cemmap
centre for microdata methods and practice

Single equation endogenous binary response models

Andrew Chesher

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP23/09

An ESRC Research Centre

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

Single equation endogenous binary response models

Andrew Chesher*
CeMMAP & UCL

March 18th 2009, revised August 17th 2009

Abstract. This paper studies single equation models for binary outcomes incorporating instrumental variable restrictions. The models are incomplete in the sense that they place no restriction on the way in which values of endogenous variables are generated. The models are set, not point, identifying. The paper explores the nature of set identification in single equation IV models in which the binary outcome is determined by a threshold crossing condition. There is special attention to models which require the threshold crossing function to be a monotone function of a linear index involving observable endogenous and exogenous explanatory variables. Identified sets can be large unless instrumental variables have substantial predictive power. A generic feature of the identified sets is that they are not connected when instruments are weak. The results suggest that the strong point identifying power of triangular "control function" models - restricted versions of the IV models considered here - is fragile, the wide expanses of the IV model's identified set awaiting in the event of failure of the triangular model's restrictions.

Keywords: Binary Response, Control functions, Endogeneity, Incomplete models, Index Restrictions, Instrumental variables, Probit Models, Set Identification, Threshold Crossing Models, Triangular Models.

JEL Codes: C10, C14, C50, C51.

1. Introduction

This paper explores the identifying power of single equation threshold-crossing models for a binary response Y generated by a structural equation as follows.

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

Here U is a scalar continuously distributed random variable. The models allow explanatory variables X to be endogenous and embody instrumental variable (IV) exclusion and independence restrictions. Probit and logit models with endogenous explanatory variables are familiar examples of parametric models to which the results of this paper apply. The analysis is essentially nonparametric but parametric restrictions are very easy to incorporate as will be demonstrated.

*I thank Martin Cripps, Sokbae Lee, Lars Nesheim, Adam Rosen and Richard Spady for stimulating comments and discussions and Konrad Smolinski for excellent research assistance. Some of the results given here were presented at seminars at Caltech, UCLA and USC in November 2007 and subsequently at the Malinvaud Seminar in Paris in December 2007, and at seminars at EUI, the University of Amsterdam, CeMMAP, the Cowles Foundation, Georgetown and Johns Hopkins Universities, at FESAMES, Tokyo 2009 and at ESEM Barcelona 2009. I gratefully acknowledge the financial support of the UK Economic and Social Research Council through a grant (RES-589-28-0001) to the ESRC Centre for Microdata Methods and Practice (CeMMAP).

These IV models place no restrictions on the genesis of endogenous explanatory variables. In this respect they are incomplete. One consequence of this is that the models are set not point identifying for deep structural features. One of the contributions of the paper is to characterize tight identified sets in nonparametric and parametric versions of the IV binary response model.

Set identification results are given for general discrete outcome IV models in Chesher (2007b, 2008). This paper studies the consequences of those results for the binary response model and considers refinements obtained when additional restrictions that may be available in the binary response case are imposed. Specifically it is shown that under additional monotonicity and single index restrictions concerning the impact of explanatory variables on the binary response it is possible to visualise the identified set of nonparametrically specified structural functions. Tight identification sets for the index coefficients are defined. Those sets can be determined by calculations in which the threshold crossing function plays no role which is computationally extremely beneficial.

Many complete models are restricted versions of the IV models studied here. A leading case of interest in view of its dominance in applied econometric practice is the triangular model which motivates widely used "control function" estimators. The software suites STATA and LIMDEP both provide commands to compute estimates using parametric versions of the control function model.[1]

Parametric and nonparametric control function models can be point identifying for deep structural features but even the nonparametric models rely on very strong restrictions concerning the genesis of potentially endogenous variables. The results of this paper allow one to see what alternative binary response structures are observationally indistinguishable from some triangular structure, possible well-supported by data, once the strong restrictions of the control function *model* are jettisoned.

A slightly depressing result of the paper is that in the endogenous binary response setting the identified sets delivered by an IV model can be large unless instrumental variables have substantial predictive power for the endogenous explanatory variables. This is in contrast to cases with less coarse discrete responses such as arise when studying ordered choice and interval censored outcomes. With continuous responses the IV model can be point identifying as shown in Chernozhukov and Hansen (2005).[2]

The message to take away from this is that in many cases with binary responses the restrictions of a point identifying triangular model which underpin control function estimation may contribute enormously to the determination of the results those estimators produce. Where those results are the basis for substantive decisions it may be prudent to consider the range of magnitudes delivered by observationally equivalent structures admitted by the less restrictive IV model, in the context of which the control function restrictions are not falsifiable. The results of this paper allow this to be done.

The triangular model and the control function idea are now briefly described. Then the encompassing single equation IV model is introduced and its identifying power is studied, first with no additional restrictions and then under a monotone index

[1] Statacorp (2007) and Greene (2007).

[2] The sensitivity of the identifying power of the IV model to varying amounts of discreteness in responses is the focus of Chesher and Smolinski (2009).

restriction in which the threshold-crossing function is a monotone function of a linear index involving all the observed explanatory variables. The results are illustrated with some exact calculations of identified sets of functions and, in parametric cases, of parameter values. Estimation of identified sets is discussed and illustrated in a small Monte Carlo experiment.

2. Triangular models and control functions

Let Y be a binary response and let X be a scalar, potentially endogenous, explanatory variable. A triangular model motivating control function estimation has structural equations as follows:

$$Y = \begin{cases} 0 \quad , & 0 \leq \; U \; \leq p(X) \\ 1 \quad , & p(X) < \; U \; \leq 1 \end{cases} \quad , \qquad X = g(Z, V) \tag{1}$$

with: (U, V) continuously distributed, the function h monotone in scalar U and the function g *strictly* monotone in scalar V, and $(U, V) \perp\!\!\!\perp Z$, a vector of instrumental variables that are excluded from p.[3] The system is *triangular* in the sense that Y does not feature in the structural equation for X.[4]

Since g is strictly monotone in V there is a one-to-one correspondence between V and X for every Z and a well defined single valued inverse function g^{-1} such that $V = g^{-1}(Z, X)$. This is known as the *control function*.

In this model X is endogenous if and only if U and V are dependently distributed. So, for variations in Y, X and Z such that $V = g^{-1}(Z, X)$ is held constant, U and X will vary independently[5] and the *ceteris paribus* effect of X on h can be identified if the function g can be identified. That is easily done. With V normalised $Unif(0, 1)$ the function g is identified as the conditional quantile function of X given Z and g^{-1} is the conditional distribution function of X given Z. See Matzkin (2003).

Since under the triangular model's restrictions $U \perp\!\!\!\perp X|V$ there is:

$$P[Y = 0|X = x, V = v] = P[U \leq p(x)|V = v] = F_{U|V}(p(x)|v)$$

which leads to the following.

$$P[Y = 0|X = x, Z = z] = F_{U|V}(p(x)|g^{-1}(z, x))$$

Here $F_{U|V}$ is the conditional distribution function of U given V. Control function estimation can proceed in a variety of ways, for example by estimating the regression

[3] The notation $A \perp\!\!\!\perp B$ has two interpretations. When B is a random variable it indicates that random variables A and B are mutually independently distributed. A and B may be vector random variables. When B is not a random variable, as would be the case if it took values purposively chosen by an expermenter or survey designer, then it indicates that the distribution function of A that applies when $B = b$ does not vary with b.

[4] If X were a vector the triangular model would have additional structural equations determining the value of X, none of them involving Y. See Chesher (2003) for an example.

[5] The triangular model comprising (1) and $(U, V) \perp\!\!\!\perp Z$ implies $U \perp\!\!\!\perp X|V$ because conditional on V, X varies only with Z and $(U, V) \perp\!\!\!\perp Z$ implies $U \perp\!\!\!\perp Z|V$.

function of Y given X and V, replacing V with a first round estimate derived from an estimate of g.

Control function methods in semi- and non-parametric settings are studied in Chesher (2003), Blundell and Powell (2003, 2004) and Imbens and Newey (2009) and in parametric settings in Rivers and Vuong (1988) and Smith and Blundell (1986). Hausman (1978) and Heckman (1979) contain early examples of the use of control function methods. STATA 10 (Statacorp (2007)) and LIMDEP 9.0 (Greene (2007)) have commands to perform parametric (probit) control function estimation in triangular models for binary responses.

The support of the covariate Z places limits on what can be known of h. Chesher (2003) shows that when X is continuous, with $x_r(z) \equiv g(z,r)$ which is identified by[6] $Q_{X|Z}(r|z)$, and with $u_p(r) \equiv Q_{U|V}(p|r)$ there is, under the restrictions of the triangular model:

$$Q_{Y|XZ}(p|Q_{X|Z}(r|z),z) = \begin{cases} 0 \ , & 0 \leq \ u_p(r) \ \leq p(x_r(z)) \\ 1 \ , & p(x_r(z)) < \ u_p(r) \ \leq 1 \end{cases} \tag{2}$$

from which it is clear that continuous variation in z is required if the threshold crossing function is to be nonparametrically identified and that limits on the extent of that variation may result in that function not being identifiable over ranges of its argument.

The restrictions of this triangular control function model can be violated in many ways, some of them set out below. In each case, relaxing the restrictions of the triangular model to accommodate the violation can result in loss of point identifying power.

1. **Discrete endogenous variable**. If X is discrete there is not a one-to-one correspondence between V and X for each value of Z and the control function does not control the value of V.[7]

2. **Direct dependence of U and X**. If U and X have dependence that arises not just through U's dependence on V then holding V fixed will not result in independent variation of U and X.

3. **Dependence between U and V is affected by Z**. If the joint independence restriction $(U,V) \perp\!\!\!\perp Z$ fails to hold, for example because the dependence between U and V varies with Z then the result $U \perp\!\!\!\perp X|V$ will fail to hold.

4. **Excess heterogeneity**. If V is not scalar, so that X is driven by more than one source of stochastic variation as in a random coefficients set up, then even if $(U,V) \perp\!\!\!\perp Z$ the model fails to identify p. One can always develop a reduced form equation for X involving a scalar error, say $\tilde{V}$, which is independent of Z but the condition $(U,\tilde{V}) \perp\!\!\!\perp Z$ will not hold in general.

[6] Here $V \perp\!\!\!\perp Z$ is normalised $Unif(0,1)$. The notation $Q_{A|B}(p|b)$ indicates the conditional p-quantile of random variable A given $B=b$.

[7] Chesher (2005) gives a set identification result in this case when there is a monotone variation restriction on the dependence of U on V, namely that $Q_{U|V}(p|v)$ be a monotone function of v.

5. **Full simultaneity**. If Y appears in the structural function g then even though a reduced form equation for X is available with scalar unobservable, say $\tilde{V}$, independent of Z the joint independence restriction $(U, \tilde{V}) \perp\!\!\!\perp Z$ will generally not hold. The simultaneous entry game model of Tamer (2003) provides an example.[8]

In each of these cases the triangular model fails to hold because it does not correctly specify some aspect of the process generating the endogenous explanatory variables. However in each case there remains as valid the instrumental variable restriction that U is distributed independently of instrumental variables Z which are excluded from the threshold-crossing function $p(X)$. The single equation IV model built on these restrictions concerning the genesis of the binary response Y *encompasses* the triangular model and extensions which accommodate the departures from the triangular model set out in 1 - 5 above.

The set identifying power of the single equation IV model is now considered.

3. Identifying power of the single equation IV binary response model

3.1. The single equation model. In the single equation IV model considered here the value of a binary variable $Y \in \{0, 1\}$ is uniquely determined by a structural function as follows.

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

Here U is an unobserved scalar continuously distributed random variable and X is a vector random variable which may be jointly dependently distributed with U. To the extent that there is dependence between X and U then elements of X are endogenous. The marginal distribution of U is normalised to be uniform on $(0, 1)$.

There are instrumental (exogenous) variables arranged in a vector Z. The model excludes these variables from p and imposes the restriction that U and Z are jointly independently distributed. For most of the analysis Z need not be regarded as a random variable and then the restriction $P[U \leq u|Z = z] = u$ for all $z \in \mathcal{Z}$ is imposed, which embodies the uniform marginal distribution normalisation. When Z is a random variable the set $\mathcal{Z}$ is the support of the random variable Z; otherwise it is a set of valid instrumental values[9] of Z.

In what follows, because all probabilities are conditioned on Z, the instrumental variables can appear as arguments of the threshold-crossing function p. Of course for a model to have informative identifying power it will have to embody some restriction on the impact of Z on p. Z will appear as an argument of p when index restrictions are considered in Section 6 but for now, mainly to simplify notation, the model will contain the restriction that Z is excluded from p.

The identifying power of this single equation model is now considered, answering the question: what can be known of the function p from knowledge of the probability distribution of Y and X given $Z = z$ when z varies within $\mathcal{Z}$?

[8] Each of Tamer's equations taken one-at-a-time along with the marginal independence restrictions implied by his model satisfy the restrictions of the single equation IV model.

[9] Chesher (2007a) gives an analysis of identification in terms of instrumental values.

4. Set identification

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

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

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

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

Consider a data generating structure $S_0 \equiv \{p_0, F^0_{UX|Z}\}$ admitted by this model in which p_0 is a threshold-crossing function and $F^0_{UX|Z}$ denotes a joint *distribution* function for U and X given Z.

To be admitted by the model $\mathcal{M}$ the distribution function $F^0_{UX|Z}$ must respect the independence property, that is:

$$F^0_{U|Z}(u|z) \equiv F^0_{UX|Z}(u, \bar{x}|z) = u$$

for all $u \in (0,1)$ and $z \in \mathcal{Z}$. Here $\bar{x}$ is the upper limit of the support of X.

Let $F^0_{YX|Z}$ denote the joint distribution function of Y and X given Z generated by the structure S_0, determined as follows.[10]

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

Let $\Pr_0$ indicate probabilities calculated with respect to this measure. Observationally equivalent structures, S^*, have threshold crossing functions p_* and distribution functions $F^*_{UX|Z}$ such that

$$F^*_{YX|Z}(0, x|z) \equiv F^*_{UX|Z}(p_*(x), x|z) = F^0_{YX|Z}(0, x|z)$$

for all $x \in \text{supp}(X)$ and $z \in \mathcal{Z}$.

Theorem 1 gives a system of inequalities which is satisfied by all threshold-crossing functions in admissible structures that are observationally equivalent to an admissible structure generating a joint distribution function $F^0_{YX|Z}$ for all values of instrumental variables z in some set of values $\mathcal{Z}$.

Theorem 1

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

$$c_{0l}(u, z; p) \equiv \Pr_0[Y = 0 \cap p(X) < u | Z = z] < u \tag{3}$$

$$c_{0u}(u, z; p) \equiv 1 - \Pr_0[Y = 1 \cap u \leq p(X) | Z = z] \geq u \tag{4}$$

[10] Since Y is binary, $F^0_{YX|Z}(1, x|z) = F^0_{X|Z}(x|z)$.

Here subscripts "l" and "u" indicate respectively *lower* and *upper* bounding probability functions. The subscript "0" indicates that a function (c_{0l} or c_{0u}) is calculated using the distribution function $F^0_{YX|Z}$ generated by the structure S_0. Note that, because there is conditioning on $Z = z$, Theorem 1 continues to hold when the threshold-crossing function p includes Z as an argument.

Proof of Theorem 1

It is first shown that (3) and (4) hold for all $u \in (0, 1)$ and all $z \in \mathcal{Z}$ when $p = p_0$.

Consider the inequality (3) with $p = p_0$. For all x such that $p_0(x) \geq u$,

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

and for all x such that $p_0(x) < u$:

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

and so for all x there is the following inequality.

$$\Pr_0[Y = 0 \cap p_0(X) < u|X = x, Z = z] < \Pr_0[U \leq u|X = x, Z = z]$$

Taking expected value over X given $Z = z$ yields the inequality (3) with $p = p_0$.

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

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

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

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

and so for all x there is the following inequality.

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

Taking expected value over X given $Z = z$ yields the inequality (4) with $p = p_0$.

The result of the Theorem now follows directly since if some p_* is an element of a structure *observationally equivalent* to S_0 then it generates the *same* probability measure as S_0 does, so (3) and (4) hold for all $u \in (0, 1)$ and all $z \in \mathcal{Z}$ when $p = p_*$. ■

Theorem 1 states that all the threshold-crossing functions identified by the single equation IV model lie in the set of functions defined by the inequalities (3) and (4) as u varies across $(0, 1)$ and z varies across $\mathcal{Z}$. Chesher (2008) shows that, when X is continuous this set is the *sharply defined identified set*, that is all functions in the set are elements of observationally equivalent structures admitted by the model $\mathcal{M}$.[11] This is also true when X is discrete, the case considered in the next Section.

