho}_{[k]} - \delta_k$

All the values are zero or negative and the result is that the lower bounds $\underline{\phi}^t_{sr}(\mathcal{Z})$ are zero. Therefore in the binary Y case the restrictions imposed by the bounds (21) for $r \neq 0$ have no force. It is shown in the next Section that the bounds obtained from (21) setting $r = 0$, equivalently the bounds (16), are identical to the bounds given in Chesher (2009, 2010) which are shown in those papers to define the identified set $\mathcal{H}(\mathcal{Z})$.

3.5. Relationship to earlier results. It is shown in Chesher (2010) that all structural functions h that lie in the set identified by a SEIV model given a particular probability distribution $F_{YX|Z}$ with $Z = z \in \mathcal{Z}$ satisfy the following inequalities for all $\tau \in (0,1)$.

$$\max_{z\in\mathcal{Z}} \Pr[Y < h(X,\tau)|Z=z] < \tau \leq \min_{z\in\mathcal{Z}} \Pr[Y \leq h(X,\tau)|Z=z] \tag{22}$$

Here probabilities are calculated using the distribution $F_{YX|Z}$.

The inequalities generated by (22) as τ varies over $(0,1)$ define a set of structural functions referred to as $\mathcal{C}(\mathcal{Z})$ in Chesher and Smolinski (2009). When X is discrete and characterized by a vector $\boldsymbol{\gamma}$ as in the previous discussion the set $\mathcal{C}(\mathcal{Z})$ is a union of convex sets, $\mathcal{C}_t(\mathcal{Z})$, one associated with each arrangement, t, of $\boldsymbol{\gamma}$.

$$\mathcal{C}(\mathcal{Z}) = \bigcup_{t=1}^{T} \mathcal{C}_t(\mathcal{Z})$$

Each set $\mathcal{C}_t(\mathcal{Z})$ is an intersection of sets obtained as z varies within $\mathcal{Z}$.

$$\mathcal{C}_t(\mathcal{Z}) = \bigcap_{z\in\mathcal{Z}} \mathcal{C}_t(z)$$

We now show that the bounds (22) are identical to those generated by (16) as n varies over $\{1,\ldots,N\}$. Chesher and Smolinski (2009) show that in the discrete endogenous variable case considered here the bounds (22) hold for all $\tau \in (0,1)$ if and only if the following inequalities hold for all $l \in \{1,\ldots,M-1\}$ and $s \in \{1,\ldots,K\}$.

$$\max_{z\in\mathcal{Z}} \sum_{k=1}^{K}\sum_{m=1}^{M-1} \rho_{mk}(z) 1(\gamma_{mk} \leq \gamma_{ls}) \leq \gamma_{ls} \leq \min_{z\in\mathcal{Z}} \sum_{k=1}^{K}\sum_{m=1}^{M} \rho_{mk}(z) 1(\gamma_{m-1,k} < \gamma_{ls})$$

Consider a particular arrangement of $\boldsymbol{\gamma}$ and its nth largest element, $\gamma_{[n]}$. Substituting $\gamma_{[n]}$ for γ_{ls} above gives the following.

$$\max_{z\in\mathcal{Z}} \sum_{k=1}^{K}\sum_{m=1}^{M-1} \rho_{mk}(z) 1(\gamma_{mk} \leq \gamma_{[n]}) \leq \gamma_{[n]} \leq \min_{z\in\mathcal{Z}} \sum_{k=1}^{K}\sum_{m=1}^{M} \rho_{mk}(z) 1(\gamma_{m-1,k} < \gamma_{[n]}) \tag{23}$$

Comparing this with (16) it can be concluded that the bounds are identical because both of the following equations are satisfied:

$$\bar{\rho}_{[d_{nk}]}(z) \equiv \lambda_{nk}(z) = \sum_{m=1}^{M-1} \rho_{mk}(z) 1(\gamma_{mk} \leq \gamma_{[n]}) \tag{24}$$

$$\bar{\rho}_{[a_{nk}]}(z) \equiv \pi_{nk}(z) = \sum_{m=1}^{M} \rho_{mk}(z) 1(\gamma_{m-1,k} < \gamma_{[n]}) \tag{25}$$

and on the right hand side of (24) and (25) are the expressions summed over k to produce the bounds in (23).

3.6. Alternative expressions for the bounds. The objects $\lambda_{nk}(z)$ and $\pi_{nk}(z)$ can be expressed in terms of probabilities involving the structural function as follows.

$$\lambda_{nk}(z) = \Pr[Y < h(x_k, \gamma_{[n+1]}) \wedge X = x_k | Z = z] \tag{26}$$

$$\pi_{nk}(z) = \Pr[Y \leq h(x_k, \gamma_{[n]}) \wedge X = x_k | Z = z] \tag{27}$$

Here, to accommodate the case $n = N + 1$, we adopt the convention that, for all k, $h(x_k, \gamma_{[N+2]})$ denotes a quantity exceeding 1 which ensures that, for all k, (26) delivers $\lambda_{N+1,k} = \bar{\rho}_{Mk} = \delta_k$.[18] The definitions already in hand ensure that for all k, (27) delivers $\pi_{N+1,k} = \bar{\rho}_{Mk} = \delta_k$.

With these expressions in hand the bounds (16) can be written as follows:

$$\lambda_n(\mathcal{Z}) \equiv \max_{z \in \mathcal{Z}} \Big(\Pr[Y < h(X, \gamma_{[n+1]}) | Z = z] \Big) \leq \gamma_{[n]} \leq \min_{z \in \mathcal{Z}} \Big(\Pr[Y \leq h(X, \gamma_{[n]}) | Z = z] \Big) \equiv \pi_n(\mathcal{Z})$$

and the term

$$\sum_{k=1}^{K} \max(0, \lambda^t_{sk}(z) - \pi^t_{rk}(z))$$

in the expression (21) can be written as follows.

$$\sum_k \max \Big(0, \Big\{ \Pr[Y < h(x_k, \gamma_{[s+1]}) \wedge X = x_k) | Z = z] - \Pr[Y \leq h(x_k, \gamma_{[r]}) \wedge X = x_k | Z = z] \Big\} \Big)$$

Table 5 shows the values of λ_{nk} and π_{nk} in the arrangement used in the example considered earlier in the paper. Dependence on the value z is no longer made explicit in the notation.

Table 6 shows the value of the structural function $h(x, u)$ for all the combinations of x and u that arise in (26) and (27) in this example. For example the entry for $n = 3$ (row) and $k = 2$ under the heading π_{n2} is $h(x_2, \gamma_{[3]}) = h(x_2, \gamma_{21}) = 2$. The entries in this Table are easily verified by referring to Figure 1.

Consider for example λ_{42}. From (26) we have

$$\lambda_{42} = \Pr[Y \leq h(x_2, \gamma_{[5]}) \wedge X = x_2 | Z = z]$$

and since $\gamma_{[5]} = \gamma_{23}$ there is, from Table 6, $h(x_2, \gamma_{23}) = 2$. Accordingly

$$\lambda_{42} = \Pr[Y < 1 \wedge X = x_2 | Z = z]$$

which is equal to $\bar{\rho}_{12}$ as shown in Table 5 in the entry for $n = 4$ and $k = 2$.

Consider for example π_{33}. From (27) we have

$$\pi_{33} = \Pr[Y \leq h(x_3, \gamma_{[3]}) \wedge X = x_3 | Z = z]$$

and since $\gamma_{[3]} = \gamma_{21}$ there is, from Table 6, $h(x_3, \gamma_{21}) = 1$. Accordingly

$$\pi_{33} = \Pr[Y \leq 1 \wedge X = x_3 | Z = z]$$

which is equal to $\bar{\rho}_{13}$ as shown in Table 5 in the entry for $n = 3$ and $k = 3$.

[18] Equivalently we can define $\lambda_{N+1,k}(z)$ as follows:

$$\lambda_{N+1,k}(z) = \Pr[Y \leq h(x_k, \gamma_{[N+1]}) \wedge X = x_k | Z = z]$$

leaving equation (26) to apply for $n \in \{0, 1, \ldots, N\}$.