The functions c_{0l} and c_{0u} in (3) and (4) are non-decreasing in u and satisfy inequalities:

$$0 \leq c_{0l}(u, z; p) \leq F^0_{Y|Z}(0|z) \leq c_{0u}(u, z; p) \leq 1 \tag{5}$$

which hold for all z and $u \in (0, 1)$. Each function attains its lower and upper bounds as u approaches respectively 0 and 1.

Since the inequalities (3) and (4) hold for all $z \in \mathcal{Z}$ a threshold-crossing function p is in the identified set if and only if for all $u \in (0, 1)$

$$c_{0l}(u; p) \equiv \max_{z \in \mathcal{Z}} c_{0l}(u, z; p) < u \leq \min_{z \in \mathcal{Z}} c_{0u}(u, z; p) \equiv c_{0u}(u; p) \tag{6}$$

The functions $c_{0u}(u; p)$ and $c_{0l}(u; p)$ are non-decreasing functions of u and it follows from (5) that for all u and any admissible p.

$$\begin{aligned} 0 &\leq c_{0l}(u; p) \leq \max_{z \in \mathcal{Z}} F^0_{Y|Z}(0|z) \\ \min_{z \in \mathcal{Z}} F^0_{Y|Z}(0|z) &\leq c_{0u}(u; p) \leq 1 \end{aligned}$$

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

Given a particular distribution for Y and X given Z and a set of instrumental values, $\mathcal{Z}$, a putative threshold-crossing function p can be assigned to the identified set of structural functions by calculating the functions $c_{0u}(u; p)$ and $c_{0l}(u; p)$ and observing whether the inequalities (6) are satisfied for all $u \in (0, 1)$.

A restricted version of the model $\mathcal{M}$ may require the threshold-crossing function, p, to be a member of a parametric family of functions. Later the case in which $p(x)$ has the "probit" form $\Phi(\alpha_0 + x'\alpha_1)$ is considered. When there are parametric restrictions the inequalities (3) and (4) sharply define the identified set of values of parameters associated with the distribution $F^0_{YX|Z}$ and the model $\mathcal{M}$.

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

The force of this semiparametric restriction is considered in Section 5. It leads to a result which allows visualisation of identified sets of nonparametrically specified

[11] Chesher (2008) works with a structural equation $Y = h(X, U)$ for a general discrete outcome and with probabilities $\text{Pr}_0[Y < h(X, u)|Z = z]$ and $\text{Pr}_0[Y \leq h(X, u)|Z = z]$. In the binary Y case studied here these probabilities are expressed in terms of the threshold-crossing function p associated with a structural function h using the following identities.

$$\{Y < h(X, u)\} = \{Y = 0 \cap p_0(X) < u\}$$

$$\{Y \leq h(X, u)\} = \overline{\{Y > h(X, u)\}} = \overline{\{Y = 1 \cap u \leq p(X)\}}$$

monotone structural functions and characterisation of identified sets of values of index coefficients.

First the case in which endogenous variables are discrete is considered and a proof of sharp set identification for that case is provided.

4.1. Discrete Endogenous Variables. The probability inequalities that appear in Theorem 1 are now given explicit representations for the case in which X is discrete.

Let X have support $\{x_1, x_2, \ldots, x_K\}$ and for $k \in \{1, \ldots, K\}$ define $\gamma_k \equiv p(x_k)$ and define $\gamma_0 \equiv 0$ and $\gamma_{K+1} \equiv 1$. Since X is discrete its dimensionality is irrelevant. In this discrete endogenous variable case it is the identification of the finite dimensional sequence $\gamma \equiv \{\gamma_k\}_{k=1}^K$ that is of interest.

The set of values of γ defined by the probability inequalities of Theorem 1 is the union of $K!$ convex sets, one associated with each permutation of γ. The sets are in general disconnected and when instruments are strong many of them can be empty. To proceed, without loss of generality let indices be assigned to the points of support of X so that: $\gamma_1 \leq \gamma_2 \leq \cdots \leq \gamma_K$. Other permutations can be accommodated by exchanging indices in what follows.

Now the inequalities (3) and (4) are expressed in terms of the γ_k's.

For $k \in \{1, \ldots, K\}$ define:

$$\alpha_k(z) \equiv \Pr[Y = 0|X = x_k, Z = z] \qquad \delta_k(z) \equiv \Pr[X = x_k|Z = z]$$

and adopt the convention that sums from 1 to 0 are zero, $\sum_{s=1}^0 (\cdot)_s \equiv 0$.

The values $\{\delta_k(z)\}_{k=1}^K$ are required to all be non-zero for all $z \in \mathcal{Z}$ in what follows. This restriction could easily be relaxed but at the cost of some additional complexity in the notation.

The probabilities that appear in the inequalities (3) and (4) have values that depend upon the location of u in the sequence $\gamma_1, \ldots . \gamma_K$. Consider a value $u \in (0, 1)$ such that for two elements in γ, γ_{k-1} and γ_k, $\gamma_{k-1} < u \leq \gamma_k$. In this case the probabilities are as follows.[12]

$$\Pr_0[Y = 0 \cap p(X) < u|Z = z] = \sum_{j=1}^{k-1} \delta_j(z)\alpha_j(z) \tag{7}$$

[12] The first expression arises because

$$\begin{aligned} \Pr_0[Y = 0 \cap p(X) < u|Z = z] &= \Pr_0[Y = 0 \cap X \in \{x : p(x) < u\}|Z = z] \\ &= \sum_{j=1}^{k-1} \Pr_0[Y = 0|X = x_j, Z = z] \Pr_0[X = x_j|Z = z] \end{aligned}$$

and the second expression arises from

$$\begin{aligned} 1 - \Pr_0[Y = 1 \cap u \leq p(X)|Z = z] &= 1 - \Pr_0[Y = 1 \cap X \in \{x : u \leq p(x)\}|Z = z] \\ &= 1 - \sum_{j=k}^{K} \Pr_0[Y = 1|X = x_j, Z = z] \Pr_0[X = x_j|Z = z] \\ &= 1 - \sum_{j=k}^{K} \delta_j(z)(1 - \alpha_j(z)) \end{aligned}$$

after substituting $1 - \sum_{j=k}^K \delta_j(z) = \sum_{j=1}^{k-1} \delta_j(z)$.

$$1 - \text{Pr}_0[Y = 1 \cap u \leq p(X)|Z = z] = \sum_{j=1}^{k-1} \delta_j(z) + \sum_{j=k}^{K} \delta_j(z)\alpha_j(z) \tag{8}$$

If there are many elements in γ equal to γ_k then their associated values $\delta_j(z)\alpha_j(z)$ all contribute to the summation from k to K in (8). If there are many elements in γ equal to γ_{k-1} then their associated values $\delta_j(z)$ and $\delta_j(z)\alpha_j(z)$ all contribute to the summations from 1 to $k-1$ in (7) and (8).

When $\gamma_{k-1} < u \leq \gamma_k$ the inequality (3) requires (i) that the probability $\text{Pr}_0[Y = 0 \cap p(X) < u|Z = z]$ in (7) be less than u for all values of u in that interval, so the inequality

$$\sum_{j=1}^{k-1} \delta_j(z)\alpha_j(z) \leq \gamma_{k-1} \tag{9}$$

must hold, and (ii) that the probability $1 - \text{Pr}_0[Y = 1 \cap u \leq p(X)|Z = z]$ in (8) be at least equal to u for all values of u in that interval, so the inequality:

$$\sum_{j=1}^{k-1} \delta_j(z) + \sum_{j=k}^{K} \delta_j(z)\alpha_j(z) \geq \gamma_k \tag{10}$$

must hold. The inequality (7) holds for all values in γ that are equal to γ_{k-1} and the inequality (8) holds for all values in γ that are equal to γ_k.

Bringing these two results together and replacing $k-1$ by k in (9) delivers the following sequence of inequalities for $k \in \{1, \ldots, K\}$.[13]

$$\sum_{j=1}^{k} \delta_j(z)\alpha_j(z) \leq \gamma_k \leq \sum_{j=1}^{k-1} \delta_j(z) + \sum_{j=k}^{K} \delta_j(z)\alpha_j(z) \tag{11}$$

If $\gamma_1 = \gamma_2 = \cdots = \gamma_K$ are equal, say to some value $\bar{\gamma}$, then the left *and* the right hand sides of all the inequalities (11) are all equal to

$$\sum_{j=1}^{K} \delta_j(z)\alpha_j(z) = \text{Pr}_0[Y = 0|Z = z] \equiv \bar{\alpha}(z)$$

and only $\bar{\gamma} = \bar{\alpha}(z)$ is admissible. If $\bar{\alpha}(z)$ varies at all as z varies in $\mathcal{Z}$ then there is no admissible constant value $\bar{\gamma}$ and threshold crossing functions $p(x)$ which do not vary with x are not admissible.

Inequality sequences for different permutations of γ are obtained by exchange of indices. For each permutation the set of values of γ defined by the inequalities (11) is precisely the subset of the identified set for the model $\mathcal{M}$ associated with the permutation. This is the subject of Theorem 2.

[13] Here the summation from 1 to k on the left hand side and the summation from k to K on the right hand side include contributions $\delta_j(z)\alpha_j(z)$ associated with all elements of γ equal to γ_k. The summation from 1 to $k-1$ on the right hand side includes contributions $\delta_j(z)$ from *all* elements of γ having values less than γ_k.

Theorem 2

For every sequence $\gamma_1 \leq \gamma_2 \leq \cdots \leq \gamma_K$ for which the system of inequalities (11) holds, there exists a distribution function $F_{U|XZ}$ such that the following conditions hold for $k \in \{1, \ldots, K\}$ and each $z \in \mathcal{Z}$.

1. *Proper conditional distribution functions:*

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

2. *Independence:*[14]

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

3. *Observational equivalence:*

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

A proof is given in the Annex to the paper.

5. Monotonicity and index restrictions

The force of a restriction requiring the threshold function to be *monotone* is now studied.

First the case in which X is scalar is considered. The threshold function is specified as $p(x)$ with p monotone but with *no* restriction on the direction of the dependence on x. The identified set of threshold functions is shown to comprise all monotone functions that lie between pairs of bounding functions; one pair is increasing, the other pair is decreasing. These functions are shown to be simple functionals of the joint distribution of the binary outcome and the endogenous variable.

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

The identified set of threshold functions is the intersection of sets determined by pairs of upper and lower bounding functions. Each distinct value of the instrumental variables generates a pair of bounds. A procedure for estimating sets defined by intersection bounds is applied to this problem and studied in a small Monte Carlo experiment.

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

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

set, bounding functions are derived which define the set of functions associated with (α^*, δ^*). These are simply obtained by applying the methods derived for the scalar X case, replacing the random variable X in that analysis by the random variable $X'\alpha^* + Z'\delta^*$. Under the monotonicity restriction there is no need to consider particular alternative functions p when developing the identified set of index coefficients which substantially simplifies the computation and estimation of that identified set.

5.1. Monotone threshold functions with scalar X. Let p^{-1} denote the inverse function of p.[15] When the threshold function is restricted to be monotone and X is scalar, events such as $\{u > p(X)\}$ can be expressed as $\{p^{-1}(u) > X\}$ if p is increasing and as $\{p^{-1}(u) < X\}$ if p is decreasing. Then the bounding functions in (3) and (4) can be written as follows when p is *increasing*:

$$\begin{aligned} c_{0l}(u, z; p) &= \text{Pr}_0[\{Y = 0\} \cap \{X < p^{-1}(u)\} \, |z] < u && (12) \\ c_{0u}(u, z; p) &= 1 - \text{Pr}_0[\{Y = 1\} \cap \{X \geq p^{-1}(u)\} \, |z] \geq u && (13) \end{aligned}$$

and as follows when p is *decreasing*.

$$\begin{aligned} c_{0l}(u, z; p) &= \text{Pr}_0[\{Y = 0\} \cap \{X > p^{-1}(u)\} \, |z] < u && (14) \\ c_{0u}(u, z; p) &= 1 - \text{Pr}_0[\{Y = 1\} \cap \{X \leq p^{-1}(u)\} \, |z] \geq u && (15) \end{aligned}$$

Substituting $\sigma = p^{-1}(u)$ the threshold function is moved out of the bounding functions to the right hand sides of the inequalities. The resulting bounding functions and inequalities written in terms of $\sigma \in \text{supp}(X)$ are, for increasing p, thus:

$$\begin{aligned} d_{0l}^{\uparrow}(\sigma, z) &= \text{Pr}_0[\{Y = 0\} \cap \{X < \sigma\} \, |z] < p(\sigma) && (16) \\ d_{0u}^{\uparrow}(\sigma, z) &= 1 - \text{Pr}_0[\{Y = 1\} \cap \{X \geq \sigma\} \, |z] \geq p(\sigma) && (17) \end{aligned}$$

and for decreasing p, thus.

$$\begin{aligned} d_{0l}^{\downarrow}(\sigma, z) &= \text{Pr}_0[\{Y = 0\} \cap \{X > \sigma\} \, |z] < p(\sigma) && (18) \\ d_{0u}^{\downarrow}(\sigma, z) &= 1 - \text{Pr}_0[\{Y = 1\} \cap \{X \leq \sigma\} \, |z] \geq p(\sigma) && (19) \end{aligned}$$

It is very convenient to have the threshold-crossing function pulled out of the bounding functions in this fashion because the bounding functions can be derived or estimated just once and then compared with any candidate threshold function, leading to visualisation of the identified set of threshold-crossing functions. An illustration follows shortly.

The functions $d_{0l}^{\uparrow}$ and $d_{0u}^{\uparrow}$ are increasing in σ; the functions $d_{0l}^{\downarrow}$ and $d_{0u}^{\downarrow}$ are decreasing in σ. There are the following inequalities with left and right hand bounds achieved as σ approaches respectively $-\infty$ and $+\infty$.

$$0 \leq d_{0l}^{\uparrow}(\sigma, z) \leq \text{Pr}_0[Y = 0|z] \leq d_{0u}^{\uparrow}(\sigma, z) \leq 1 \qquad (20)$$

[15] For weakly monotonic functions p, define p^{-1} as follows

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

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

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

$$1 \geq d_{0u}^{\downarrow}(\sigma, z) \geq \Pr_0[Y = 0|z] \geq d_{0l}^{\downarrow}(\sigma, z) \geq 0 \tag{21}$$

An *increasing* function p is in the identified set of threshold functions if it satisfies (16) and (17) for all $\sigma \in \text{supp}(X)$ and all $z \in \mathcal{Z}$. A *decreasing* function p is in the identified set of threshold functions if it satisfies (18) and (19) for all $\sigma \in \text{supp}(X)$ and all $z \in \mathcal{Z}$. The identified set of threshold functions is therefore the union of two sets of functions: one comprising all *increasing* functions p that satisfy

$$\max_{z \in \mathcal{Z}} d_{0l}^{\uparrow}(\sigma, z) \equiv d_{0l}^{\uparrow}(\sigma) < p(\sigma) \leq d_{0u}^{\uparrow}(\sigma) \equiv \min_{z \in \mathcal{Z}} d_{0u}^{\uparrow}(\sigma, z) \tag{22}$$

for all $\sigma \in \text{supp}(X)$, the other comprising all *decreasing* functions p that satisfy

$$\max_{z \in \mathcal{Z}} d_{0l}^{\downarrow}(\sigma, z) \equiv d_{0l}^{\downarrow}(\sigma) < p(\sigma) \leq d_{0u}^{\downarrow}(\sigma) \equiv \min_{z \in \mathcal{Z}} d_{0u}^{\downarrow}(\sigma, z) \tag{23}$$

for all $\sigma \in \text{supp}(X)$. One of these sets may be empty and this will tend to happen when instruments are strong with rich support as illustrated shortly.

If a model further restricts p to lie in a parametric family then only parameter values leading to functions in the family that lie within the set defined by (22) and (23) fall in the identified set of parameter values. Parametric (probit) restrictions are considered shortly in an illustrative example.

In view of the definitions of the tight inequalities (20) and (21) there are the following results on the large and small σ behaviour of the tight bounding functions.

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

Here the inequality is strict unless z has *no effect* on $\Pr_0[Y = 0|z]$ for all $z \in \mathcal{Z}$.

It follows that the constant function $p(\sigma) = c$ free of σ *does not lie in the identified set* unless the structure generating the probability measure has $\Pr_0[Y = 0|z]$ for all $z \in Z$. However there can be both increasing *and* decreasing functions in the identified set. In this respect the identified set may be disconnected. When p is parametrically restricted this leads to identified sets of parameter values which may be disconnected. The next Section illustrates.