			$k=1$		$k=2$		$k=3$	
n	$\gamma_{(n)}$	$\gamma_{(n+1)}$	λ_{n1}	π_{n1}	λ_{n2}	π_{n2}	λ_{n3}	π_{n3}
0	0	γ_{11}	0	$\bar{\rho}_{11}$	0	$\bar{\rho}_{12}$	0	$\bar{\rho}_{13}$
1	γ_{11}	γ_{12}	$\bar{\rho}_{11}$	$\bar{\rho}_{11}$	0	$\bar{\rho}_{12}$	0	$\bar{\rho}_{13}$
2	γ_{12}	γ_{21}	$\bar{\rho}_{11}$	$\bar{\rho}_{21}$	$\bar{\rho}_{12}$	$\bar{\rho}_{12}$	0	$\bar{\rho}_{13}$
3	γ_{21}	γ_{13}	$\bar{\rho}_{21}$	$\bar{\rho}_{21}$	$\bar{\rho}_{12}$	$\bar{\rho}_{22}$	0	$\bar{\rho}_{13}$
4	γ_{13}	γ_{23}	$\bar{\rho}_{21}$	$\bar{\rho}_{31}$	$\bar{\rho}_{12}$	$\bar{\rho}_{22}$	$\bar{\rho}_{13}$	$\bar{\rho}_{13}$
5	γ_{23}	γ_{22}	$\bar{\rho}_{21}$	$\bar{\rho}_{31}$	$\bar{\rho}_{12}$	$\bar{\rho}_{22}$	$\bar{\rho}_{23}$	$\bar{\rho}_{23}$
6	γ_{22}	1	$\bar{\rho}_{21}$	$\bar{\rho}_{31}$	$\bar{\rho}_{22}$	$\bar{\rho}_{22}$	$\bar{\rho}_{23}$	$\bar{\rho}_{33}$
7	1		$\bar{\rho}_{31}$	$\bar{\rho}_{31}$	$\bar{\rho}_{32}$	$\bar{\rho}_{32}$	$\bar{\rho}_{33}$	$\bar{\rho}_{33}$

Table 5: Values of λ_{nk} and π_{nk} in the arrangement used in the example.

			$k=1$		$k=2$		$k=3$	
n	$\gamma_{(n)}$	$\gamma_{(n+1)}$	A: λ_{n1}	B: π_{n1}	A: λ_{n2}	B: π_{n2}	A: λ_{n3}	B: π_{n3}
0	0	γ_{11}	1	1	1	1	1	1
1	γ_{11}	γ_{12}	2	1	1	1	1	1
2	γ_{12}	γ_{21}	2	2	2	1	1	1
3	γ_{21}	γ_{13}	3	2	2	2	1	1
4	γ_{13}	γ_{23}	3	3	2	2	2	1
5	γ_{23}	γ_{22}	3	3	2	2	3	2
6	γ_{22}	1	3	3	3	2	3	3
7	1		3	3	3	3	3	3

Table 6: For the arrangement used in the example these are the values of A: $h(x_k, \gamma_{[n+1]})$ appearing in the definition of λ_{nk} and of B: $h(x_k, \gamma_{[n]})$ appearing in the definition of π_{nk}.

4. ILLUSTRATIVE CALCULATIONS, COMPUTATION AND ESTIMATION

4.1. Examples of bounds. We enumerate the bounds for a case with $M = 3$ and $K = 3$ and the arrangement of $\boldsymbol{\gamma}$ shown in equation (7) that has been considered throughout the paper.

Table 7 shows the values of lower bounds on $\gamma_{[s]} - \gamma_{[r]}$ for s (columns) and r (rows) varying in $\{1, \ldots, 7\}$. For example the entry in the row for $\gamma_{[1]}$ and the column for $\gamma_{[3]}$ gives the bound

$$\gamma_{[3]} - \gamma_{[1]} \geq \rho_{[2]}$$

that is

$$\gamma_{21} - \gamma_{11} \geq \rho_{21}$$

and note that this must hold for all $z \in \mathcal{Z}$. As z varies ρ_{21} varies and making this dependence explicit and dependence on the arrangement explicit too there is the bound

$$\gamma_{21} - \gamma_{11} \geq \max_{z \in \mathcal{Z}} \rho^t_{21}(z)$$

which contributes to the bounds defining $\mathcal{E}_t(\mathcal{Z})$.

The model places no restrictions on some of the differences other than those arising because of the ordering in the arrangement under consideration. An example is $\gamma_{[4]} - \gamma_{[2]}$ which is only required to be non-negative. Some of the restrictions that define a set $\mathcal{E}_t(\mathcal{Z})$ render others redundant. For example in Table 7 there is the restriction

$$\gamma_{[5]} - \gamma_{[4]} \geq \rho_{[5]} \tag{28}$$

which when satisfied ensures that two other restrictions are satisfied as follows.

$$\begin{aligned} \gamma_{[5]} - \gamma_{[3]} &\geq \rho_{[5]} \\ \gamma_{[5]} - \gamma_{[2]} &\geq \rho_{[5]} \end{aligned}$$

The restrictions $\gamma_{[6]} - \gamma_{[3]} \geq \rho_{[5]}$ and $\gamma_{[6]} - \gamma_{[4]} \geq \rho_{[5]}$ are also redundant, both being implied by the restriction (28).

In the final column lies $\gamma_{[7]}$ which is equal to 1. The entries in this column give lower bounds on $1 - \gamma_{[r]}$ where r varies from 1 to 6 down the rows of the Table. Subtracting these entries from 1 (i.e. eliminating the leading unit terms and changing the signs of what remains) delivers *upper* bounds on $\gamma_{[r]}$ for $r \in \{1, \ldots, 6\}$.

Lower bounds on the $\gamma_{[r]}$'s are simply, for each $r \in \{1, \ldots, 6\}$

$$\sum_{r'=1}^{r} \rho_{[r']} \leq \gamma_{[r]}$$

as shown in Section 3.1.

Adding the negative of the upper bound for $\gamma_{[r]}$ to the lower bound for $\gamma_{[s]}$ delivers a lower bound on $\gamma_{[s]} - \gamma_{[r]}$ which we can compare with the bounds shown in Table 7. Doing this we find that the lower bounds on $\gamma_{[4]} - \gamma_{[1]}$, $\gamma_{[5]} - \gamma_{[1]}$, $\gamma_{[6]} - \gamma_{[1]}$ and $\gamma_{[6]} - \gamma_{[2]}$ in Table 7 are exactly the bounds obtained by comparing lower and upper bounds on individual elements of $\boldsymbol{\gamma}$.

The only inequality in Table 7 that survives these various eliminations is $\gamma_{[5]} - \gamma_{[4]} \geq \rho_{[5]}$. So for this arrangement the set $\mathcal{E}_t(\mathcal{Z})$ is defined by this inequality and the lower and upper bounds on the individual elements of $\boldsymbol{\gamma}$ and the inequalities that express the ordering of the elements of $\boldsymbol{\gamma}$ in this arrangement.

In the $M = 3$, $K = 3$ example considered in detail in this paper there are 90 admissible arrangements of $\boldsymbol{\gamma}$ of which 15 are fundamental in the sense that each of these 15 generates $3! = 6$ arrangements by permuting the index identifying the three values of the conditioning variable. Annex 2 shows the bounds on $\gamma_{[s]} - \gamma_{[r]}$ just as in Table 7 for each of these 15 fundamental arrangements. In the sequence presented there the arrangement considered in this Section is number 8.

Comparisons amongst the inequalities on differences of elements of $\boldsymbol{\gamma}$ and comparing those inequalities with the implications of the lower and upper bounds on elements of $\boldsymbol{\gamma}$ leads to elimination of large numbers of the entries in the tables that refer to differences $\gamma_{[s]} - \gamma_{[r]}$ for s and r in $\{1, \ldots, 6\}$. In Arrangement 1 all such inequalities on differences disappear. In Arrangement 2 only the inequality $\gamma_{[5]} - \gamma_{[3]} \geq \rho_{[5]}$ remains. In Arrangement 3 only the inequality $\gamma_{[4]} - \gamma_{[2]} \geq \rho_{[4]}$ remains. In other cases there are more survivors. For example in Arrangement 11 the following three inequalities on differences of elements of γ survive.

$$\begin{aligned} \gamma_{[3]} - \gamma_{[2]} &\geq \rho_{[3]} \\ \gamma_{[5]} - \gamma_{[4]} &\geq \rho_{[5]} \\ \gamma_{[6]} - \gamma_{[2]} &\geq \rho_{[5]} + \rho_{[6]} \end{aligned}$$

In the example considered here the number of discrete outcomes is $M = 3$ and the number of points of support of the endogenous variables is $K = 3$. When M or K are larger there are many more contributions to the definitions of sets $\mathcal{E}_t(\mathcal{Z})$ coming from inequalities involving differences of elements of $\boldsymbol{\gamma}$.

4.2. A `Mathematica` notebook. This paper is accompanied by a `Mathematica` notebook which is viewable in the freeware `Math Player 7`.[19] The notebook does symbolic calculation of bounds as set out in Table 7. The user provides values for M, the number of discrete outcomes and K the number of points of support of the endogenous variables. A stylised graphical display of the $M - 1$ threshold functions appears with the values associated with the K points of support of an endogenous variable X highlighted. The user can manipulate these thereby generating particular arrangements of $\boldsymbol{\gamma}$. For each arrangement t selected, the notebook produces a table like Table 7 showing in symbolic form the inequalities defining a set $\mathcal{E}_t(z)$.

4.3. Computation and estimation. When M and K are both large computation of the set $\mathcal{E}(\mathcal{Z})$ is challenging because of the large number of potential arrangements of $\boldsymbol{\gamma}$, that is of the K values of the $M - 1$ threshold functions, that may arise. For example when $M = K = 4$ there are 369,600 admissible arrangements rising to over 300 billion when $M = K = 5$. *Shape restrictions* are helpful in reducing the scale of the problem.

[19] The notebook can be downloaded from www.cemmap.ac.uk/wps/seiv.nb. `Math Reader 7` is available at: http://www.wolfram.com/products/player/download.cgi .

		γ_{11}	γ_{12}	γ_{21}	γ_{13}	γ_{23}	γ_{22}	1
		$\gamma_{[1]}$	$\gamma_{[2]}$	$\gamma_{[3]}$	$\gamma_{[4]}$	$\gamma_{[5]}$	$\gamma_{[6]}$	$\gamma_{[7]}$
γ_{11}	$\gamma_{[1]}$	$\cdot$	0	$\rho_{[2]}$	$\rho_{[3]}$	$\rho_{[3]}+\rho_{[5]}$	$\rho_{[3]}+\rho_{[5]}+\rho_{[6]}$	$1-\rho_{[1]}-\rho_{[2]}-\rho_{[4]}$
γ_{12}	$\gamma_{[2]}$	$\cdot$	$\cdot$	0	0	$\rho_{[5]}$	$\rho_{[5]}+\rho_{[6]}$	$1-\rho_{[1]}-\rho_{[2]}-\rho_{[3]}-\rho_{[4]}$
γ_{21}	$\gamma_{[3]}$	$\cdot$	$\cdot$	$\cdot$	0	$\rho_{[5]}$	$\rho_{[5]}$	$1-\rho_{[1]}-\rho_{[2]}-\rho_{[3]}-\rho_{[4]}-\rho_{[6]}$
γ_{13}	$\gamma_{[4]}$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\rho_{[5]}$	$\rho_{[5]}$	$1-\delta_1-\rho_{[2]}-\rho_{[4]}-\rho_{[6]}$
γ_{23}	$\gamma_{[5]}$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	0	$1-\delta_1-\rho_{[2]}-\rho_{[4]}-\rho_{[5]}-\rho_{[6]}$
γ_{22}	$\gamma_{[6]}$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$1-\delta_1-\delta_3-\rho_{[2]}-\rho_{[6]}$

Table 7: Values of lower bounds on $\gamma_{[s]}-\gamma_{[r]}$ for s (in columns) and r (in rows) for the example arrangement.

In the binary outcome SEIV model a monotonicity restriction coupled with a single index restriction, requiring that the threshold function is a monotone function of a scalar valued function of endogenous and exogenous variables, brings great simplification as shown in Chesher (2009). The use and benefit of restrictions on threshold functions such as monotonicity, concavity, convexity and single-peakedness coupled with index restrictions is the subject of current research.

Shape restrictions can also be introduced by employing constrained sieve approximations. Parametric restrictions cut down the scale of the problem and provide a link to classical likelihood based analysis of discrete outcome data. This is illustrated in Chesher and Smolinski (2009) where ordered probit structural functions are employed with a coefficient on a scalar endogenous variable X that is common across threshold functions whose "intercept terms" differ. This model embodies strong shape restrictions, requiring threshold functions to be monotone in X and parallel after applying the inverse normal distribution function transformation.

As a prelude to consideration of methods for estimating identified sets in parametric or otherwise shape constrained models, first consider a *theoretical* analysis in which one has to hand probability distributions $\Pr[Y=m \wedge X=x_k|Z=z]$ for each value $z \in \mathcal{Z}$. Suppose there is a parametric model or sieve approximating model for the structural function with parameter vector Θ. For any value θ there is an associated value of $\boldsymbol{\gamma}$ denoted by $\boldsymbol{\gamma}(\theta)$ which is in some arrangement denoted by $t(\theta)$. The values of $\boldsymbol{\gamma}(\theta)$ and $t(\theta)$ are easy to compute. The value θ is in the identified set of parameter values, denoted by $\mathcal{H}^{\Theta}(\mathcal{Z})$, if and only if $\boldsymbol{\gamma}(\theta) \in \mathcal{E}_{t(\theta)}(\mathcal{Z})$.

Define the non-negative valued distance measure $D(\theta)$ as follows.

$$D(\theta) \equiv \min_{w \in \mathcal{E}_{t(\theta)}(\mathcal{Z})} \left((\gamma(\theta)-w)' (\gamma(\theta)-w) \right) \tag{29}$$

This is the squared Euclidean distance from $\gamma(\theta)$ to the point in the set $\mathcal{E}_{t(\theta)}(\mathcal{Z})$ closest to $\gamma(\theta)$ as the crow flies. The measure is zero if and only if $\boldsymbol{\gamma}(\theta) \in \mathcal{E}_{t(\theta)}(\mathcal{Z})$ and so zero if and only if $\theta \in \mathcal{H}^{\Theta}(\mathcal{Z})$. The value of $D(\theta)$ is easily found using a quadratic programming algorithm and the expressions for the linear half spaces defining the sets $\mathcal{E}_t(\mathcal{Z})$ that we have given in this paper.[20] The set of values of θ that

[20] An L1 norm and a linear programming calculation could be employed instead.

minimise the function $D(\cdot)$ is the identified set $\mathcal{H}^{\Theta}(\mathcal{Z})$.

$$\mathcal{H}^{\Theta}(\mathcal{Z}) = \{\theta : \theta = \arg\min_{s} D(s)\} = \{\theta : D(\theta) = 0\}.$$

In applied econometric work there will be *estimates* of the probability distributions, $\Pr[Y = m \wedge X = x_k | Z = z]$ for $z \in \mathcal{Z}$, and so estimates of the sets $\mathcal{E}_t(\mathcal{Z})$. Let $\hat{D}(\theta)$ be the distance measure arising when $\mathcal{E}_t(\mathcal{Z})$ in (29) is replaced by an estimate $\hat{\mathcal{E}}_t(\mathcal{Z})$. The distance measure $\hat{D}(\theta)$ has the properties required of Chernozhukov, Hong and Tamer's (2007) "econometric criterion function" $Q(\theta)$ and their methods can be employed to estimate, and develop confidence regions for, the set $\mathcal{H}^{\Theta}(\mathcal{Z})$.

It will be prudent to use bias corrected estimates of the sets $\mathcal{E}_t(\mathcal{Z})$. Bias arises because the sets $\mathcal{E}_t(\mathcal{Z})$ arise as *intersections* of sets $\mathcal{E}_t(z)$ across values $z \in \mathcal{Z}$. The issue is explained in Chernozhukov, Lee and Rosen (2009) (CLR) where a solution is proposed. This is directly applicable in the case that arises here.