This, at first sight, paradoxical result arises because the exclusion and independence restrictions of the model require that Z affects Y only *via* the endogenous X so even the smallest dependence of the outcome Y on the instruments Z implies that the threshold function delivering the value of the binary outcome *does depend on the endogenous variable*. However, with sufficiently feeble instruments, threshold functions exhibiting positive or negative dependence on elements of X are capable of delivering the probability measure used to calculate the identified set which in consequence can be disconnected.

5.2. Illustration. These results are illustrated using probability measures generated by a triangular Gaussian structure which satisfies the restrictions of the single equation IV model. The structural function for binary Y has a probit form with an endogenous explanatory variable. This choice makes the calculation of the bounding functions easy, it highlights the relative power of the control function model which would be point identifying in this case, and it places us in familiar applied econometrics territory.[16]

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

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

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

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

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

from which it is straightforward to calculate the probabilities that appear in (16) - (19) as bivariate normal orthant probabilities.[17] The (monotone) threshold function for the structures employed in this example is $p(x) = \Phi(-a_0 - a_1 x)$ where Φ denotes the standard normal distribution function.

5.3. Nonparametric model. The identifying power of the following nonparametric model is considered.

$$Y = h(X, U) = \begin{cases} 0 \ , & 0 \leq U \leq p(X) \\ 1 \ , & p(X) < U \leq 1 \end{cases} \qquad U \perp\!\!\!\perp Z \qquad p \text{ monotone} \qquad (24)$$

The following graphs show the bounding functions (16) - (19) varying with σ for specific values of the instrument z and their envelope functions, (22) and (23). The functions are calculated using the probability measure generated by Gaussian triangular structures as defined above with the parameter values as shown in the first row of Table 1 and with z taking 10 equally spaced values in $[-1, +1]$. At these parameter values the structural threshold function is the standard normal distribution function $\Phi(x)$.

In Figure 1 the value of b_1, the coefficient on the instrumental variable in the equation for endogenous X, is 0.3. The upper pane shows the increasing bounding functions (16) - (17); the lower pane shows the decreasing functions (18) - (19). These functions are drawn in blue.

The envelope bounding functions (22) and (23) are obtained at each value of σ as the maximum of the lower bounding functions and the minimum of the upper bounding functions. They are drawn as dashed red lines. The identified set of structural threshold functions comprises all increasing functions which pass between the upper and lower envelope bounding functions in the upper pane and all decreasing

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

[17] The function pmvnorm in the mvtnorm package of R (Ihaka and Gentleman (1996)) is used.

Parameter	a_0	a_1	b_0	b_1	s_{wv}	s_{vv}
Figure 1	0	-1	0	0.3	0.5	1
Figure 2	0	-1	0	0.4	0.5	1
Figure 3	0	-1	0	0.3	0.05	0.1
Figure 4	0	-1	0	0.3	0.005	0.01
Figure 5	0	-1	0	0.3	0.0005	0.001
Figure 6	0	-1	0	0.6	0.0005	0.001
Figure 7	0	-1	0	0.3 0.6 0.9 1.2 1.5	0.15	0.3

Table 1: Parameter values used in Figures 1 - 7

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

The structural threshold function in the Gaussian triangular structure used to generate the probability measure employed in these calculations, is the increasing dashed line passing between the upper and lower bounding functions in the upper pane. *Any* monotone increasing (decreasing) function passing between the red dashed lines in the upper (lower) pane in Figure 1, together with a suitable chosen (typically non-Gaussian) distribution for U and X given Z *also* generates the *same* probability measure. A construction for producing one such distribution for U and X given Z is given in Chesher (2008).

When the power of the instrument is increased by setting the parameter $b_1 = 0.4$ the identified set is reduced as shown in Figure 2. The envelope bounding functions in the lower pane now intersect - no decreasing function can pass between these functions. The effect of the instrument is now sufficiently strong to eliminate *all* monotone decreasing functions from the identified set. However the identified set of increasing functions is little affected.

For Figure 3 the coefficient b_1 is reset to its Figure 1 value, 0.3, and the strength of the instrument is increased by drastically raising its *predictive power*, a situation achieved by reducing s_{vv} tenfold, from 1 to 0.1, while reducing s_{wv} to 0.05 so that the correlation between W and V is unchanged at 0.25. This strengthening of the instrument also serves to remove decreasing functions from the identified set and produces a noticeable narrowing of the bounds around increasing functions but the situation is still a long way from point identification even with this small value of s_{vv}.

To investigate the extreme situation that arises as the instrument approaches the state of being a perfect predictor of endogenous X, Figures 4 and 5 show the effect of further substantial increases in the predictive power of the instrument with $s_{vv} = 0.01$ in Figure 4 and $s_{vv} = 0.001$ in Figure 5. The correlation between W and V is kept constant at 0.25 as s_{vv} is reduced. The bounds narrow very considerably but there is still some significant degree of variation in functions within the identified set. In Figure 5 particularly it is clear that the support restriction on Z is very influential. In the example Z is restricted to lie within $[-1,1]$ and takes a coefficient of 0.3 and with the very small value of s_{vv} there is only information delivered about the structural function by the probability measure (as opposed to the monotonicity restriction) for $\sigma \in [-.3, .3]$. Outside this range essentially every increasing function lies in the identified set.

Figures 6 shows the effect on the identified set shown in Figure 5 of doubling b_1 (to 0.6), the coefficient on Z in the equation for X in the triangular structure used here to generate probability measures. At this very small value of s_{vv} the only effect this has is to extend, by a factor of 2, the range of values of σ effectively covered by the identified set.

Figure 7 shows the effect of changing b_1 with the other parameter values set as shown in the final row of Table 1. As b_1 is increased the identified set is reduced in extent but the effect is virtually all at extreme rather than central values of σ. Changing b_1 is equivalent to changing the units of measurement of, and so the range, of Z. This increases the range of values of X for which the identified set is informative but has almost no effect on the width of the identified set over that range.

5.4. Estimation. This Section considers estimation of the identified set of threshold functions when they are restricted to be monotone. The method proposed in Chernozhukov, Lee and Rosen (2009) (henceforth in this Section CLR) is employed. The procedure delivers approximately pointwise median unbiased estimators of bounding functions when these are, as here, defined as an infimum or supremum over a set of functions whose members correspond to different values of instrumental variables. It also produces an approximate confidence region for the identified set of functions.

The method is easiest described in the context of estimation of one of the bounding functions introduced in Section 5.1. So, consider the upper bounding function for monotone increasing functions defined in equation (17) as:

$$d_{0u}^{\uparrow}(\sigma) \equiv \min_{z \in \mathcal{Z}} d_{0u}^{\uparrow}(\sigma, z)$$

where $d_{0u}^{\uparrow}(\sigma, z)$ is defined as follows.

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

In the case studied here there are n independent realisations from $F_{YX|Z}^0$, $\{Y_i, X_i\}_{i=1}^n$, at instrumental values $\{Z_i\}_{i=1}^n$. Inferences are made conditional on these instrumental values. There are K values of Z, that is $\mathcal{Z} = \{z_{(1)}, \ldots, z_{(K)}\}$, with $n_k \equiv \sum_{i=1}^n 1[Z_i = z_{(k)}]$ realisations at $z_{(k)}$ and $\sum_{k=1}^K n_k = n$. The analysis is carried out pointwise in the argument σ.

At each value of σ there are the following analog estimates.

$$\hat{d}_{0u}^{\uparrow}(\sigma, z_{(k)}) \equiv 1 - n_k^{-1} \sum_{i=1}^{n} 1[Y_i = 0 \cap X_i \geq \sigma \cap Z_i = z_{(k)}]$$

It is noted in CLR that the naive estimator of $d_{0u}^{\uparrow}(\sigma)$ defined as the infimum of these estimators over $k \in \{1, \ldots, K\}$ is downward biased. A bias-corrected estimator is proposed in CLR, obtained by adding a correction term to each estimator $\hat{d}^{\uparrow}_{0u}(\sigma, z_{(k)})$, which depends on the precision of each estimate, less precise terms receiving larger positive corrections.

In an initial step a data-dependent subset of the estimates is chosen. This converges in probability to a non-stochastic set which contains the infimum of the set: $\{d^{\uparrow}_{0u}(\sigma, z_{(k)})\}_{k=1}^{K}$. Let $\hat{\mathcal{K}}$ denote the subset of $\{1, \ldots, K\}$ which indexes the estimators contained in this set. The following estimator is suggested in CLR:

$$\hat{\mathcal{K}} = \{k \in \{1, \ldots, K\} : \hat{d}^{\uparrow}_{0u}(\sigma, z_{(k)}) \leq \min_{k \in K} \{\hat{d}^{\uparrow}_{0u}(\sigma, z_{(k)})\}_{k=1}^{K} + \lambda_n$$

for some choice λ_n where $\lambda_n \to 0$ and $n^{1/2}\lambda_n \to \infty$.

The bias corrected estimator is the infimum of the precision adjusted estimators that have indexes $k \in \hat{\mathcal{K}}$. In the case studied here the correction term applied to an estimator $\hat{d}^{\uparrow}_{0u}(\sigma, z_{(k)})$ is a constant $\kappa(p)$ multiplied by an estimate of the standard error of the estimators $\hat{d}^{\uparrow}_{0u}(\sigma, z_{(k)}), k \in \hat{\mathcal{K}}$. Since these are a linear transformations of binomial random variables the estimated (squared) standard errors are as follows:

$$n_k^{-1} \hat{d}^{\uparrow}_{0u}(\sigma, z_{(k)}) \left(1 - \hat{d}^{\uparrow}_{0u}(\sigma, z_{(k)})\right) \qquad k \in \hat{\mathcal{K}}$$

and the estimators are asymptotically independently normally distributed.

In this circumstance the factor $\kappa(p)$ is the p-quantile of the maximum of $\hat{K}$ standard independent normal variates where $\hat{K}$ is the number of indexes in $\hat{\mathcal{K}}$. Define $\tilde{d}^{\uparrow}_{0u(p)}(\sigma)$ as follows.

$$\tilde{d}^{\uparrow}_{0u(p)}(\sigma) \equiv \min_{k \in \hat{\mathcal{K}}} \left\{ \hat{d}^{\uparrow}_{0u}(\sigma, z_{(k)}) + \kappa(p) \times \left(n_k^{-1} \hat{d}^{\uparrow}_{0u}(\sigma, z_{(k)}) \left(1 - \hat{d}^{\uparrow}_{0u}(\sigma, z_{(k)})\right)\right)^{1/2} \right\}$$

Choosing $p = 0.5$ yields an approximately median unbiased estimator of $d^{\uparrow}_{0u}(\sigma)$. Choosing $p = 1 - \alpha$ yields an approximate one sided $(1 - \alpha)$ confidence region for $d^{\uparrow}_{0u}(\sigma)$.[18] Regularity conditions, propositions and proofs and a more thorough explanation are given in CLR.

Tables 3 - 6 report results of four Monte Carlo experiments (1000 replications each) intended to demonstrate the feasibility of estimating bounding functions and identified sets of functions in moderate sized samples using the CLR procedure.

The setting is as in Section 5.2 with the parameter values given in the row headed "Figure 1" in Table 1 and with either 5 or 10 values of z equally spaced in $[-1, 1]$. In experiments MC1 and MC2 the sample size is $n = 100$; in experiments MC3 and MC4 the sample size is $n = 400$. In experiments MC1 and MC3 there are $K = 5$ equally spaced values of Z with equal numbers of realisations at each value. In experiments MC2 and MC4 there are $K = 10$ equally spaced values in Z with equal number of realisations on each value. The value of the tuning parameter is adjusted accordingly.

Tables 3 - 6 show in the upper part, results for increasing bounding functions and in the lower part results for decreasing bounding functions. Results for lower and upper bounding functions are shown respectively to the left and the right. Results

[18] In the sense that for large n, $P[d^{\uparrow}_{0u}(\sigma) \leq \tilde{d}^{\uparrow}_{0u(1-\alpha)}(\sigma)] \geq 1 - \alpha$.

are given for estimates at five values of σ (the argument of the monotone threshold function). Columns headed $d_l^{\uparrow}(\sigma)$ etc., give exact values of bounding functions at the chosen values of σ. Columns headed b_N and b_C give Monte Carlo estimates of median bias (multiplied by 100) of respectively the naive (N) estimator and the bias corrected (C) estimator at the indicated value of σ. Columns headed $RMSE_N$ and $RMSE_C$ give Monte Carlo estimates of the root mean squared error of respectively the naive (N) estimator and the bias corrected (C) estimator.

Figures 12 - 15 accompany the tables of results and are helpful in interpreting them. Upper and lower panes show results for estimated sets of respectively increasing and decreasing functions. In each case lines coloured brown show medians of naive estimates across the 1000 replications, calculated pointwise over values of σ. In all cases expect MC3 these suggest that the identified set contains *no* decreasing functions. The shaded blue areas indicate the median position of the boundaries of the identified sets once the bias correction is applied. In all cases these *bias adjusted* estimated sets contain increasing and decreasing functions. Red dashed curves show the exact bounding functions. It is evident that the bias correction is quite effective. Finally the outer grey lines show the pointwise median position of upper and lower one sided 95% confidence regions for the estimated bounding functions.

The bias in the naive estimators is quite substantial when there are 10 values of Z (MC2 and MC4) rather than 5 and in these cases the bias correction is very effective. The confidence regions are tolerably small in the larger sample size cases.

5.5. Parametric model. When this nonparametric model is augmented with parametric restrictions the identified set is reduced to the subset of the identified set of nonparametric functions in which lie only functions that are members of the family of functions specified in the parametric model. To illustrate, consider the identifying power of the following probit parametric model,

$$Y = h(X,U) = \begin{cases} 0 \quad , & 0 \leq \; U \; \leq \Phi(-\alpha_0 - \alpha_1 X) \\ 1 \quad , & \Phi(-\alpha_0 - \alpha_1 X) < \; U \; \leq 1 \end{cases} \qquad U \perp\!\!\!\perp Z$$

when Y and X are determined by the structure used to produce Figure 1 for which the parameter values are given in the first row of Table 1. At these parameter values the structural threshold-crossing function is $\Phi(x)$ corresponding to a negative value $\alpha_1 = -1$ in the parameterisation used here.

The identified set of parameter values comprises the set of values of (α_0, α_1) which deliver functions $\Phi(-\alpha_0 - \alpha_1 X)$ that lie between the envelope upper and lower bounding functions graphed (red dashed) in Figure 1. Figure 8 shows this set - the set is not connected. The small set in the upper part of the graph corresponds to the monotone decreasing functions in Figure 1.

Figure 9 redraws Figure 1 and superimposes some of the probit functions that lie in the identified set. In the upper pane monotone increasing functions ($\alpha_1 < 0$) are drawn. Functions drawn in violet, black and green have intercept term α_0 equal to respectively -0.4, 0 and $+0.4$. In the lower pane, which shows decreasing functions ($\alpha_1 > 0$), only functions with $\alpha_0 = 0$ are shown.

In Figure 10 the identified set (shaded light blue) obtained when b_1 is increased to 0.4 is superimposed on the set obtained when $b_1 = 0.3$ (shaded dark blue). The nonparametric bounding functions for this case are shown in Figure 2 and it can be seen that there are no decreasing functions in the identified set with this larger value of b_1. As a result the light blue set shown in Figure 10 is connected.

The symmetry in these identified sets arises because of essential symmetry in the probability measure used in this example. Figure 11 shows asymmetric identified sets of parameters in the parametric probit model obtained under a different probability measure. Here the structure generating the probability measure has been modified so that it no longer satisfies the *full* set of triangular model restrictions. The triangular form of the structural equations *is* maintained and the unobservable variables *are* jointly Gaussian but the covariance of the unobservables conditional on the instrumental variable (now written as $s_{wv}(z)$) now depends on the instrumental variable's value, as follows.[19]

$$s_{wv}(z) = s_{vv}^{1/2} \times \frac{\exp(\pi_0 + \pi_1 z) - 1}{\exp(\pi_0 + \pi_1 z) + 1}$$

The unobservables (W, V) are now not *jointly* independent of Z unless $\pi_1 = 0$ but they are *marginally* independent of Z. The value of π_0 is set equal to $\ln(3)$ which gives $s_{wv}(0) = 0.5$ and $\pi_1 = 1.5$ so that the conditional covariance is an increasing function of z.

In Figure 11 the smaller light blue shaded set is obtained with parameters set as in Figure 2 (see Table 1) which has a relatively strong instrument with $b_1 = 0.4$. This set is connected. The dark blue shaded set is obtained with parameters set as in Figure 1 and has a relatively weak instrument with $b_1 = 0.3$. This set is not connected being the union of two connected (indeed convex) sets one of which contains only negative values of α_1 while the other contains only positive values. The symmetry evident in Figure 10 is not present in Figure 11.