Define $\rho^t(z) \equiv \{\rho^t_{[1]}(z), \ldots, \rho^t_{[N]}(z)\}$. With $\gamma_{[0]} \equiv 0$ and $\gamma_{[N+1]} \equiv 1$ all the constraints defining a set $\mathcal{E}_t(\mathcal{Z})$ have the form

$$\gamma_{[s]} - \gamma_{[r]} \geq \max_{z \in \mathcal{Z}} \left(\alpha_{sr} \centerdot \rho^t(z)\right) \tag{30}$$

for certain pairs of indices $s > r$ selected from $\{0, 1, \ldots, N+1\}$. Here α_{sr} is a vector of integers specific to the $s - r$ comparison.

The proposal in CLR is to calculate an estimate of $l^t(\alpha_{sr}, \mathcal{Z}) \equiv \max_{z \in \mathcal{Z}} \left(\alpha_{sr} \centerdot \rho^t(z)\right)$ by calculating the maximum over $z \in \mathcal{Z}$ of *precision corrected* estimates as follows.

$$\hat{l}^t(\alpha_{sr}, \mathcal{Z}) = \max_{z \in \hat{\mathcal{Z}}} \left(\alpha_{sr} \centerdot \hat{\rho}^t(z) + \kappa \sigma^t(\alpha_{sr}, z)\right)$$

Here $\sigma^t(\alpha_{sr}, z)$ is the standard error of $\alpha_{sr} \centerdot \hat{\rho}^t(z)$, $\hat{\mathcal{Z}}$ is a data dependent set of values of z that converges in probability to a non-stochastic set which contains $\arg\max_{z \in \mathcal{Z}} \left(\alpha_{sr} \centerdot \rho^t(z)\right)$ and κ is an estimate of

$$\text{median}\left(\inf_{z \in \hat{\mathcal{Z}}} \frac{\alpha_{sr} \centerdot \rho^t(z) - \alpha_{sr} \centerdot \hat{\rho}^t(z)}{\sigma^t(\alpha_{sr}, z)}\right)$$

proposals for which are given in CLR.

The result is an asymptotically upward median unbiased estimate of the bound in (30). Proceeding in this way gives bias corrected estimates of all bounds and thus bias corrected estimated sets $\hat{\mathcal{E}}_t(\mathcal{Z})$ which will be used in the calculation of the distance measure $\hat{D}(\theta)$. An example of inference using the CLR method in a binary outcome case is given in Chesher (2009).

5. CONCLUDING REMARKS

We have studied identification of a nonparametrically specified structural function in a discrete outcome - discrete endogenous variable setting. The single equation instrumental variable (SEIV) model we have considered is attractive because it places no restrictions on the process generating values of endogenous variables. Commonly

used control function alternatives based on triangular models do not deliver point identification when, as here, endogenous variables are discrete unless there are strong parametric restrictions.

The SEIV model *set* identifies the structural function. In the M outcome case the structural function is characterised by $M-1$ threshold functions. In general the identified set is a union of a large number of convex sets. In principle there is one such set associated with each admissible ordering of the K values taken by $M-1$ threshold functions as endogenous variables pass across their K points of support.

Each convex component of the identified set is the intersection of collections of linear half spaces, each value of the instrumental variables generating one such collection. The number and extent of the convex components of the identified set depends on the strength and support of the instrumental variables. When these are good predictors of the values of endogenous variables the identified set may comprise just a small number of convex components, perhaps just one.

We have developed expressions for a set $\mathcal{E}(\mathcal{Z})$ which can be calculated for any probability distribution of the outcome Y and endogenous variables X given instruments Z taking values in a set $\mathcal{Z}$. We have shown that the set identified by the SEIV model, $\mathcal{H}(\mathcal{Z})$, is equal to $\mathcal{E}(\mathcal{Z})$. We provide a `Mathematica` notebook which conducts symbolic calculation of convex components of the identified set.

Unrestricted nonparametric estimation and inference pose challenging problems once M or K are at all large. Parametric restrictions or shape restrictions reduce the scale of the estimation problem. We have defined an easy-to-compute criterion function which can be employed in estimation using the methods proposed in Chernozhukov, Hong and Tamer (2007) with bias corrected estimates of bounds as proposed in Chernozhukov, Lee and Rosen (2009).

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 (2001): "An IV Model of Quantile Treatment Effects," Working Paper 02-06, Department of Economics, MIT. Available at http://ssrn.com/abstract=298879 or doi: 10.2139/ssrn.298879.

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

Chernozhukov, Victor, Han Hong and Elie Tamer (2007): "Estimation and Confidence Regions for Parameter Sets in Econometric Models," *Econometrica*, 75, 1243-1284.

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., (2007): "Endogeneity and Discrete Outcomes," CeMMAP Working Paper 05/07.
CHESHER, ANDREW D., (2009): "Single Equation Endogenous Binary Response Models," CeMMAP Working Paper 23/09.
CHESHER, ANDREW D., (2010): "Instrumental Variable Models for Discrete Outcomes," *Econometrica*, 78, 575-602.
CHESHER, ANDREW D., AND KONRAD SMOLINSKI (2009): "IV Models of Ordered Choice," CeMMAP Working Paper 37/09.
DAS, MITALI (2005): "Instrumental Variables Estimators of Nonparametric Models with Discrete Endogenous Regressors," *Journal of Econometrics*, 124, 335-361.
HECKMAN, JAMES J., (1978): "Dummy Endogenous Variables in a Simultaneous Equations System," *Econometrica*, 46, 931-959.
IMBENS, GUIDO W., AND WHITNEY K. NEWEY (2009): "Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity," *Econometrica*, 77, 1481–1512.
JUN, SUNG JAE, JORIS PINKSE AND HAIQING XU (2009): "Tighter Bounds in Triangular Systems," unpublished paper.
LEWBEL, ARTHUR (2007). "Coherency And Completeness Of Structural Models Containing A Dummy Endogenous Variable," *International Economic Review*, 48, 1379-1392.
TAMER, ELIE (2003). "Incomplete Simultaneous Discrete Response Model with Multiple Equilibria," *Review of Economic Studies*, 70, 147–165.
ZIEGLER, GUNTER M., (2007): *Lectures on Polytopes.* (Updated seventh printing of the first edition*)*. New York: Springer Science.

Annex 1: Proof of Proposition 1

1. (a) This follows directly from (12) which states that for all i and j, $\eta(i,j)$ is either equal to $G(i,j)$ or, equal to $\max(0, R(i,j))$ if this is less than $G(i,j)$.

   (b) For any i and j, if $G(i,j) \geq 0$ then $\eta(i,j) \geq 0$ because (12) states that $\eta(i,j)$ is at least equal to $G(i,j)$ or a non-negative quantity, namely $\max(0, R(i,j))$. The recursion (13) taken together with (a) and $G(i,j) \geq 0 \implies \eta(i,j) \geq 0$ implies that, for any i and j, if $G(i,j-1) \geq 0$ then $G(i,j) \geq 0$. Since for all i, $G(i,1) = \Delta\gamma_{[i]} \geq 0$, the result follows by induction letting j pass from 2 to K.

   (c) As noted in the proof of (b), for all i and j, $G(i,j) \geq 0 \implies \eta(i,j) \geq 0$ and the result follows because the result (b) states that for all i and j, indeed, $G(i,j) \geq 0$.

   (d) This follows directly from (13) and (a) and (c).

   (e) If $G(i,j) > \eta(i,j)$ then $G(i,j) > 0$ and since $G(i,j)$ is a non-increasing function of j, for all $j' \leq j$, $G(i,j') > 0$. Therefore, for all $j' < j$, from (13), $G(i,j') > \eta(i,j')$ which by assumption also holds for $j' = j$. From (12), if $G(i,j') > \eta(i,j')$ then $\eta(i,j') = \max(0, R(i,j'))$.