6. MONOTONE INDEX RESTRICTION

Now consider cases in which X may be a vector and the exogenous variables, Z, may appear in the structural function, possibly subject to some restrictions.

Consider models in which there is a *monotone index restriction*, namely that for all values, x and z, of X and Z the threshold crossing function can be written as $p(\alpha' x + \delta' z)$ for some constant finite dimensional vectors α and δ, where p is a monotone function. The resulting monotone (linear) index binary outcome model is as follows.

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

There will typically be a restriction excluding some elements of Z from this index, that is requiring some elements of δ to be zero. There will be a normalisation; for example one might set equal to 1 an element of δ corresponding to an exogenous variable whose coefficient is restricted to be non-zero.

[19]This functional form respects the condition that the correlation between U and V given Z lie between -1 and 1.

Consider a threshold function $p(\alpha' x + \delta' z)$ which lies in the identified set for the probability measure Pr_0.

Analogous to (12) and (13) there is, for increasing p:

$$c_{0l}(u, z; h) = \text{Pr}_0[\{Y = 0\} \cap \left\{\alpha' X + \delta' Z < p^{-1}(u)\right\} |z] < u \quad (25)$$

$$c_{0u}(u, z; h) = 1 - \text{Pr}_0[\{Y = 1\} \cap \left\{\alpha' X + \delta' Z \geq p^{-1}(u)\right\} |z] \geq u \quad (26)$$

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

Continuing along the lines taken in Section 5.1 there is, on substituting $u = p(\sigma)$, for increasing p:

$$d_{0l}^{\uparrow}(\sigma, z; \alpha, \delta) \equiv \text{Pr}_0[\{Y = 0\} \cap \left\{\alpha' X < \sigma - \delta' z\right\} |z] < p(\sigma) \quad (27)$$

$$d_{0u}^{\uparrow}(\sigma, z; \alpha, \delta) \equiv 1 - \text{Pr}_0[\{Y = 1\} \cap \left\{\alpha' X \geq \sigma - \delta' z\right\} |z] \geq p(\sigma) \quad (28)$$

and for decreasing p:

$$d_{0l}^{\downarrow}(\sigma, z; \alpha, \delta) \equiv \text{Pr}_0[\{Y = 0\} \cap \left\{\alpha' X > \sigma - \delta' z\right\} |z] < p(\sigma) \quad (29)$$

$$d_{0u}^{\downarrow}(\sigma, z; \alpha, \delta) \equiv 1 - \text{Pr}_0[\{Y = 1\} \cap \left\{\alpha' X \leq \sigma - \delta' z\right\} |z] \geq p(\sigma) \quad (30)$$

which both hold for all $z \in Z$ and $\sigma \in \text{supp}(X)$. Since there is conditioning on $Z = z$ it is the random variables Y and $\alpha' X$ that are involved in the probability calculations.

If (and only if) $p(\alpha' x + \delta' z)$ lies in the identified set these inequalities hold at each $\sigma \in \text{supp}(X)$ *for all* $z \in Z$. So it is the largest and smallest values of the respectively lower and upper bounding probabilities that are relevant. Defining:

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

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

there are the following inequalities:

$$\text{increasing } p: \quad d_{0l}^{\uparrow}(\sigma; \alpha, \delta) < p(\sigma) \leq d_{0u}^{\uparrow}(\sigma; a, \delta) \quad (31)$$

$$\text{decreasing } p: \quad d_{0l}^{\downarrow}(\sigma; \alpha, \delta) < p(\sigma) \leq d_{0u}^{\downarrow}(\sigma; \alpha, \delta) \quad (32)$$

which hold for all $\sigma \in \text{supp}(X)$ and all (and only) structural functions $p(\alpha' x + \delta' z)$ in the identified set under the monotone index restriction.

The identified set I_0 associated with a structure S_0 that generates a probability measure $F_{YX|Z}^0$ (indicated by Pr_0) comprises all (p, α, δ) for which one of the inequalities (31) and (32) hold for all $\sigma \in \text{supp}(X)$.

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

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

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

$$A_0^{\uparrow}(\alpha,\delta) \equiv \{p : d_{0l}^{\uparrow}(\sigma;\alpha,\delta) < p(\sigma) \leq d_{0u}^{\uparrow}(\sigma;\alpha,\delta) \quad \forall \sigma \in \text{supp}(X)\}$$

$$A_0^{\downarrow}(\alpha,\delta) \equiv \{p : d_{0l}^{\downarrow}(\sigma;\alpha,\delta) < p(\sigma) \leq d_{0u}^{\downarrow}(\sigma;\alpha,\delta) \quad \forall \sigma \in \text{supp}(X)\}$$

If a pair of upper and lower bounding functions, $(d_{0l}^{\uparrow}, d_{0u}^{\uparrow})$ or $(d_{0l}^{\downarrow}, d_{0u}^{\downarrow})$, intersect then the corresponding set (respectively $A_0^{\uparrow}$ and $A_0^{\downarrow}$) is empty. Let ϕ denote the empty set. The component of the identified set which relates to the finite dimensional parameters is defined as follows.

$$I_0^{\alpha\delta} \equiv \{\alpha,\delta : I_0^p(\alpha,\delta) \neq \phi\}$$

To summarise: all (and only) values (α,δ) which generate bounding functions between which can pass monotone functions p (increasing, decreasing or both) are in the identified set and each such value (α^*,δ^*) is associated with all the monotone functions that can pass between the bounding functions that are generated by (α^*,δ^*).

If for a value (α^*,δ^*) the sets $A_0^{\uparrow}(\alpha^*,\delta^*)$ and $A_0^{\downarrow}(\alpha^*,\delta^*)$ are both empty, which will happen if and only if the upper and lower envelope bounding functions defining each set intersect, then (α^*,δ^*) is not in the identified set of parameter values. If interest is centred on the finite dimensional parameters (α,δ) then only the identified set $I_0^{\alpha\delta}$ is of interest and it can be determined as the set of values (α,δ) such that *at least one* of the following inequalities holds for all $\sigma \in \text{supp}(X)$.

$$\begin{aligned} d_{0u}^{\uparrow}(\sigma;\alpha,\delta) - d_{0l}^{\uparrow}(\sigma;\alpha,\delta) &\geq 0 \\ d_{0u}^{\downarrow}(\sigma;\alpha,\delta) - d_{0l}^{\downarrow}(\sigma;\alpha,\delta) &\geq 0 \end{aligned}$$

The monotonicity restriction delivers enormous computational benefits because it allows the identified set of index coefficient values to be characterised without reference to the unknown threshold crossing function, p.

6.1. Illustration: specification. The probability measures used in this illustration are, as earlier, generated by triangular structures with a single endogenous variable. There are two exogenous variables (instruments), $Z = (Z_1.Z_2)$ with Z_1 excluded from the structural equation for Y.

$$Y = 1[Y^* > 0] \qquad Y^* = a_0 + a_1X + d_1Z_1 + d_2Z_2 + W \qquad X = b_0 + b_1Z_1 + b_2Z_2 + V$$

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

There is the normalisation $Var(W) = 1$.

The model whose identifying power is considered is as follows.

$$Y = \begin{cases} 0 \ , & 0 < \ U \ \leq p(\alpha_1X + Z_2) \\ 1 \ , & p(\alpha_1X + Z_2) < \ U \ \leq 1 \end{cases} \qquad U \perp\!\!\!\perp Z \quad p \text{ monotone}$$

The coefficient on Z_2 is normalised equal to 1.

To calculate the identified set the joint distribution of Y^* and $\alpha_1 X$ given $Z = z$ is required. Here α_1 is a trial value for inclusion in the identified set $I_0^{\alpha\delta}$. The distribution is $N(\mu, \Sigma)$ with parameters as follows.

$$\mu \equiv \begin{bmatrix} a_1 b_0 + (a_1 b_1 + d_1)\, z_1 + (a_1 b_2 + d_2)\, z_2 \\ \alpha_1 (b_0 + b_1 z_1 + b_2 z_2) \end{bmatrix}$$

$$\Sigma \equiv \begin{bmatrix} 1 + 2a_1 s_{wv} + a_1^2 s_{vv} & \alpha_1 (s_{wv} + a_1 s_{vv}) \\ \alpha_1 (s_{wv} + a_1 s_{vv}) & \alpha_1^2 s_{vv} \end{bmatrix}$$

Given this distribution it is straightforward to compute the bounding functions (27), (28), (29) and (30) as bivariate normal orthant probabilities and the envelope bounding functions that appear in (31) and (32) are obtained by finding minimum and maximum values for $z \equiv (z_1, z_2) \in \mathcal{Z} \equiv \mathcal{Z}_1 \times \mathcal{Z}_2$. In this illustration $\mathcal{Z}_1 = [-2, 2]$ and two intervals $\mathcal{Z}_2$ are considered: $[-2, 2]$ and $[-3, 3]$.

The parameter values used in the illustrative calculations are as follows.

$$a_0 = 0 \quad a_1 = -1 \quad d_1 = 0 \quad d_1 = 1 \quad b_0 = 0 \quad b_1 \in [0.17, 1.5] \quad b_2 = 0 \quad s_{wv} = 0.5 \quad s_{vv} = 1$$

There is an exclusion restriction, $d_1 = 0$, and d_2 is set equal to -1 which is consistent with the normalisation employed in the model.

The coefficient b_2 is zero, so in this illustration X is uncorrelated with Z_2. The variable Z_2 effectively provides a scale against which the impact of endogenous X on the index is measured. As already noted, two ranges of values of Z_2 are considered, $[-2, 2]$ and $[-3, 3]$. If Z_2 were not present, for example because it exhibited no variation at all or because d_2 were actually zero, then the model would not have any identifying power for α_1. This suggests that identified sets will be smaller when Z_2 exhibits more variation.

In the structures employed in this illustration the structural function is $\Phi(-a_1 X - d_2 Z_2)$ which is $\Phi(X + Z_2)$ for the parameter values employed.

6.2. Illustration: results. The identified sets are shown in Table 2. For small values of b_1 (the actual value of the coefficient on Z_1 in the equation for endogenous X) the identified sets are not connected; there is an interval containing negative values of α_1 (the coefficient on endogenous X in the structural function) and an interval containing positive values. The value $\alpha_1 = 0$ and values close to zero *never* lie in the identified set. This is because, as explained earlier, in the structure that generates the probability measure in this illustration the distribution of the outcome Y *does depend on the instrumental variable*, Z.

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

b_1	$z_2 \in [-2, 2]$				$z_2 \in [-3, 3]$			
	L < 0	U < 0	L > 0	U > 0	L < 0	U < 0	L > 0	U > 0
0.170	-2.38	-0.38	0.11	8.69	-2.38	-0.21	0.09	8.14
0.175	-2.21	-0.41	0.11	8.41	-2.16	-0.26	0.10	7.90
0.190	-1.57	-0.49	0.12	7.67	-1.69	-0.37	0.11	7.26
0.250	·	·	0.17	5.72	·	·	0.15	5.51
0.500	·	·	0.33	3.15	·	·	0.31	2.94
0.750	·	·	0.45	2.82	·	·	0.44	2.21
1.000	·	·	0.54	2.83	·	·	0.54	1.83
1.500	·	·	0.60	2.87	·	·	0.66	1.75
2.000	·	·	0.60	2.80	·	·	0.70	1.77

Table 2: Identified sets for α_1 for a sequence of values of b_1 and two ranges of values of z_2. At small values of b_1 the set is the union of two disjoint sets one containing negative values, one containing positive values .

7. Concluding remarks

A single equation IV threshold crossing model for a binary response is set, not point, identifying for the threshold function even when it is parametrically restricted.

When the predictive power of the instrumental variables is not very great, having low predictive power for the endogenous variable, the identified sets can be large in extent and they may not be connected. In this situation the identifying power of the additional restrictions embodied in the triangular model that motivates control function estimation is very substantial.

If there is doubt about the validity of the triangular model's restrictions then it is prudent to consider the sets identified by the single equation IV model. Sharp identifying sets have been characterized in this paper and estimation has been shown to be feasible. Requiring the threshold function to satisfy a monotonicity restriction yields very substantial computational benefits.

Suppose a triangular binary response structure is supported by data. The identified set for the single equation IV model calculated using the distribution of Y and X given Z from which that data is generated characterises the *observationally equivalent* structures that the data supports under the IV model's restrictions. Because these structures are observationally equivalent *no* features of data can ever distinguish any of these structures from the triangular structure. The results of this paper show the extent of these observationally equivalent structures.

If the estimates delivered by the triangular model for a binary response are used it will be because one has *faith* in that model's restrictions since no evidence will be ever be found in data to support those restrictions in the context of the encompassing single equation IV model.

References

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

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

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

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

Chesher, Andrew D., (2003): "Identification in Nonseparable Models," *Econometrica*, 71, 1405-1441.

Chesher, Andrew D., (2005): "Nonparametric Identification under Discrete Variation," *Econometrica*, 73, 1525-1550.

Chesher, Andrew D., (2007a): "Instrumental Values," *Journal of Econometrics*, 139, 15 – 34.

Chesher, Andrew D., (2007b): "Endogeneity and Discrete Outcomes," CeMMAP Working Paper 05/07.

Chesher, Andrew D., (2008): "Instrumental Variable Models for Discrete Outcomes," CeMMAP Working Paper 30/08.

Chesher, Andrew D., and Konrad Smolinski, (2009): "Ordered Choice Models with Endogenous Explanatory Variables," in preparation, presented at the conference "Identification and Decisions", May 8-9, Northwestern University.

Greene, William (2007): *LIMDEP 9.0 Reference Guide.* Econometric Software, Inc., New York.

Hausman, Jerry A., (1978): "Specification Tests in Econometrics," *Econometrica* 46, 1251-1271.

Heckman, James J., (1979): "Sample Selection Bias as a Specification Error," *Econometrica*, 47, 153-161

Ihaka, Ross, and Robert Gentleman (1996): "R: A Language for Data Analysis and Graphics," *Journal of Computational and Graphical Statistics*, 5, 299-314.

Imbens, Guido W., and Whitney K. Newey (2009): "Identification and Estimation of Triangular Simultaneous Equations Models without Additivity," forthcoming in *Econometrica.*

Matzkin, Rosa L. (2003): "Nonparametric Estimation of Nonadditive Random Functions," *Econometrica*, 71, 1339 - 1376.

Rivers, Douglas and Quang Vuong (1988): "Limited Information Estimators and Exogeneity Tests for Simultaneous Probit Models," *Journal of Econometrics* 39, 347-366.

Smith, Richard J., and Richard W. Blundell (1986): "An Exogeneity Test for a Simultaneous Equation Tobit Model with an Application to Labor Supply," *Econometrica* 54, 679-685.

Statacorp (2007): "Stata Statistical Software: Release 10." College Station, TX: StataCorp LP.
Tamer, Elie (2003): "Incomplete Simultaneous Discrete Response Model with Multiple Equilibria," *Review of Economic Studies*, 70, 147-165.

Annex: Proof of Theorem 2

Theorem 2 is proved by constructing a conditional distribution function with the required properties.[20] The construction is done for a representative value $z \in \mathcal{Z}$ and a particular permutation of $\gamma = \{\gamma_k\}_{k=1}^K$. Without loss of generality it is assumed that indices are assigned so that $\gamma_1 \leq \gamma_2 \leq \cdots \leq \gamma_K$. Assume the system of inequalities associated with this permutation given in equation (11) holds.[21] Define $\gamma_0 \equiv 0$ and $\gamma_{K+1} \equiv 1$.

In order to simplify notation, dependence of various conditional probabilities on z is not made explicit in the notation. Thus $\delta_k(z)$ is written as δ_k and $\alpha_k(z)$ is written as α_k. Define

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

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

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

and note that

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

and that the inequalities (11) imply the following inequality.

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

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

$$\delta_j \tilde{\beta}_{kj} = \min\left\{\delta_j \alpha_j, \max\left\{0, \tilde{\gamma}_k - \sum_{s=1}^{j-1} \delta_s \tilde{\beta}_{ks}\right\}\right\} \tag{33}$$

$$\delta_j \hat{\beta}_{kj} = \min\left\{\delta_j \left(1 - \alpha_j\right), \max\left\{0, \hat{\gamma}_k - \sum_{s=1}^{j-1} \delta_s \hat{\beta}_{ks}\right\}\right\} \tag{34}$$

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

$$F_{U|XZ}(\gamma_k|x_j, z) \equiv \tilde{\beta}_{kj} + \hat{\beta}_{kj} \tag{35}$$

which implies[22] $F_{U|XZ}(0|x_j, z) = 0$, $F_{U|XZ}(1|x_j, z) = 1$. The distribution function is endowed with non-decreasing line segments between each suucessive distinct pair of elements in γ.[23]

[20] The construction used here was proposed by Martin Cripps.