2. Suppose that for some $j \leq K$, $G(i,j) \leq \max(0, R(i,j))$. Then $\eta(i,j) = G(i,j)$ and from (13) $G(i,j+1) = 0$ and by repeated application of (13), for all $j' > j$, $\eta(i,j') = G(i,j') = 0$ and so

$$G(i,K+1) = \Delta\gamma_{[i]} - \sum_{k=1}^{K} \eta_{[i]k} = 0.$$

   Suppose that there is no $j \leq K$ such that $G(i,j) \leq \max(0, R(i,j))$. Then, considering $j = K$,

$$G(i,K) > \max(0, R(i,K))$$

   so

$$\eta(i,K) = \max(0, R(i,K))$$

   and so from (13)

$$G(i,K+1) = \Delta\gamma_{[i]} - \sum_{k=1}^{K} \eta_{[i]k} = G(i,K) - \eta(i,K) > 0$$

   and so

$$\sum_{k=1}^{K} \eta_{[i]k} < \Delta\gamma_{[i]}.$$

Annex 2

This Annex provides tables like Table 7 giving lower bounds on differences $\gamma_{[s]} - \gamma_{[r]}$ for the 15 fundamental arrangements of $\boldsymbol{\gamma}$ in the $M = 3$, $K = 3$ case. In each case the final column gives lower bounds on $1 - \gamma_{[r]}$ for $r \in \{1, \ldots, 6\}$. Subtracting 1 from each of these expressions and changing sign gives upper bounds on $\gamma_{[r]}$. Lower bounds are simply $\gamma_{[r]} \geq \sum_{r'=1}^{r} \rho_{[r']}$.

$A.1$		γ_{11}	γ_{12}	γ_{13}	γ_{21}	γ_{22}	γ_{23}	1
		$\gamma_{[1]}$	$\gamma_{[2]}$	$\gamma_{[3]}$	$\gamma_{[4]}$	$\gamma_{[5]}$	$\gamma_{[6]}$	$\gamma_{[7]}$
γ_{11}	$\gamma_{[1]}$	$\cdot$	0	0	$\rho_{[4]}$	$\rho_{[4]}+\rho_{[5]}$	$\rho_{[4]}+\rho_{[5]}+\rho_{[6]}$	$1-\rho_{[1]}-\rho_{[2]}-\rho_{[3]}$
γ_{12}	$\gamma_{[2]}$	$\cdot$	$\cdot$	0	0	$\rho_{[5]}$	$\rho_{[5]}+\rho_{[6]}$	$1-\rho_{[1]}-\rho_{[2]}-\rho_{[3]}-\rho_{[4]}$
γ_{13}	$\gamma_{[3]}$	$\cdot$	$\cdot$	$\cdot$	0	0	$\rho_{[6]}$	$1-\rho_{[1]}-\rho_{[2]}-\rho_{[3]}-\rho_{[4]}-\rho_{[5]}$
γ_{21}	$\gamma_{[4]}$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	0	0	$1-\rho_{[1]}-\rho_{[2]}-\rho_{[3]}-\rho_{[4]}-\rho_{[5]}-\rho_{[6]}$
γ_{22}	$\gamma_{[5]}$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	0	$1-\delta_{[1]}-\rho_{[2]}-\rho_{[3]}-\rho_{[5]}-\rho_{[6]}$
γ_{23}	$\gamma_{[6]}$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$1-\delta_{[1]}-\delta_{[2]}-\rho_{[3]}-\rho_{[6]}$

$A.2$		γ_{11}	γ_{12}	γ_{13}	γ_{21}	γ_{23}	γ_{22}	1
		$\gamma_{[1]}$	$\gamma_{[2]}$	$\gamma_{[3]}$	$\gamma_{[4]}$	$\gamma_{[5]}$	$\gamma_{[6]}$	$\gamma_{[7]}$
γ_{11}	$\gamma_{[1]}$	$\cdot$	0	0	$\rho_{[4]}$	$\rho_{[4]}+\rho_{[5]}$	$\rho_{[4]}+\rho_{[5]}+\rho_{[6]}$	$1-\rho_{[1]}-\rho_{[2]}-\rho_{[3]}$
γ_{12}	$\gamma_{[2]}$	$\cdot$	$\cdot$	0	0	$\rho_{[5]}$	$\rho_{[5]}+\rho_{[6]}$	$1-\rho_{[1]}-\rho_{[2]}-\rho_{[3]}-\rho_{[4]}$
γ_{13}	$\gamma_{[3]}$	$\cdot$	$\cdot$	$\cdot$	0	$\rho_{[5]}$	$\rho_{[5]}$	$1-\rho_{[1]}-\rho_{[2]}-\rho_{[3]}-\rho_{[4]}-\rho_{[6]}$
γ_{21}	$\gamma_{[4]}$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	0	0	$1-\rho_{[1]}-\rho_{[2]}-\rho_{[3]}-\rho_{[4]}-\rho_{[5]}-\rho_{[6]}$
γ_{23}	$\gamma_{[5]}$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	0	$1-\delta_{[1]}-\rho_{[2]}-\rho_{[3]}-\rho_{[5]}-\rho_{[6]}$
γ_{22}	$\gamma_{[6]}$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$1-\delta_{[1]}-\delta_{[3]}-\rho_{[2]}-\rho_{[6]}$

$A.3$		γ_{11}	γ_{12}	γ_{13}	γ_{22}	γ_{21}	γ_{23}	1
		$\gamma_{[1]}$	$\gamma_{[2]}$	$\gamma_{[3]}$	$\gamma_{[4]}$	$\gamma_{[5]}$	$\gamma_{[6]}$	$\gamma_{[7]}$
γ_{11}	$\gamma_{[1]}$	$\cdot$	0	0	$\rho_{[4]}$	$\rho_{[4]}+\rho_{[5]}$	$\rho_{[4]}+\rho_{[5]}+\rho_{[6]}$	$1-\rho_{[1]}-\rho_{[2]}-\rho_{[3]}$
γ_{12}	$\gamma_{[2]}$	$\cdot$	$\cdot$	0	$\rho_{[4]}$	$\rho_{[4]}$	$\rho_{[4]}+\rho_{[6]}$	$1-\rho_{[1]}-\rho_{[2]}-\rho_{[3]}-\rho_{[5]}$
γ_{13}	$\gamma_{[3]}$	$\cdot$	$\cdot$	$\cdot$	0	0	$\rho_{[6]}$	$1-\rho_{[1]}-\rho_{[2]}-\rho_{[3]}-\rho_{[4]}-\rho_{[5]}$
γ_{22}	$\gamma_{[4]}$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	0	0	$1-\rho_{[1]}-\rho_{[2]}-\rho_{[3]}-\rho_{[4]}-\rho_{[5]}-\rho_{[6]}$
γ_{21}	$\gamma_{[5]}$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	0	$1-\delta_{[2]}-\rho_{[1]}-\rho_{[3]}-\rho_{[5]}-\rho_{[6]}$
γ_{23}	$\gamma_{[6]}$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$1-\delta_{[1]}-\delta_{[2]}-\rho_{[3]}-\rho_{[6]}$

Table 8:

$A.4$		γ_{11}	γ_{12}	γ_{13}	γ_{22}	γ_{23}	γ_{21}	1
		$\gamma_{[1]}$	$\gamma_{[2]}$	$\gamma_{[3]}$	$\gamma_{[4]}$	$\gamma_{[5]}$	$\gamma_{[6]}$	$\gamma_{[7]}$
γ_{11}	$\gamma_{[1]}$	$\cdot$	0	0	$\rho_{[4]}$	$\rho_{[4]}+\rho_{[5]}$	$\rho_{[4]}+\rho_{[5]}+\rho_{[6]}$	$1-\rho_{[1]}-\rho_{[2]}-\rho_{[3]}$
γ_{12}	$\gamma_{[2]}$	$\cdot$	$\cdot$	0	$\rho_{[4]}$	$\rho_{[4]}+\rho_{[5]}$	$\rho_{[4]}+\rho_{[5]}$	$1-\rho_{[1]}-\rho_{[2]}-\rho_{[3]}-\rho_{[6]}$
γ_{13}	$\gamma_{[3]}$	$\cdot$	$\cdot$	$\cdot$	0	$\rho_{[5]}$	$\rho_{[5]}$	$1-\rho_{[1]}-\rho_{[2]}-\rho_{[3]}-\rho_{[4]}-\rho_{[6]}$
γ_{22}	$\gamma_{[4]}$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	0	0	$1-\rho_{[1]}-\rho_{[2]}-\rho_{[3]}-\rho_{[4]}-\rho_{[5]}-\rho_{[6]}$
γ_{23}	$\gamma_{[5]}$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	0	$1-\delta_{[2]}-\rho_{[1]}-\rho_{[3]}-\rho_{[5]}-\rho_{[6]}$
γ_{21}	$\gamma_{[6]}$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$1-\delta_{[2]}-\delta_{[3]}-\rho_{[1]}-\rho_{[6]}$

$A.5$		γ_{11}	γ_{12}	γ_{13}	γ_{23}	γ_{21}	γ_{22}	1
		$\gamma_{[1]}$	$\gamma_{[2]}$	$\gamma_{[3]}$	$\gamma_{[4]}$	$\gamma_{[5]}$	$\gamma_{[6]}$	$\gamma_{[7]}$
γ_{11}	$\gamma_{[1]}$	$\cdot$	0	0	$\rho_{[4]}$	$\rho_{[4]}+\rho_{[5]}$	$\rho_{[4]}+\rho_{[5]}+\rho_{[6]}$	$1-\rho_{[1]}-\rho_{[2]}-\rho_{[3]}$
γ_{12}	$\gamma_{[2]}$	$\cdot$	$\cdot$	0	$\rho_{[4]}$	$\rho_{[4]}$	$\rho_{[4]}+\rho_{[6]}$	$1-\rho_{[1]}-\rho_{[2]}-\rho_{[3]}-\rho_{[5]}$
γ_{13}	$\gamma_{[3]}$	$\cdot$	$\cdot$	$\cdot$	$\rho_{[4]}$	$\rho_{[4]}$	$\rho_{[4]}$	$1-\rho_{[1]}-\rho_{[2]}-\rho_{[3]}-\rho_{[5]}-\rho_{[6]}$
γ_{23}	$\gamma_{[4]}$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	0	0	$1-\rho_{[1]}-\rho_{[2]}-\rho_{[3]}-\rho_{[4]}-\rho_{[5]}-\rho_{[6]}$
γ_{21}	$\gamma_{[5]}$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	0	$1-\delta_{[3]}-\rho_{[1]}-\rho_{[2]}-\rho_{[5]}-\rho_{[6]}$
γ_{22}	$\gamma_{[6]}$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$1-\delta_{[1]}-\delta_{[3]}-\rho_{[2]}-\rho_{[6]}$

$A.6$		γ_{11}	γ_{12}	γ_{13}	γ_{23}	γ_{22}	γ_{21}	1
		$\gamma_{[1]}$	$\gamma_{[2]}$	$\gamma_{[3]}$	$\gamma_{[4]}$	$\gamma_{[5]}$	$\gamma_{[6]}$	$\gamma_{[7]}$
γ_{11}	$\gamma_{[1]}$	$\cdot$	0	0	$\rho_{[4]}$	$\rho_{[4]}+\rho_{[5]}$	$\rho_{[4]}+\rho_{[5]}+\rho_{[6]}$	$1-\rho_{[1]}-\rho_{[2]}-\rho_{[3]}$
γ_{12}	$\gamma_{[2]}$	$\cdot$	$\cdot$	0	$\rho_{[4]}$	$\rho_{[4]}+\rho_{[5]}$	$\rho_{[4]}+\rho_{[5]}$	$1-\rho_{[1]}-\rho_{[2]}-\rho_{[3]}-\rho_{[6]}$
γ_{13}	$\gamma_{[3]}$	$\cdot$	$\cdot$	$\cdot$	$\rho_{[4]}$	$\rho_{[4]}$	$\rho_{[4]}$	$1-\rho_{[1]}-\rho_{[2]}-\rho_{[3]}-\rho_{[5]}-\rho_{[6]}$
γ_{23}	$\gamma_{[4]}$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	0	0	$1-\rho_{[1]}-\rho_{[2]}-\rho_{[3]}-\rho_{[4]}-\rho_{[5]}-\rho_{[6]}$
γ_{22}	$\gamma_{[5]}$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	0	$1-\delta_{[3]}-\rho_{[1]}-\rho_{[2]}-\rho_{[5]}-\rho_{[6]}$
γ_{21}	$\gamma_{[6]}$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$1-\delta_{[2]}-\delta_{[3]}-\rho_{[1]}-\rho_{[6]}$

Table 9:

$A.7$		γ_{11}	γ_{12}	γ_{21}	γ_{13}	γ_{22}	γ_{23}	1
		$\gamma_{[1]}$	$\gamma_{[2]}$	$\gamma_{[3]}$	$\gamma_{[4]}$	$\gamma_{[5]}$	$\gamma_{[6]}$	$\gamma_{[7]}$
γ_{11}	$\gamma_{[1]}$	$\cdot$	0	$\rho_{[3]}$	$\rho_{[3]}$	$\rho_{[3]}+\rho_{[5]}$	$\rho_{[3]}+\rho_{[5]}+\rho_{[6]}$	$1-\rho_{[1]}-\rho_{[2]}-\rho_{[4]}$
γ_{12}	$\gamma_{[2]}$	$\cdot$	$\cdot$	0	0	$\rho_{[5]}$	$\rho_{[5]}+\rho_{[6]}$	$1-\rho_{[1]}-\rho_{[2]}-\rho_{[3]}-\rho_{[4]}$
γ_{21}	$\gamma_{[3]}$	$\cdot$	$\cdot$	$\cdot$	0	0	$\rho_{[6]}$	$1-\rho_{[1]}-\rho_{[2]}-\rho_{[3]}-\rho_{[4]}-\rho_{[5]}$
γ_{13}	$\gamma_{[4]}$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	0	$\rho_{[6]}$	$1-\delta_{[1]}-\rho_{[2]}-\rho_{[4]}-\rho_{[5]}$
γ_{22}	$\gamma_{[5]}$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	0	$1-\delta_{[1]}-\rho_{[2]}-\rho_{[4]}-\rho_{[5]}-\rho_{[6]}$
γ_{23}	$\gamma_{[6]}$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$1-\delta_{[1]}-\delta_{[2]}-\rho_{[4]}-\rho_{[6]}$

$A.8$		γ_{11}	γ_{12}	γ_{21}	γ_{13}	γ_{23}	γ_{22}	1
		$\gamma_{[1]}$	$\gamma_{[2]}$	$\gamma_{[3]}$	$\gamma_{[4]}$	$\gamma_{[5]}$	$\gamma_{[6]}$	$\gamma_{[7]}$
γ_{11}	$\gamma_{[1]}$	$\cdot$	0	0	$\rho_{[3]}$	$\rho_{[3]}+\rho_{[5]}$	$\rho_{[3]}+\rho_{[5]}+\rho_{[6]}$	$1-\rho_{[1]}-\rho_{[2]}-\rho_{[4]}$
γ_{12}	$\gamma_{[2]}$	$\cdot$	$\cdot$	0	0	$\rho_{[5]}$	$\rho_{[5]}+\rho_{[6]}$	$1-\rho_{[1]}-\rho_{[2]}-\rho_{[3]}-\rho_{[4]}$
γ_{21}	$\gamma_{[3]}$	$\cdot$	$\cdot$	$\cdot$	0	$\rho_{[5]}$	$\rho_{[5]}$	$1-\rho_{[1]}-\rho_{[2]}-\rho_{[3]}-\rho_{[4]}-\rho_{[6]}$
γ_{13}	$\gamma_{[4]}$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\rho_{[5]}$	$\rho_{[5]}$	$1-\delta_{[1]}-\rho_{[2]}-\rho_{[4]}-\rho_{[6]}$
γ_{23}	$\gamma_{[5]}$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	0	$1-\delta_{[1]}-\rho_{[2]}-\rho_{[3]}-\rho_{[5]}-\rho_{[6]}$
γ_{22}	$\gamma_{[6]}$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$1-\delta_{[1]}-\delta_{[3]}-\rho_{[2]}-\rho_{[6]}$