[21] The systems of inequalities associated with other permutations of γ are obtained simply by exchange of indices.

[22] Since $\gamma_0 \equiv 0$, $\tilde{\gamma}_0 = \hat{\gamma}_0 = 0$, so $\tilde{\beta}_{0j} = 0$ and $\hat{\beta}_{0j} = 0$ for all j which yields $F_{U|XZ}(0|x_j, z) = 0$. Since $\gamma_{K+1} = 1$, $\tilde{\gamma}_{K+1} = \bar{\alpha}$ and $\hat{\gamma}_{K+1} = 1 - \bar{\alpha}$, so for all j, $\tilde{\beta}_{K+1j} = \alpha_j$ and $\hat{\beta}_{K+1j} = 1 - \alpha_j$ which yields $F_{U|XZ}(1|x_j, z) = 1$.

[23] Linear segments will deliver piecewise uniform conditional distributions of U given X and Z.

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

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

$$\tilde{j}(k) = \max\{j : \tilde{\gamma}_k - \sum_{s=1}^{j-1}\delta_s\alpha_s \geq 0\} \tag{36}$$

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

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

$$\hat{j}(k) = \max\{j : \hat{\gamma}_k - \sum_{s=1}^{j-1}\delta_s\,(1-\alpha_s) \geq 0\} \tag{37}$$

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

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

$$\delta = \left[\begin{array}{ccccc} 0.1 & 0.2 & 0.3 & 0.1 & 0.3 \end{array}\right]$$

$$\alpha = \left[\begin{array}{ccccc} 0.5 & 0.3 & 0.4 & 0.5 & 0.8 \end{array}\right]$$

$$\gamma = \left[\begin{array}{ccccc} 0.3 & 0.4 & 0.5 & 0.6 & 0.8 \end{array}\right]$$

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

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

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

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

The proof now proceeds by showing the distribution function (35) is: (1) proper, (2) satisfies the independence restriction, and, (3) has an observational equivalence property. The properness and independence conditions are satisfied by construction, as will be shown. Satisfaction of the observational equivalence condition relies on the elements of γ satisfying the system of inequalities (11).

1. Proper conditional distributions

The proposed conditional distribution functions are proper if, for all j:

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

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

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

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

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

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

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

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

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

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

from which it follows that for all k:

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

as required.

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

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

The inequalities (11) can be written as follows.

$$\sum_{j=1}^{k} \delta_j \alpha_j \leq \gamma_k \leq \bar{\alpha} + \sum_{j=1}^{k-1} \delta_j(1-\alpha_j) \tag{38}$$

There are two cases to consider.

First suppose $\gamma_k \leq \bar{\alpha}$. Then $\tilde{\gamma}_k = \gamma_k$ and $\hat{\gamma}_k = 0$. From (38) there is on substituting $\gamma_k = \tilde{\gamma}_k$:

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

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

Now suppose that $\gamma_k > \bar{\alpha}$. Then $\tilde{\gamma}_k = \bar{\alpha}$, $\hat{\gamma}_k = \gamma_k - \bar{\alpha}$. Since $\bar{\alpha} - \sum_{j=1}^{k-1} \delta_j \alpha_j \geq \delta_k \alpha_k$, $\delta_k \tilde{\beta}_{kk} = \delta_k \alpha_k$. From (38) there is on substituting $\gamma_k - \bar{\alpha} = \hat{\gamma}_k$:

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

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

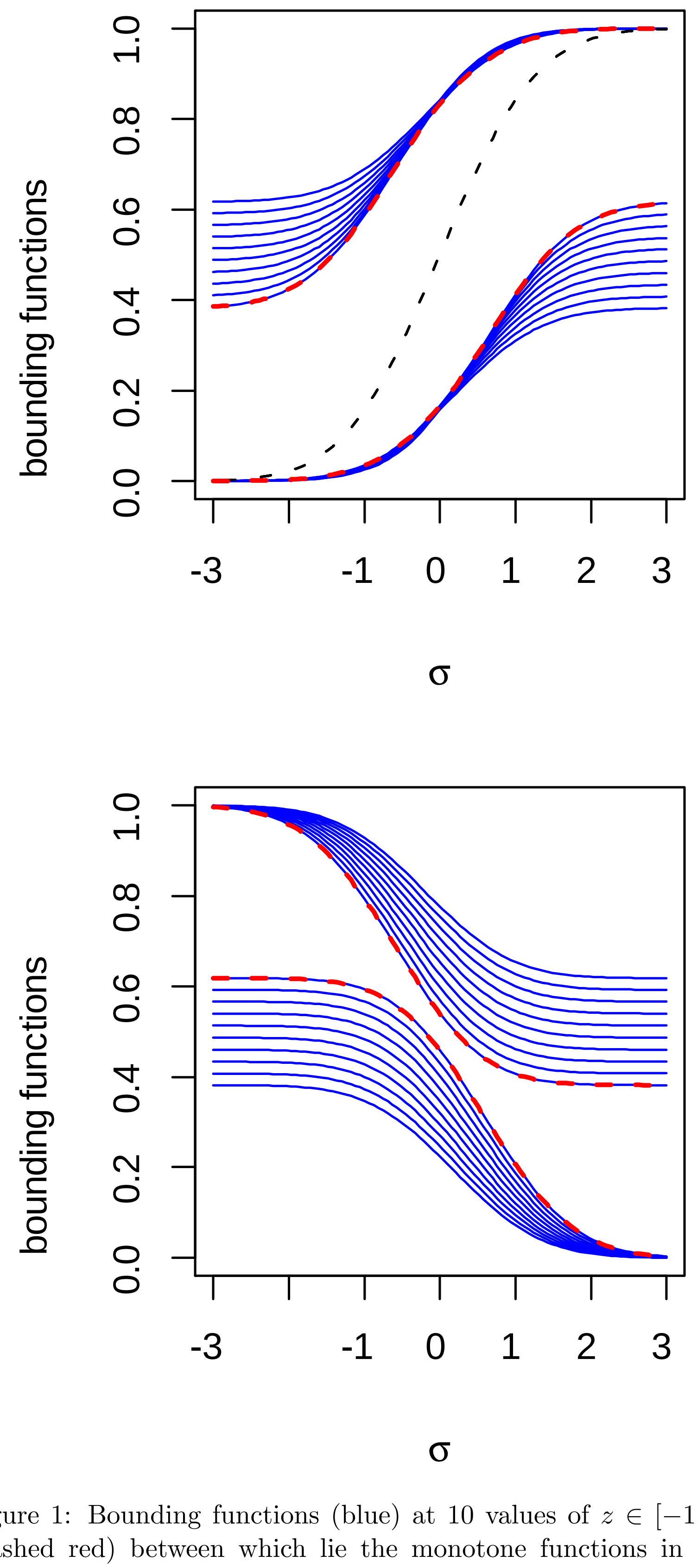


Figure 1: Bounding functions (blue) at 10 values of $z \in [-1, 1]$ with tight bounds (dashed red) between which lie the monotone functions in the identified set. A relatively weak instrument with $b_1 = 0.3$.

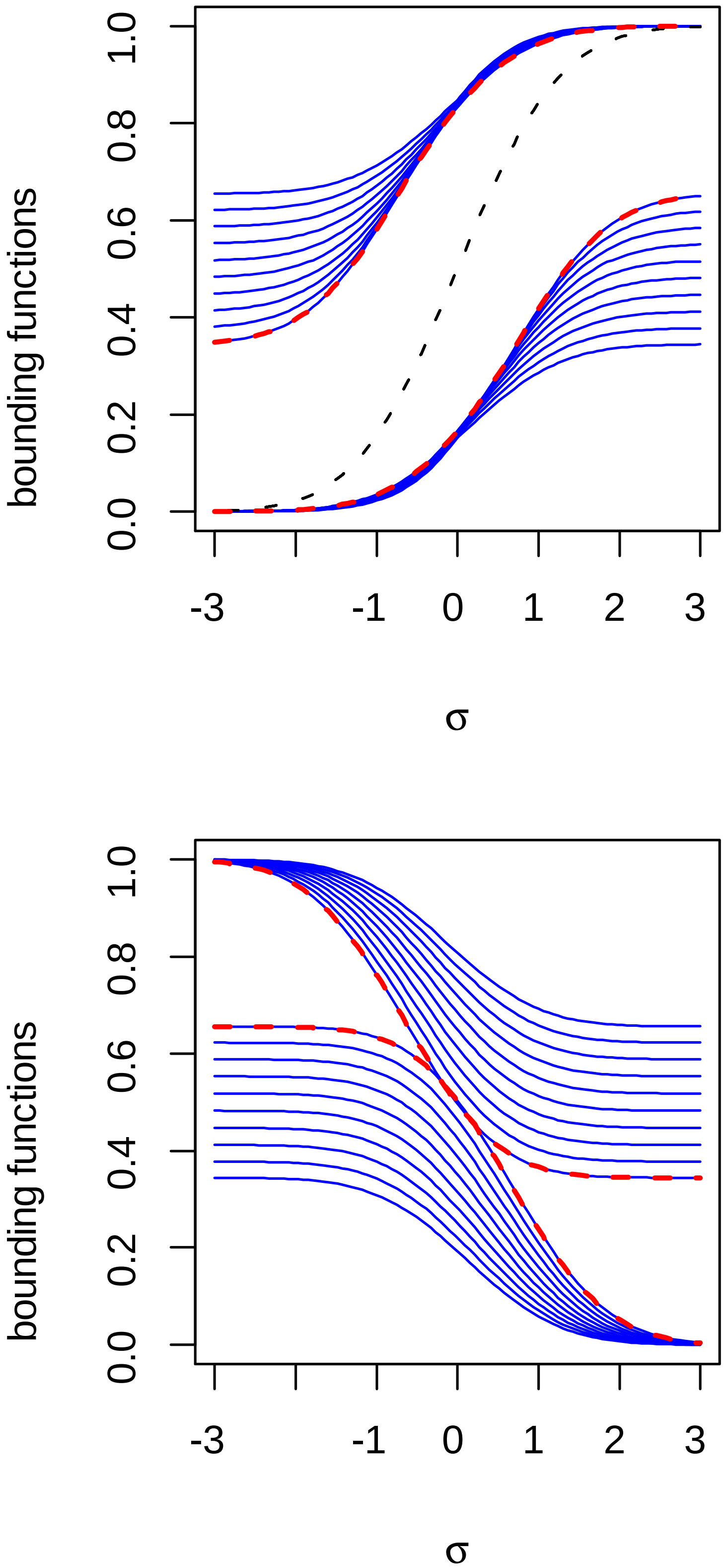


Figure 2: Bounding functions (blue) at 10 values of $z \in [-1, 1]$ with tight bounds (dashed red) between which lie the monotone functions in the identified set. A slightly stronger instrument than in Figure 1 with $b_1 = 0.4$. The set of decreasing functions (lower pane) is empty because the tight bonds intersect.

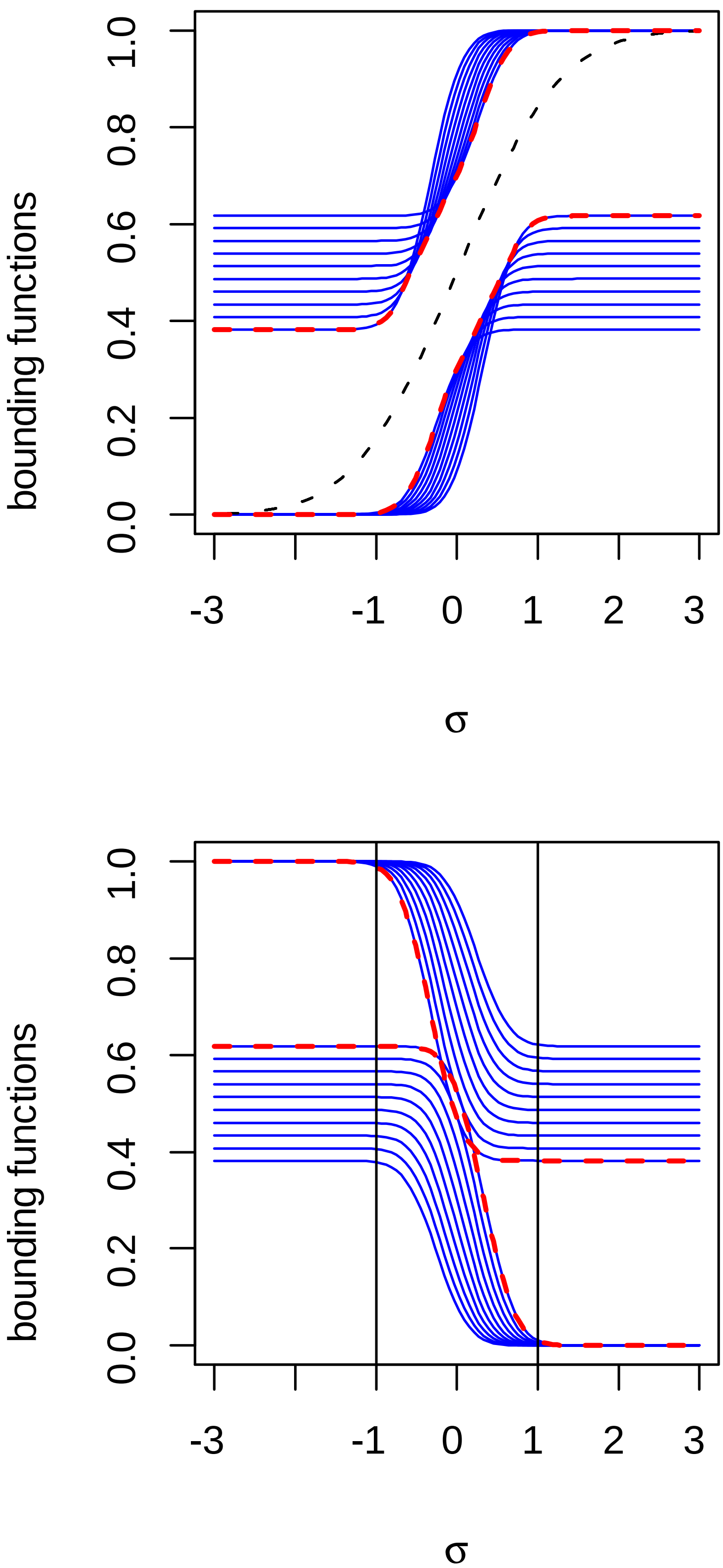


Figure 3: Bounding functions (blue) at 10 values of $z \in [-1, 1]$ with tight bounds (dashed red) between which lie the monotone functions in the identified set. A slightly stronger instrument than in Figure 1 with more predictive power for X: $b_1 = 0.3$ as in Figure 1 but $\sigma_{wv} = 0.05$, $\sigma_{vv} = 0.1$ (0.5 and 1 in Figures 1 and 2). The set of decreasing functions (lower pane) is empty.