$A.9$		γ_{11}	γ_{12}	γ_{21}	γ_{22}	γ_{13}	γ_{23}	1
		$\gamma_{[1]}$	$\gamma_{[2]}$	$\gamma_{[3]}$	$\gamma_{[4]}$	$\gamma_{[5]}$	$\gamma_{[6]}$	$\gamma_{[7]}$
γ_{11}	$\gamma_{[1]}$	$\cdot$	0	$\rho_{[3]}$	$\rho_{[3]}+\rho_{[4]}$	$\rho_{[3]}+\rho_{[4]}$	$\rho_{[3]}+\rho_{[4]}+\rho_{[6]}$	$1-\rho_{[1]}-\rho_{[2]}-\rho_{[5]}$
γ_{12}	$\gamma_{[2]}$	$\cdot$	$\cdot$	0	$\rho_{[4]}$	$\rho_{[4]}$	$\rho_{[4]}+\rho_{[6]}$	$1-\rho_{[1]}-\rho_{[2]}-\rho_{[3]}-\rho_{[5]}$
γ_{21}	$\gamma_{[3]}$	$\cdot$	$\cdot$	$\cdot$	0	0	$\rho_{[6]}$	$1-\rho_{[1]}-\rho_{[2]}-\rho_{[3]}-\rho_{[4]}-\rho_{[5]}$
γ_{22}	$\gamma_{[4]}$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	0	$\rho_{[6]}$	$1-\delta_{[1]}-\rho_{[2]}-\rho_{[4]}-\rho_{[5]}$
γ_{13}	$\gamma_{[5]}$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\rho_{[6]}$	$1-\delta_{[1]}-\delta_{[2]}-\rho_{[5]}$
γ_{23}	$\gamma_{[6]}$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$1-\delta_{[1]}-\delta_{[2]}-\rho_{[5]}-\rho_{[6]}$

Table 10:

$A.10$		γ_{11}	γ_{12}	γ_{22}	γ_{13}	γ_{21}	γ_{23}	1
		$\gamma_{[1]}$	$\gamma_{[2]}$	$\gamma_{[3]}$	$\gamma_{[4]}$	$\gamma_{[5]}$	$\gamma_{[6]}$	$\gamma_{[7]}$
γ_{11}	$\gamma_{[1]}$	$\cdot$	0	$\rho_{[3]}$	$\rho_{[3]}$	$\rho_{[3]}+\rho_{[5]}$	$\rho_{[3]}+\rho_{[5]}+\rho_{[6]}$	$1-\rho_{[1]}-\rho_{[2]}-\rho_{[4]}$
γ_{12}	$\gamma_{[2]}$	$\cdot$	$\cdot$	$\rho_{[3]}$	$\rho_{[3]}$	$\rho_{[3]}$	$\rho_{[3]}+\rho_{[6]}$	$1-\rho_{[1]}-\rho_{[2]}-\rho_{[4]}-\rho_{[5]}$
γ_{22}	$\gamma_{[3]}$	$\cdot$	$\cdot$	$\cdot$	0	0	$\rho_{[6]}$	$1-\rho_{[1]}-\rho_{[2]}-\rho_{[3]}-\rho_{[4]}-\rho_{[5]}$
γ_{13}	$\gamma_{[4]}$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	0	$\rho_{[6]}$	$1-\delta_{[2]}-\rho_{[1]}-\rho_{[4]}-\rho_{[5]}$
γ_{21}	$\gamma_{[5]}$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	0	$1-\delta_{[2]}-\rho_{[1]}-\rho_{[4]}-\rho_{[5]}-\rho_{[6]}$
γ_{23}	$\gamma_{[6]}$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$1-\delta_{[1]}-\delta_{[2]}-\rho_{[4]}-\rho_{[6]}$

$A.11$		γ_{11}	γ_{12}	γ_{22}	γ_{13}	γ_{23}	γ_{21}	1
		$\gamma_{[1]}$	$\gamma_{[2]}$	$\gamma_{[3]}$	$\gamma_{[4]}$	$\gamma_{[5]}$	$\gamma_{[6]}$	$\gamma_{[7]}$
γ_{11}	$\gamma_{[1]}$	$\cdot$	0	$\rho_{[3]}$	$\rho_{[3]}$	$\rho_{[3]}+\rho_{[5]}$	$\rho_{[3]}+\rho_{[5]}+\rho_{[6]}$	$1-\rho_{[1]}-\rho_{[2]}-\rho_{[4]}$
γ_{12}	$\gamma_{[2]}$	$\cdot$	$\cdot$	$\rho_{[3]}$	$\rho_{[3]}$	$\rho_{[3]}+\rho_{[5]}$	$\rho_{[5]}+\rho_{[6]}$	$1-\rho_{[1]}-\rho_{[2]}-\rho_{[4]}-\rho_{[6]}$
γ_{22}	$\gamma_{[3]}$	$\cdot$	$\cdot$	$\cdot$	0	$\rho_{[5]}$	$\rho_{[5]}$	$1-\rho_{[1]}-\rho_{[2]}-\rho_{[3]}-\rho_{[4]}-\rho_{[6]}$
γ_{13}	$\gamma_{[4]}$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\rho_{[5]}$	$\rho_{[5]}$	$1-\delta_{[2]}-\rho_{[1]}-\rho_{[4]}-\rho_{[6]}$
γ_{23}	$\gamma_{[5]}$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	0	$1-\delta_{[2]}-\rho_{[1]}-\rho_{[3]}-\rho_{[5]}-\rho_{[6]}$
γ_{21}	$\gamma_{[6]}$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$1-\delta_{[2]}-\delta_{[3]}-\rho_{[1]}-\rho_{[6]}$

$A.12$		γ_{11}	γ_{12}	γ_{22}	γ_{21}	γ_{13}	γ_{23}	1
		$\gamma_{[1]}$	$\gamma_{[2]}$	$\gamma_{[3]}$	$\gamma_{[4]}$	$\gamma_{[5]}$	$\gamma_{[6]}$	$\gamma_{[7]}$
γ_{11}	$\gamma_{[1]}$	$\cdot$	0	$\rho_{[3]}$	$\rho_{[3]}+\rho_{[4]}$	$\rho_{[3]}+\rho_{[4]}$	$\rho_{[3]}+\rho_{[4]}+\rho_{[6]}$	$1-\rho_{[1]}-\rho_{[2]}-\rho_{[5]}$
γ_{12}	$\gamma_{[2]}$	$\cdot$	$\cdot$	$\rho_{[3]}$	$\rho_{[3]}$	$\rho_{[3]}$	$\rho_{[3]}+\rho_{[6]}$	$1-\rho_{[1]}-\rho_{[2]}-\rho_{[4]}-\rho_{[5]}$
γ_{22}	$\gamma_{[3]}$	$\cdot$	$\cdot$	$\cdot$	0	0	$\rho_{[6]}$	$1-\rho_{[1]}-\rho_{[2]}-\rho_{[3]}-\rho_{[4]}-\rho_{[5]}$
γ_{21}	$\gamma_{[4]}$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	0	$\rho_{[6]}$	$1-\delta_{[2]}-\rho_{[1]}-\rho_{[4]}-\rho_{[5]}$
γ_{13}	$\gamma_{[5]}$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\rho_{[6]}$	$1-\delta_{[1]}-\delta_{[2]}-\rho_{[5]}$
γ_{23}	$\gamma_{[6]}$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$1-\delta_{[1]}-\delta_{[2]}-\rho_{[5]}-\rho_{[6]}$

Table 11:

$A.13$		γ_{11}	γ_{21}	γ_{12}	γ_{13}	γ_{22}	γ_{23}	1
		$\gamma_{[1]}$	$\gamma_{[2]}$	$\gamma_{[3]}$	$\gamma_{[4]}$	$\gamma_{[5]}$	$\gamma_{[6]}$	$\gamma_{[7]}$
γ_{11}	$\gamma_{[1]}$	$\cdot$	0	$\rho_{[2]}$	$\rho_{[2]}$	$\rho_{[2]}+\rho_{[5]}$	$\rho_{[2]}+\rho_{[5]}+\rho_{[6]}$	$1-\rho_{[1]}-\rho_{[3]}-\rho_{[4]}$
γ_{21}	$\gamma_{[2]}$	$\cdot$	$\cdot$	0	0	$\rho_{[5]}$	$\rho_{[5]}+\rho_{[6]}$	$1-\rho_{[1]}-\rho_{[2]}-\rho_{[3]}-\rho_{[4]}$
γ_{12}	$\gamma_{[3]}$	$\cdot$	$\cdot$	$\cdot$	0	$\rho_{[5]}$	$\rho_{[5]}+\rho_{[6]}$	$1-\delta_{[1]}-\rho_{[3]}-\rho_{[4]}$
γ_{13}	$\gamma_{[4]}$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	0	$\rho_{[6]}$	$1-\delta_{[1]}-\rho_{[3]}-\rho_{[4]}-\rho_{[5]}$
γ_{22}	$\gamma_{[5]}$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	0	$1-\delta_{[1]}-\rho_{[3]}-\rho_{[4]}-\rho_{[5]}-\rho_{[6]}$
γ_{23}	$\gamma_{[6]}$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$1-\delta_{[1]}-\delta_{[2]}-\rho_{[4]}-\rho_{[6]}$

$A.14$		γ_{11}	γ_{21}	γ_{12}	γ_{13}	γ_{23}	γ_{22}	1
		$\gamma_{[1]}$	$\gamma_{[2]}$	$\gamma_{[3]}$	$\gamma_{[4]}$	$\gamma_{[5]}$	$\gamma_{[6]}$	$\gamma_{[7]}$
γ_{11}	$\gamma_{[1]}$	$\cdot$	$\rho_{[2]}$	$\rho_{[2]}$	$\rho_{[2]}$	$\rho_{[2]}+\rho_{[5]}$	$\rho_{[2]}+\rho_{[5]}+\rho_{[6]}$	$1-\rho_{[1]}-\rho_{[3]}-\rho_{[4]}$
γ_{21}	$\gamma_{[2]}$	$\cdot$	$\cdot$	0	0	$\rho_{[3]}+\rho_{[5]}$	$\rho_{[5]}+\rho_{[6]}$	$1-\rho_{[1]}-\rho_{[2]}-\rho_{[3]}-\rho_{[4]}$
γ_{12}	$\gamma_{[3]}$	$\cdot$	$\cdot$	$\cdot$	0	$\rho_{[5]}$	$\rho_{[5]}+\rho_{[6]}$	$1-\delta_{[1]}-\rho_{[3]}-\rho_{[4]}$
γ_{13}	$\gamma_{[4]}$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\rho_{[5]}$	$\rho_{[5]}$	$1-\delta_{[1]}-\rho_{[3]}-\rho_{[4]}-\rho_{[6]}$
γ_{23}	$\gamma_{[5]}$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	0	$1-\delta_{[1]}-\rho_{[3]}-\rho_{[4]}-\rho_{[5]}-\rho_{[6]}$
γ_{22}	$\gamma_{[6]}$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$1-\delta_{[1]}-\delta_{[3]}-\rho_{[3]}-\rho_{[6]}$

$A.15$		γ_{11}	γ_{21}	γ_{12}	γ_{22}	γ_{13}	γ_{23}	1
		$\gamma_{[1]}$	$\gamma_{[2]}$	$\gamma_{[3]}$	$\gamma_{[4]}$	$\gamma_{[5]}$	$\gamma_{[6]}$	$\gamma_{[7]}$
γ_{11}	$\gamma_{[1]}$	$\cdot$	$\rho_{[2]}$	$\rho_{[2]}$	$\rho_{[2]}+\rho_{[4]}$	$\rho_{[2]}+\rho_{[4]}$	$\rho_{[2]}+\rho_{[4]}+\rho_{[6]}$	$1-\rho_{[1]}-\rho_{[3]}-\rho_{[5]}$
γ_{21}	$\gamma_{[2]}$	$\cdot$	$\cdot$	0	$\rho_{[4]}$	$\rho_{[4]}$	$\rho_{[4]}+\rho_{[6]}$	$1-\rho_{[1]}-\rho_{[2]}-\rho_{[3]}-\rho_{[5]}$
γ_{12}	$\gamma_{[3]}$	$\cdot$	$\cdot$	$\cdot$	$\rho_{[4]}$	$\rho_{[4]}$	$\rho_{[4]}+\rho_{[6]}$	$1-\delta_{[1]}-\rho_{[3]}-\rho_{[5]}$
γ_{22}	$\gamma_{[4]}$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	0	$\rho_{[6]}$	$1-\delta_{[1]}-\rho_{[3]}-\rho_{[4]}-\rho_{[5]}$
γ_{13}	$\gamma_{[5]}$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\rho_{[6]}$	$1-\delta_{[1]}-\delta_{[2]}-\rho_{[5]}$
γ_{23}	$\gamma_{[6]}$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$1-\delta_{[1]}-\delta_{[2]}-\rho_{[5]}-\rho_{[6]}$

Table 12:

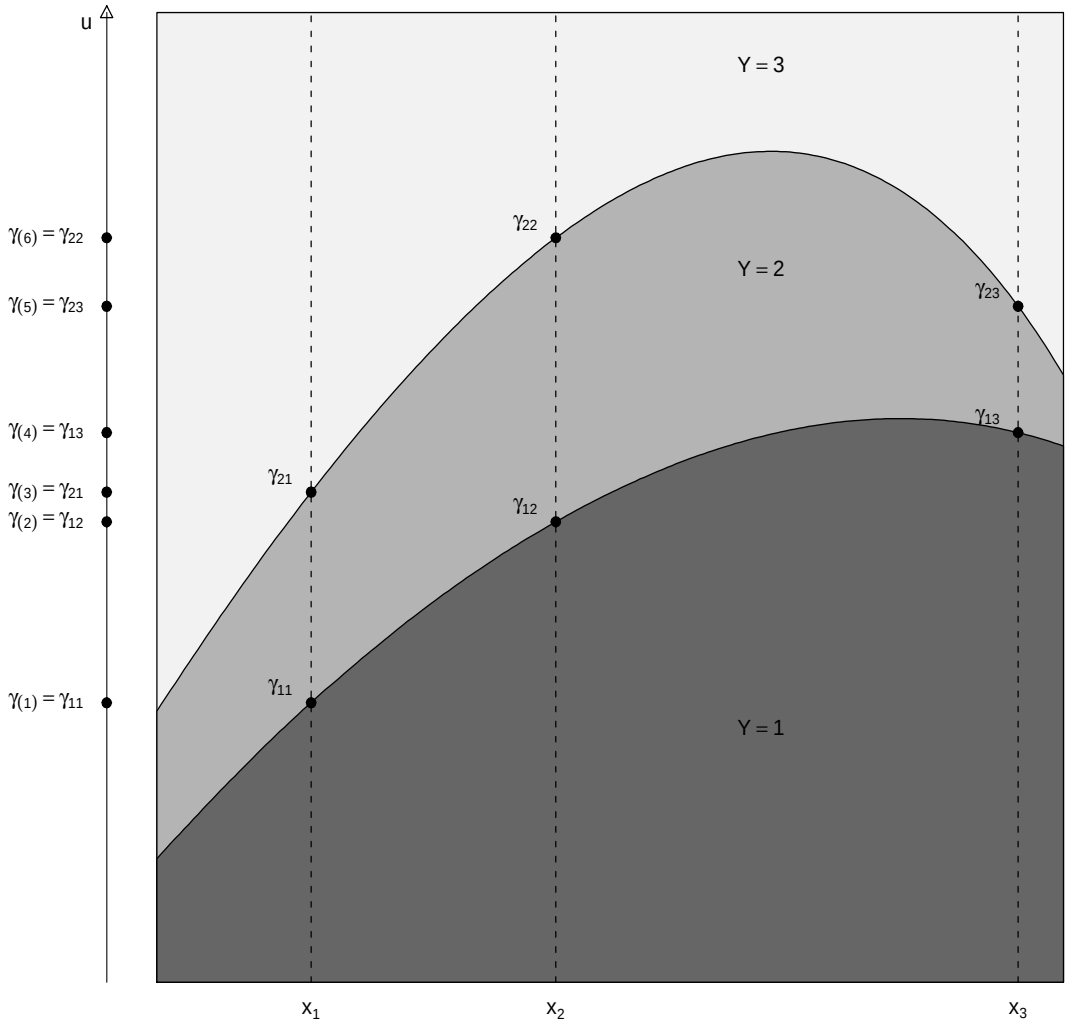

Figure 1: Examples of two threshold functions for the case $M = 3$ and $K = 3$ that are consonant with the arrangement of elements of γ shown on the vertical axis. The outcome Y takes the value 1 below the lowest threshold in the dark shaded region and the value 3 above the highest threshold in the light shaded region. The vertical scale is the unit interval $[0, 1]$.