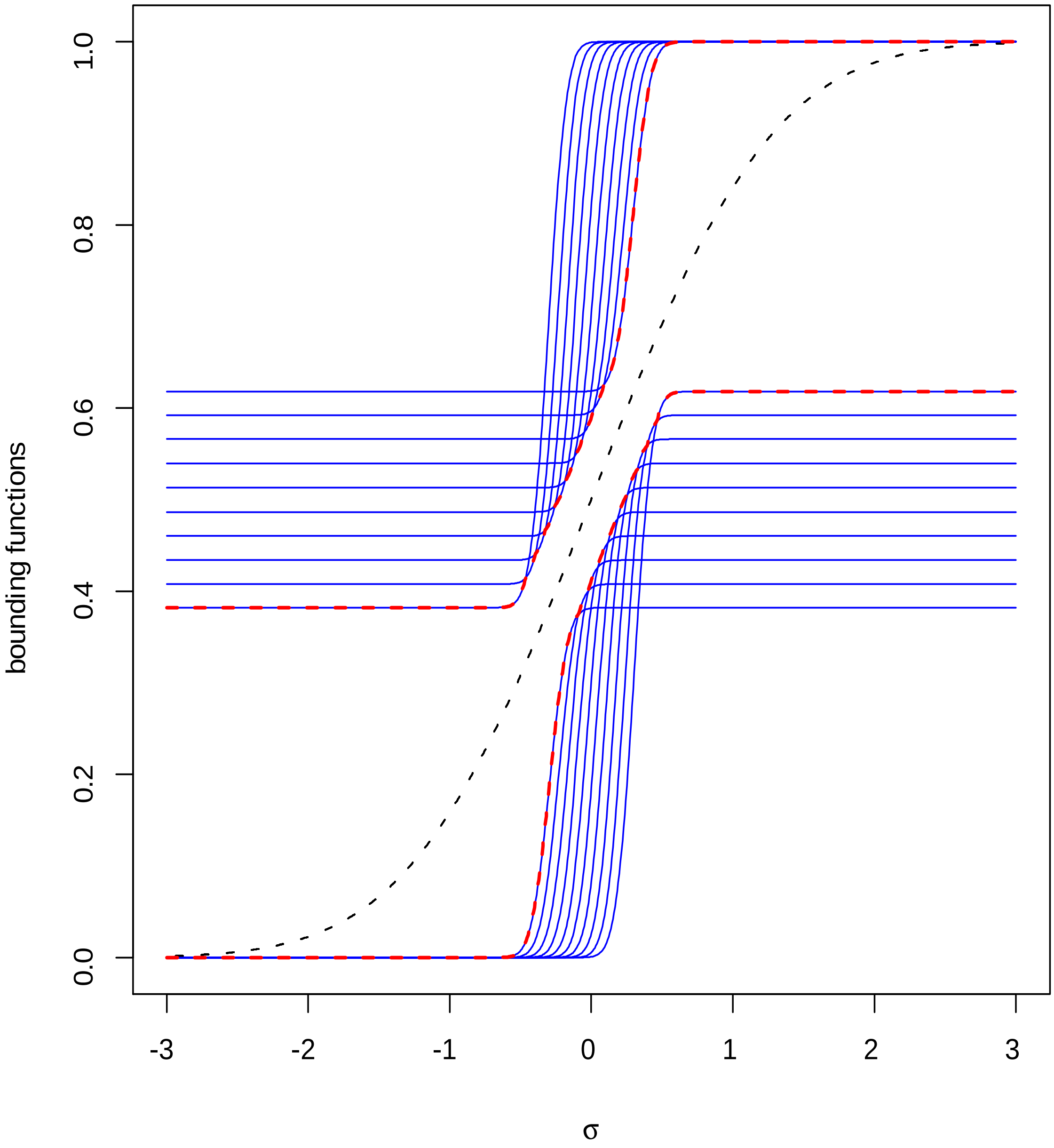


Figure 4: Bounding functions (blue) at 10 values of $z \in [-1, 1]$ with tight bounds (dashed red) between which lie the monotone functions in the identified set. An instrument with great predictive power: $b_1 = 0.3$ as in Figure 1 but $\sigma_{wv} = 0.005$, $\sigma_{vv} = 0.01$ (0.5 and 1 in Figures 1 and 2). There are no decreasing functions in the identified set.

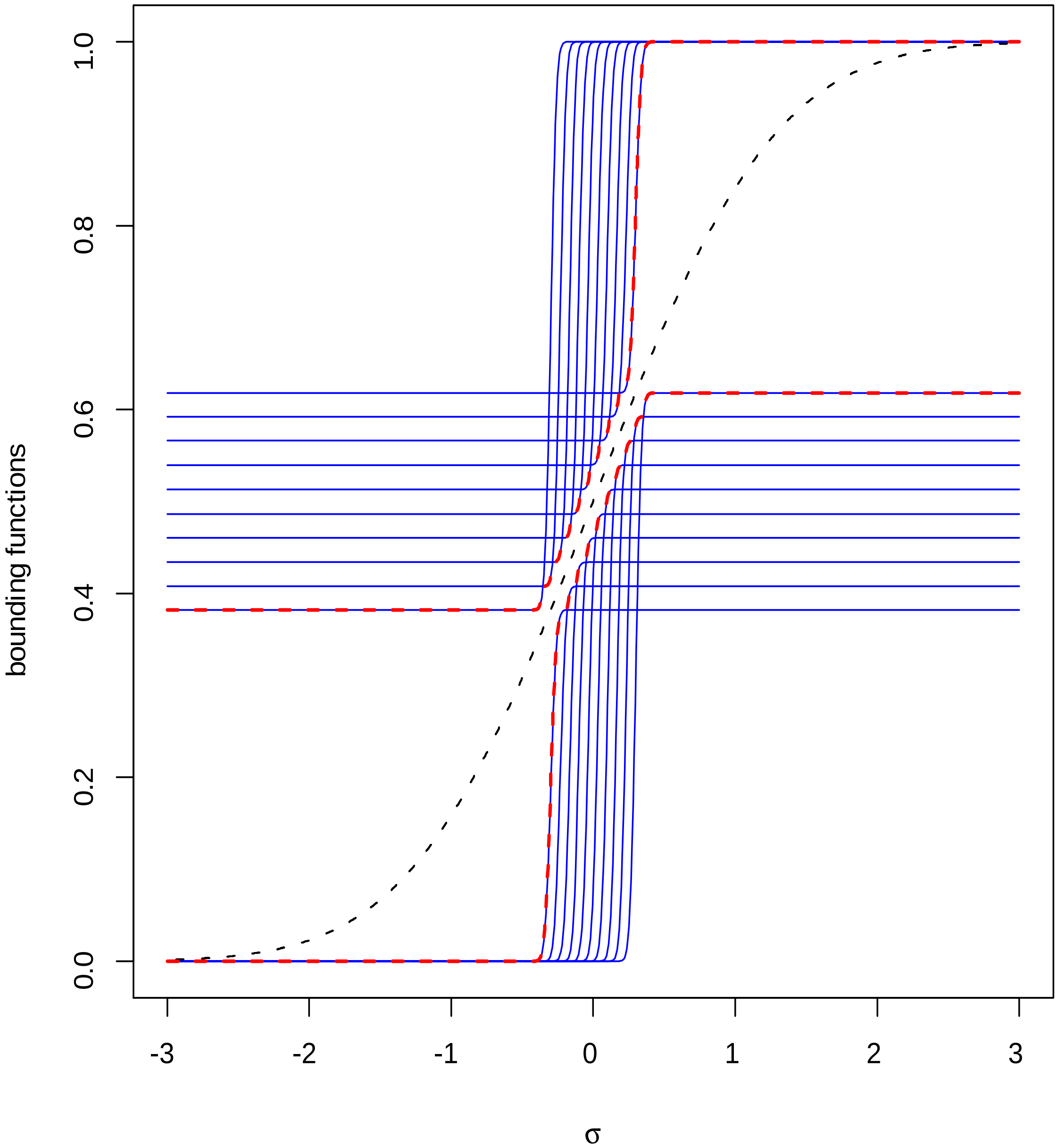


Figure 5: Bounding functions (blue) at 10 values of $z \in [-1, 1]$ with tight bounds (dashed red) between which lie the monotone functions in the identified set. An instrument with very great predictive power: $b_1 = 0.3$ as in Figure 1 but $\sigma_{wv} = 0.0005$, $\sigma_{vv} = 0.001$ (0.5 and 1 in Figures 1 and 2). There are no decreasing functions in the identified set.

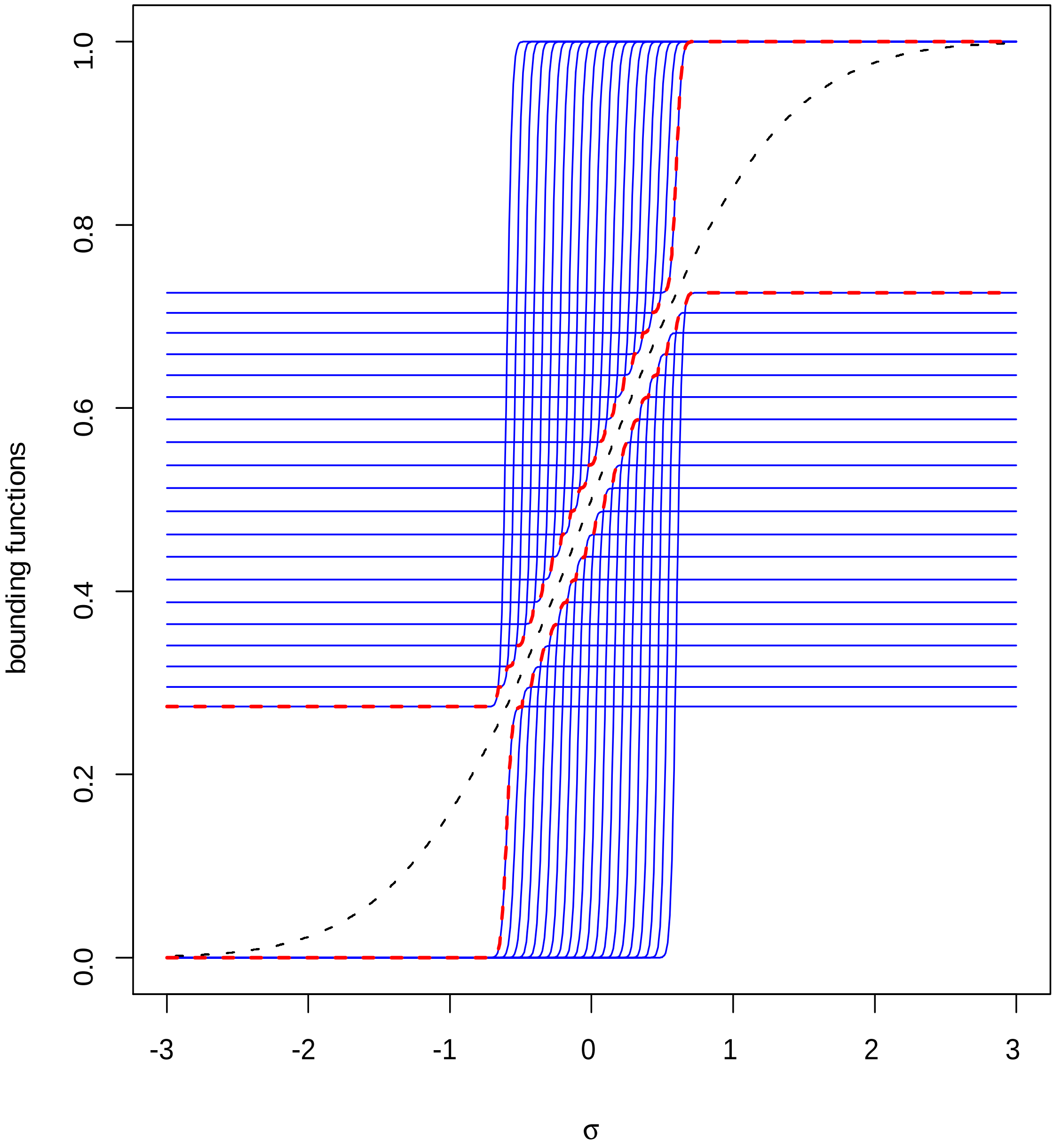


Figure 6: Bounding functions (blue) at 10 values of $z \in [-1, 1]$ with tight bounds (dashed red) between which lie the monotone functions in the identified set. An instrument with very great predictive power: $b_1 = 0.6$ twice the value in Figure 1 and $\sigma_{wv} = 0.0005$, $\sigma_{vv} = 0.001$ (0.5 and 1 in Figures 1 and 2). There are no decreasing functions in the identified set.

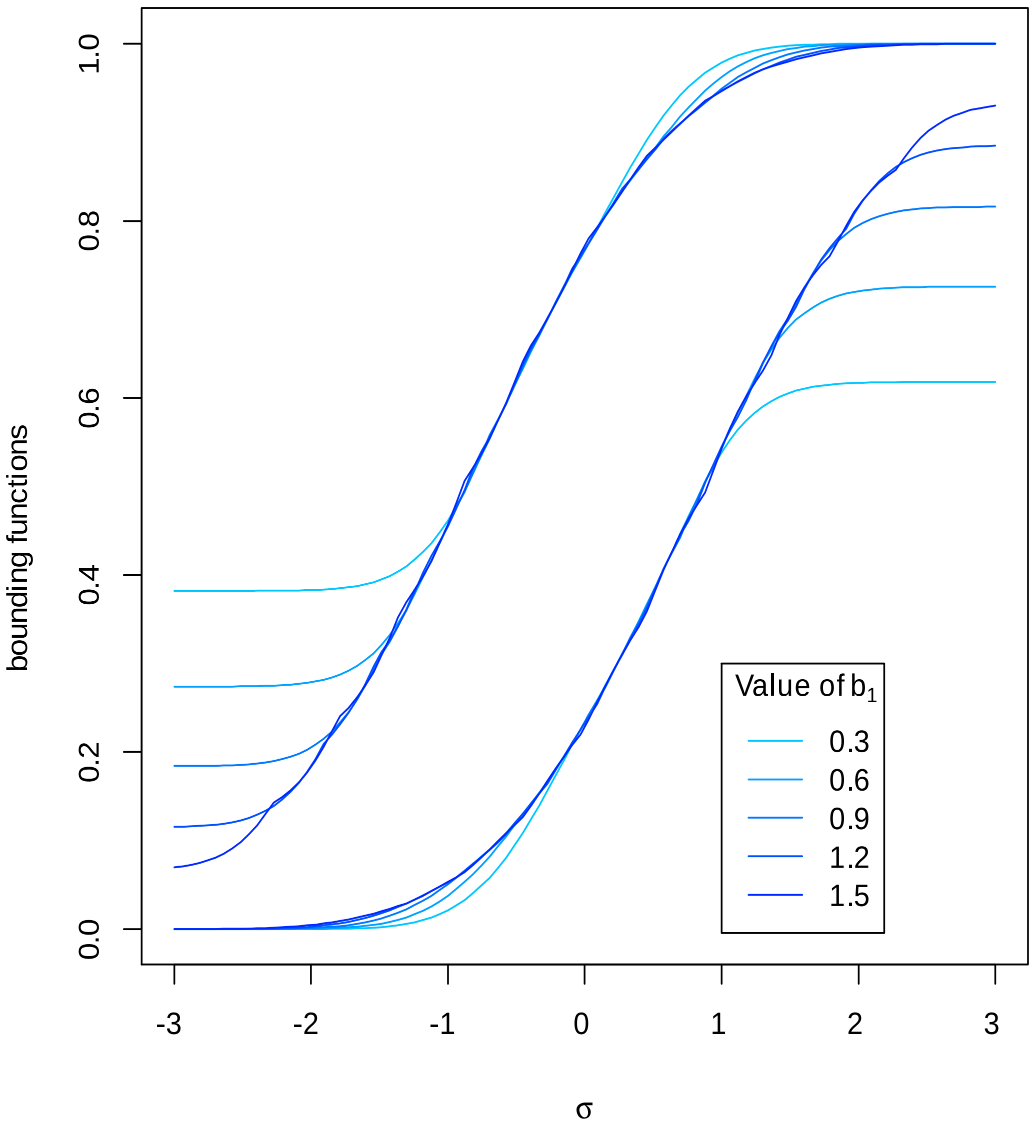


Figure 7: Tight bounds on the identified set drawn for a sequence of values of b_1.

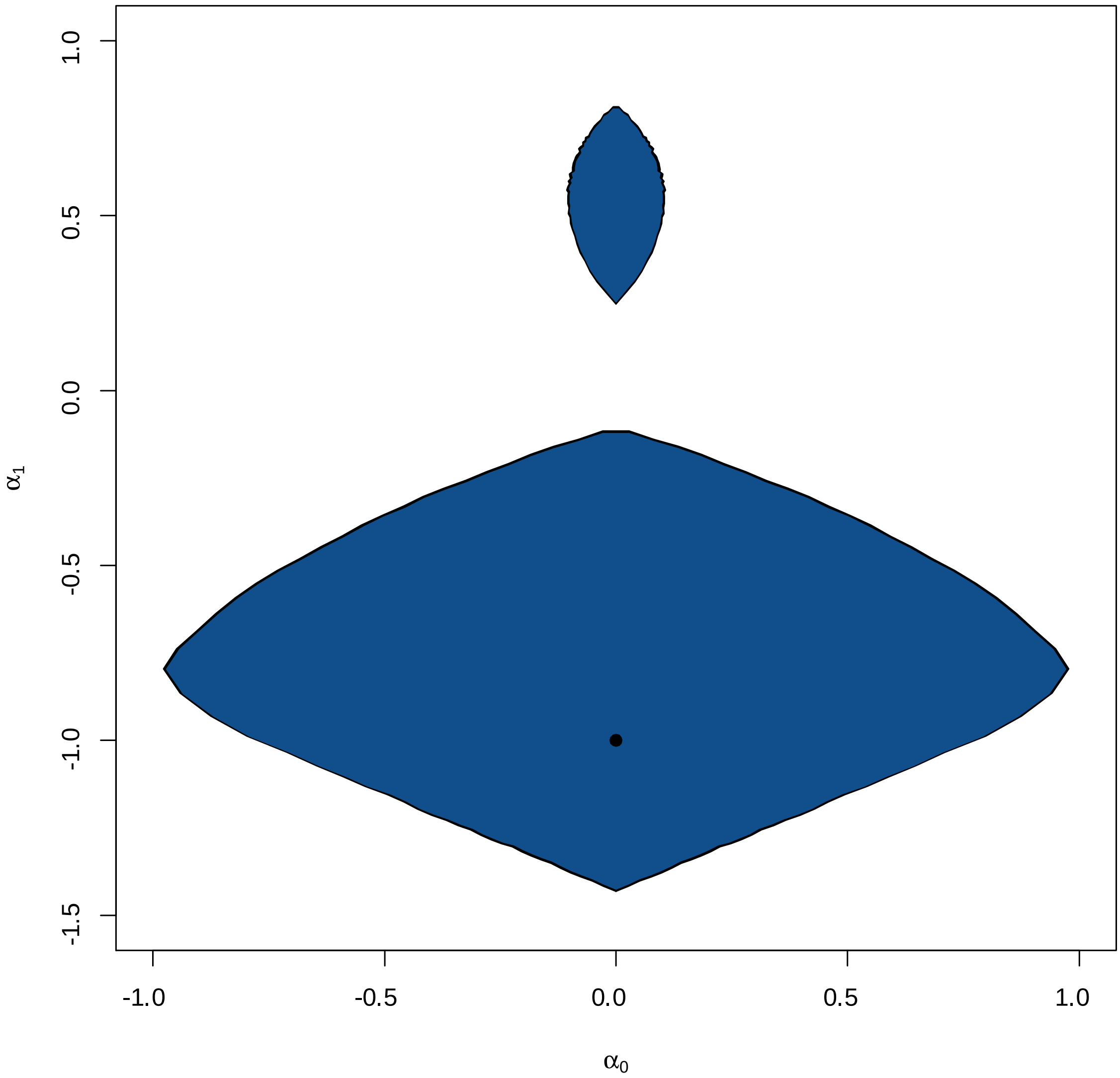


Figure 8: Identified sets for a parametric probit model for the structure set out in the first row of Table 1.

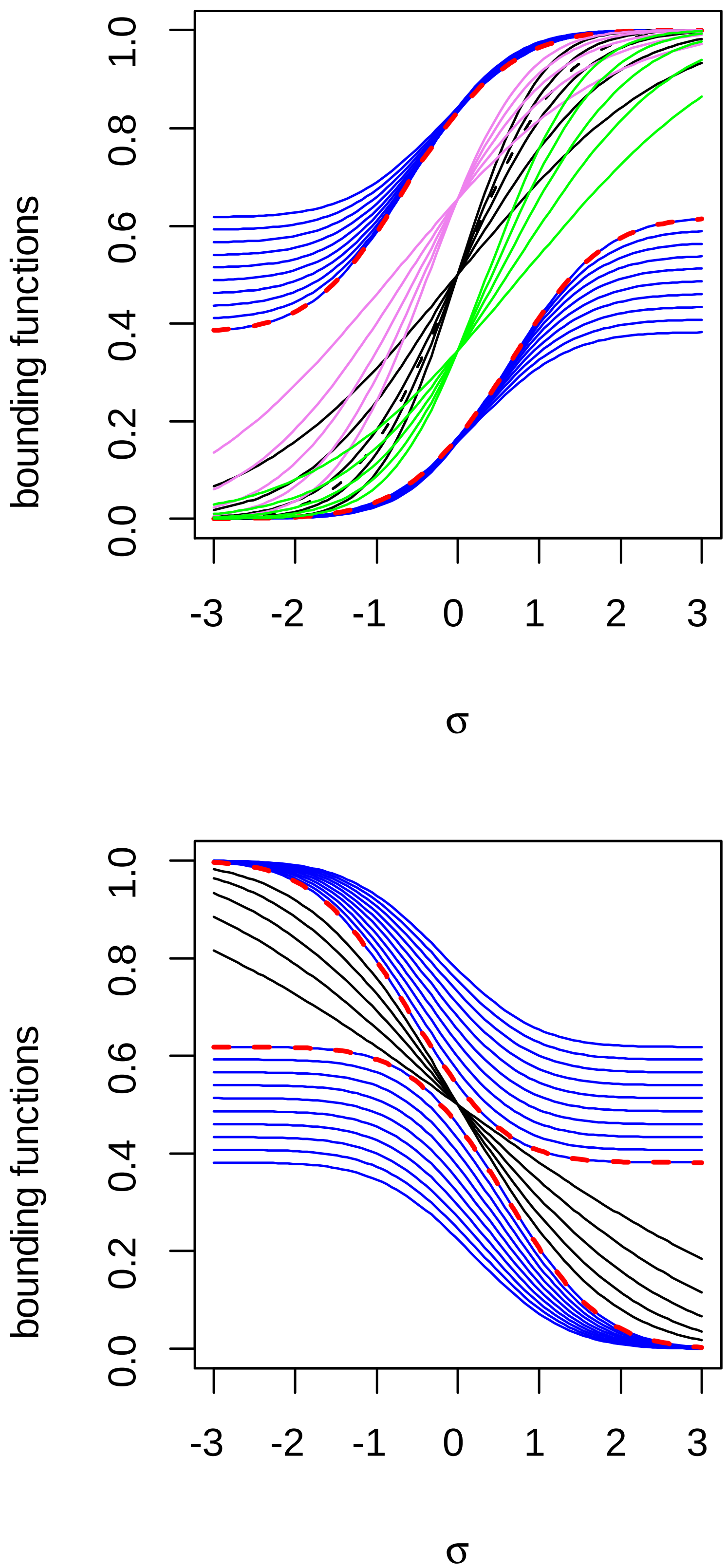


Figure 9: Some parametric probit functions falling in the identified set when $b_1 = 0.3$ with α_0 equal to -0.4 (violet), 0 (black), $+0.4$ (green)

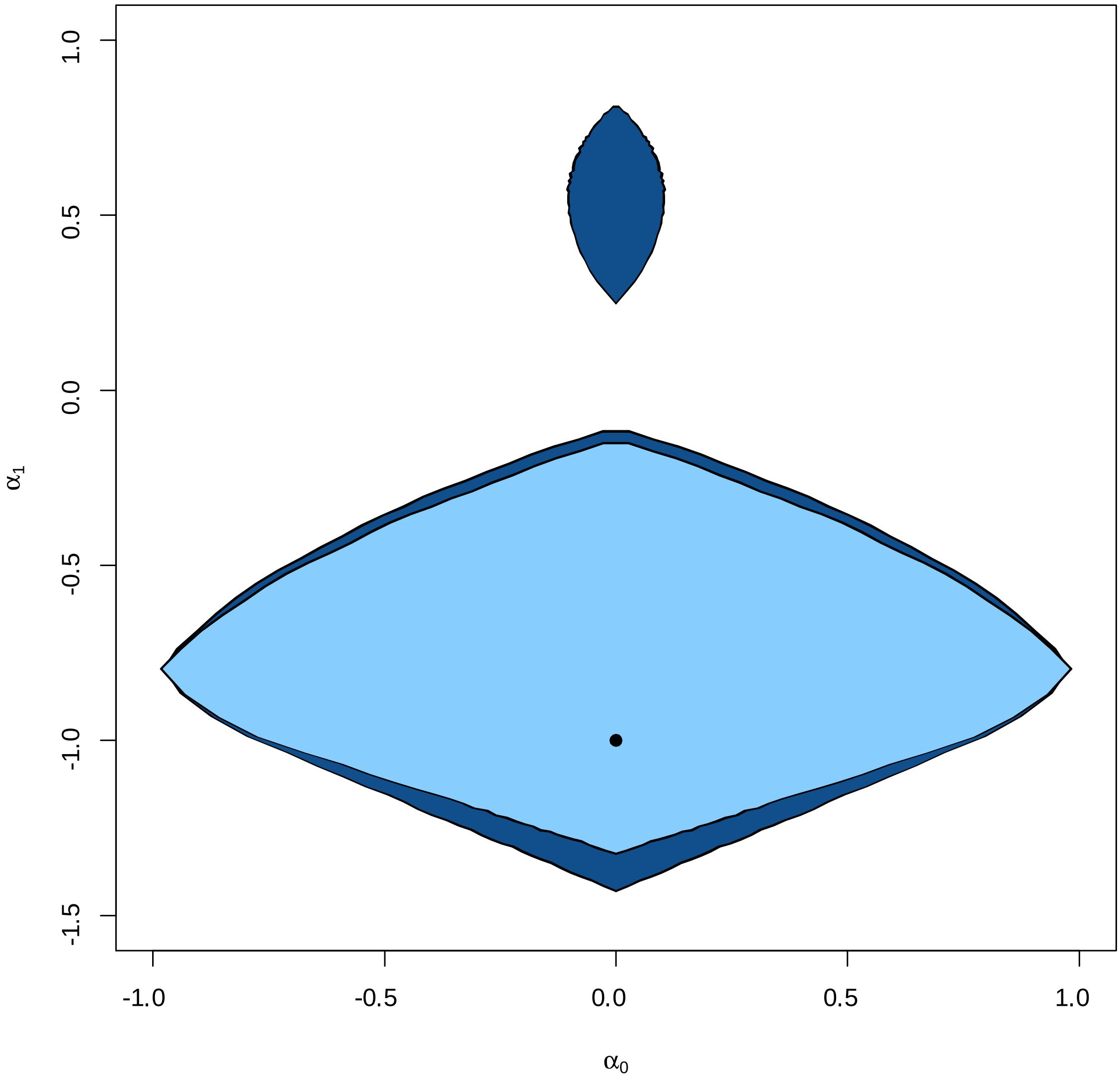


Figure 10: Identified sets for a parametric probit model for the structure set out in the first row of Table 1. with $b_1 = 0.3$ (dark blue) and $b_1 = 0.4$ (light blue)

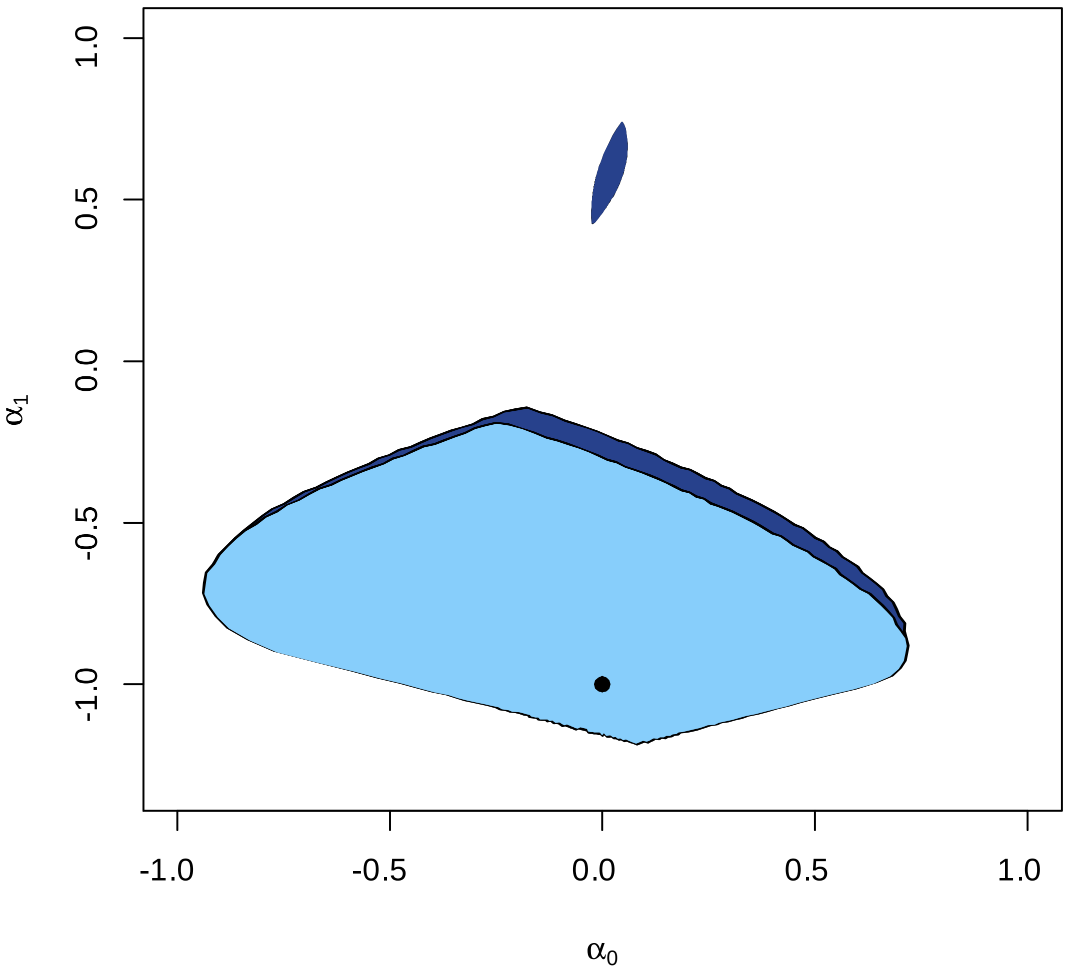


Figure 11: Identified sets for a parametric probit model when a probability measure is generated by a structure not satisfying triangular model conditions

MC1: $R = 1000 \quad n = 100 \quad K = 5 \quad \gamma_n = 0.124$												
	lower						upper					
$\uparrow$	σ	$d_l^{\uparrow}(\sigma)$	$b_N \times 100$	$b_C \times 100$	$RMSE_N$	$RMSE_C$	σ	$d_u^{\uparrow}(\sigma)$	$b_N \times 100$	$b_C \times 100$	$RMSE_N$	$RMSE_C$
1	-2.37	0.00	-0.08	-0.08	0.01	0.00	-2.37	0.40	-5.04	0.77	0.09	0.11
2	-0.91	0.04	5.76	-1.81	0.06	0.04	-0.91	0.61	-11.13	-1.97	0.12	0.11
3	0.03	0.17	7.75	-0.19	0.12	0.09	0.03	0.84	-8.91	-0.98	0.12	0.09
4	0.91	0.39	11.13	1.97	0.12	0.11	0.91	0.96	-5.76	1.81	0.06	0.04
5	2.37	0.60	5.04	-0.77	0.10	0.11	2.37	1.00	0.08	0.08	0.01	0.00
$\downarrow$	σ	$d_l^{\downarrow}(\sigma)$	$b_N \times 100$	$b_C \times 100$	$RMSE_N$	$RMSE_C$	σ	$d_u^{\downarrow}(\sigma)$	$b_N \times 100$	$b_C \times 100$	$RMSE_N$	$RMSE_C$
6	-2.37	0.62	3.24	-2.57	0.10	0.11	-2.37	0.98	-3.17	1.83	0.04	0.02
7	-0.91	0.59	6.27	0.46	0.10	0.11	-0.91	0.77	-2.08	3.19	0.09	0.10
8	0.03	0.45	4.71	-1.38	0.10	0.12	0.03	0.53	-3.44	2.65	0.10	0.12
9	0.91	0.23	2.08	-3.19	0.09	0.10	0.91	0.41	-1.27	2.46	0.09	0.11
10	2.37	0.02	3.17	-1.83	0.04	0.02	2.37	0.38	-3.24	2.57	0.09	0.11

Table 3: Monte Carlo results: MC1. R is a number of MC replications; n is a number of observations; K is a number of equally spaced points in the support of $z \in (-1, 1)$; γ_n is the tuning parameter defining the estimator of the set containing the minimizer. DGP parameters: $a_0 = 0, a_1 = -1, d_1 = 0, d_2 = 0, b_0 = 0, b_1 = 0.3, b_2 = 0$, $s_{ww} = 1, s_{vv} = 1$ and $s_{wv} = 0.5$. Columns 'lower' and 'upper' correspond to lower and upper bounding functions. Sets of increasing and decreasing functions are indicated by $\uparrow$ and $\downarrow$ respectively. For each value of σ we report true values $d(\sigma)$ of envelope probabilities; median bias of naive estimates b_N and of corrected estimates b_C; root mean square error of naive estimates $RMSE_N$, and of corrected estimates $RMSE_C$.

MC2: $R = 1000$ $n = 100$ $K = 10$ $\gamma_n = 0.25$												
	lower						upper					
$\uparrow$	σ	$d(\sigma)$	$b_N \times 100$	$b_C \times 100$	$RMSE_N$	$RMSE_C$	σ	$d(\sigma)$	$b_N \times 100$	$b_C \times 100$	$RMSE_N$	$RMSE_C$
1	-2.37	0.00	-0.08	-0.08	0.02	0.00	-2.37	0.40	-10.04	1.84	0.18	0.15
2	-0.91	0.04	5.76	-4.24	0.12	0.06	-0.91	0.61	-21.13	-3.64	0.23	0.15
3	0.03	0.17	22.75	2.38	0.21	0.12	0.03	0.84	-13.91	5.14	0.21	0.13
4	0.91	0.39	21.13	3.64	0.22	0.15	0.91	0.96	-5.76	4.24	0.12	0.05
5	2.37	0.60	10.04	-1.84	0.18	0.15	2.37	1.00	0.08	0.08	0.02	0.00
$\downarrow$	σ	$d(\sigma)$	$b_N \times 100$	$b_C \times 100$	$RMSE_N$	$RMSE_C$	σ	$d(\sigma)$	$b_N \times 100$	$b_C \times 100$	$RMSE_N$	$RMSE_C$
6	-2.37	0.62	18.24	2.67	0.17	0.15	-2.37	0.98	-8.17	1.83	0.07	0.02
7	-0.91	0.59	11.27	-3.19	0.18	0.15	-0.91	0.77	-7.08	10.76	0.15	0.14
8	0.03	0.45	14.71	-2.78	0.18	0.16	0.03	0.53	-13.44	4.05	0.17	0.15
9	0.91	0.23	7.08	-9.28	0.16	0.14	0.91	0.41	-11.27	3.19	0.17	0.15
10	2.37	0.02	8.17	-1.83	0.07	0.02	2.37	0.38	-18.24	-2.67	0.17	0.14

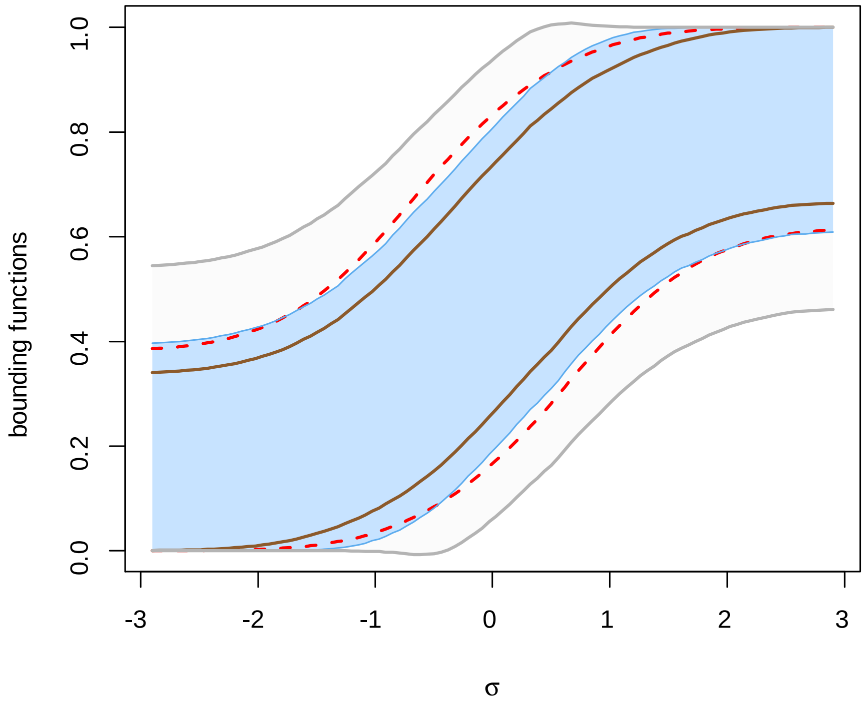
Table 4: Monte Carlo results: MC2. R is a number of MC replications; n is a number of observations; K is a number of equally spaced points in the support of $z \in (-1, 1)$; γ_n is the tuning parameter defining the estimator of the set containing the minimizer. DGP parameters: $a_0 = 0, a_1 = -1, d_1 = 0, d_2 = 0, b_0 = 0, b_1 = 0.3, b_2 = 0$, $s_{ww} = 1, s_{vv} = 1$ and $s_{wv} = 0.5$. Columns 'lower' and 'upper' correspond to lower and upper bounding functions. Sets of increasing and decreasing functions are indicated by $\uparrow$ and $\downarrow$ respectively. For each value of σ we report true values $d(\sigma)$ of envelope probabilities; median bias of naive estimates b_N and of corrected estimates b_C; root mean square error of naive estimates $RMSE_N$, and of corrected estimates $RMSE_C$.

MC3: $R = 1000$ $n = 400$ $K = 5$ $\gamma_n = 0.05$													
	lower						upper						
$\uparrow$	σ	$d(\sigma)$	$b_N \times 100$	$b_C \times 100$	$RMSE_N$	$RMSE_C$	σ	$d(\sigma)$	$b_N \times 100$	$b_C \times 100$	$RMSE_N$	$RMSE_C$
1	-2.37	0.00	-0.08	-0.08	0.01	0.00	-2.37	0.40	-1.29	-0.04	0.05	0.06
2	-0.91	0.04	2.01	-1.04	0.03	0.02	-0.91	0.61	-3.63	-1.13	0.05	0.06
3	0.03	0.17	4.00	1.42	0.06	0.05	0.03	0.84	-3.91	-1.42	0.06	0.05
4	0.91	0.39	3.63	1.13	0.05	0.05	0.91	0.96	-2.01	0.69	0.03	0.02
5	2.37	0.60	1.29	-0.19	0.05	0.06	2.37	1.00	0.08	0.08	0.01	0.00
$\downarrow$	σ	$d(\sigma)$	$b_N \times 100$	$b_C \times 100$	$RMSE_N$	$RMSE_C$	σ	$d(\sigma)$	$b_N \times 100$	$b_C \times 100$	$RMSE_N$	$RMSE_C$
6	-2.37	0.62	0.74	-0.09	0.05	0.06	-2.37	0.98	-0.67	1.30	0.02	0.02
7	-0.91	0.59	1.27	0.02	0.05	0.06	-0.91	0.77	-0.83	0.56	0.04	0.05
8	0.03	0.45	0.96	-0.29	0.05	0.06	0.03	0.53	-0.94	0.31	0.05	0.06
9	0.91	0.23	0.83	-0.49	0.04	0.05	0.91	0.41	-1.27	-0.02	0.05	0.06
10	2.37	0.02	0.67	-1.30	0.01	0.02	2.37	0.38	-0.74	-0.34	0.05	0.06

Table 5: Monte Carlo results: MC3. R is a number of MC replications; n is a number of observations; K is a number of equally spaced points in the support of $z \in (-1, 1)$; γ_n is the tuning parameter defining the estimator of the set containing the minimizer. DGP parameters: $a_0 = 0, a_1 = -1, d_1 = 0, d_2 = 0, b_0 = 0, b_1 = 0.3, b_2 = 0$, $s_{ww} = 1, s_{vv} = 1$ and $s_{wv} = 0.5$. Columns 'lower' and 'upper' correspond to lower and upper bounding functions. Sets of increasing and decreasing functions are indicated by $\uparrow$ and $\downarrow$ respectively. For each value of σ we report true values $d(\sigma)$ of envelope probabilities; median bias of naive estimates b_N and of corrected estimates b_C; root mean square error of naive estimates $RMSE_N$, and of corrected estimates $RMSE_C$.

MC4: $R = 1000$ $n = 400$ $K = 10$ $\gamma_n = 0.104$												
	lower						upper					
$\uparrow$	σ	$d(\sigma)$	$b_N \times 100$	$b_C \times 100$	$RMSE_N$	$RMSE_C$	σ	$d(\sigma)$	$b_N \times 100$	$b_C \times 100$	$RMSE_N$	$RMSE_C$
1	-2.37	0.00	-0.08	-0.08	0.01	0.00	-2.37	0.40	-5.04	1.14	0.08	0.08
2	-0.91	0.04	3.26	-2.98	0.06	0.03	-0.91	0.61	-8.63	-0.75	0.10	0.08
3	0.03	0.17	7.75	0.03	0.10	0.06	0.03	0.84	-8.91	0.09	0.10	0.06
4	0.91	0.39	8.63	-0.06	0.10	0.07	0.91	0.96	-3.26	2.98	0.05	0.03
5	2.37	0.60	5.04	-1.36	0.08	0.08	2.37	1.00	0.08	0.08	0.01	0.00
$\downarrow$	σ	$d(\sigma)$	$b_N \times 100$	$b_C \times 100$	$RMSE_N$	$RMSE_C$	σ	$d(\sigma)$	$b_N \times 100$	$b_C \times 100$	$RMSE_N$	$RMSE_C$
6	-2.37	0.62	5.74	-1.65	0.07	0.08	-2.37	0.98	-0.67	1.83	0.03	0.02
7	-0.91	0.59	3.77	-2.24	0.07	0.08	-0.91	0.77	-4.58	3.39	0.06	0.07
8	0.03	0.45	4.71	-1.77	0.07	0.08	0.03	0.53	-5.94	1.94	0.08	0.08
9	0.91	0.23	4.58	-3.39	0.07	0.07	0.91	0.41	-3.77	1.25	0.07	0.08
10	2.37	0.02	0.67	-1.83	0.03	0.02	2.37	0.38	-5.74	0.87	0.08	0.08

Table 6: Monte Carlo results: MC4. R is a number of MC replications; n is a number of observations; K is a number of equally spaced points in the support of $z \in (-1, 1)$; γ_n is the tuning parameter defining the estimator of the set containing the minimizer. DGP parameters: $a_0 = 0, a_1 = -1, d_1 = 0, d_2 = 0, b_0 = 0, b_1 = 0.3, b_2 = 0$, $s_{ww} = 1, s_{vv} = 1$ and $s_{wv} = 0.5$. Columns 'lower' and 'upper' correspond to lower and upper bounding functions. Sets of increasing and decreasing functions are indicated by $\uparrow$ and $\downarrow$ respectively. For each value of σ we report true values $d(\sigma)$ of envelope probabilities; median bias of naive estimates b_N and of corrected estimates b_C; root mean square error of naive estimates $RMSE_N$, and of corrected estimates $RMSE_C$.



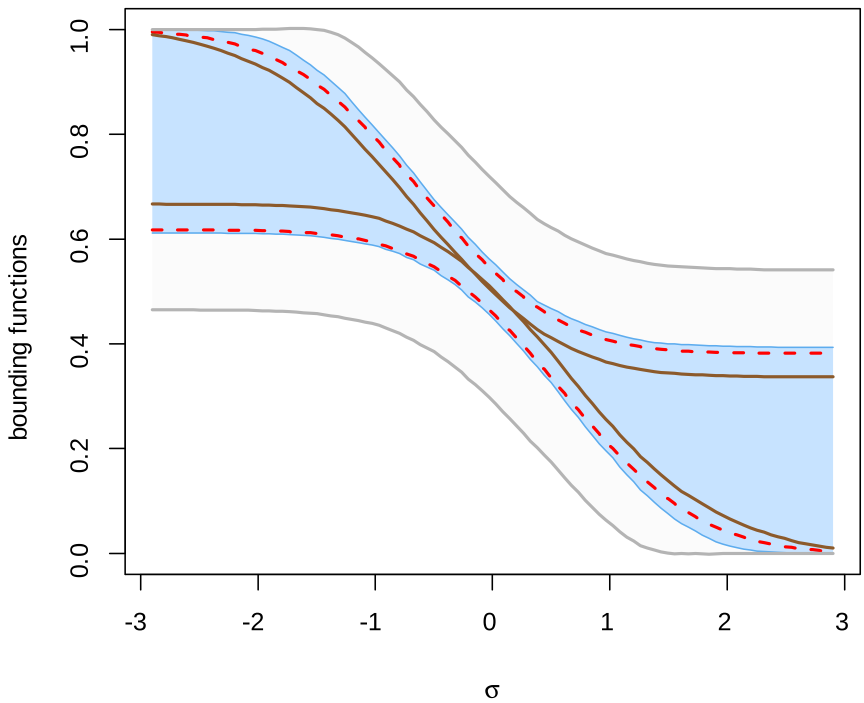
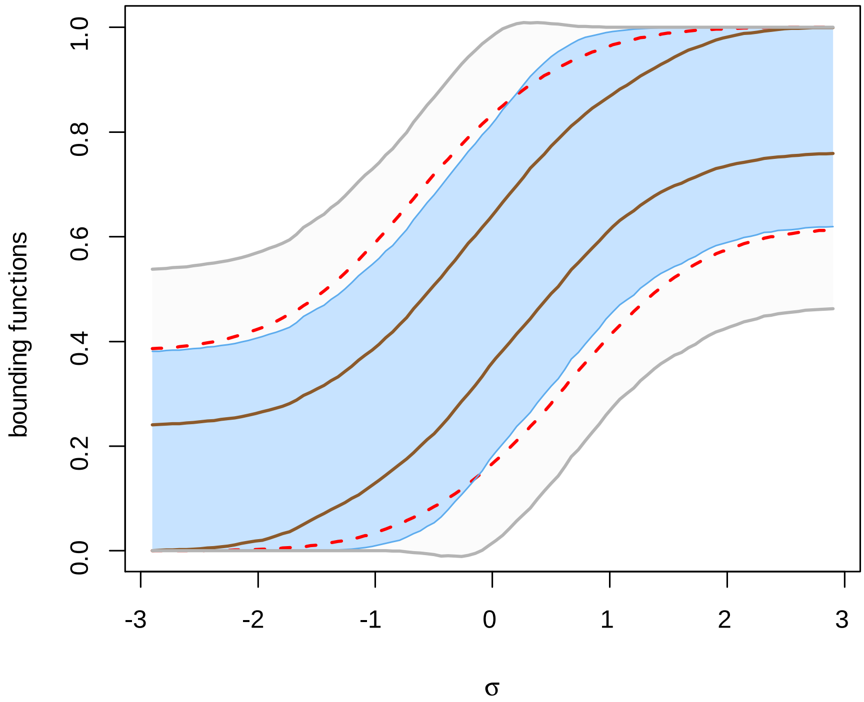


Figure 12: Monte Carlo experiment MC1. Bounding functions (dashed red), medians of naive estimates (brown), medians of corrected estimates (enclosing the shaded blue areas) and medians of boundaries of upper and lower 95% confidence regions.



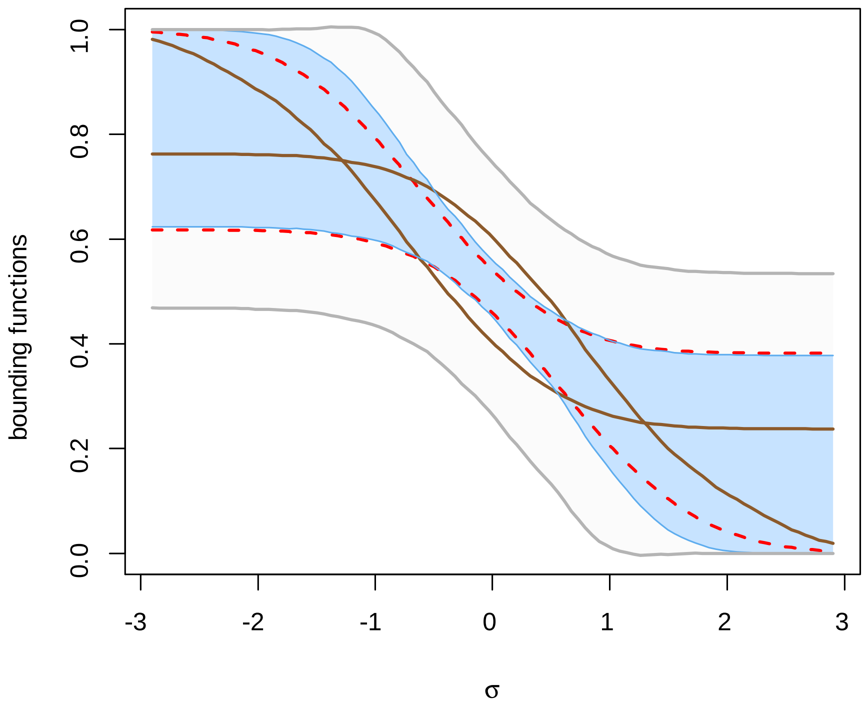


Figure 13: Monte Carlo experiment MC2. Bounding functions (dashed red), medians of naive estimates (brown), medians of corrected estimates (enclosing the shaded blue areas) and medians of boundaries of upper and lower 95% confidence regions.

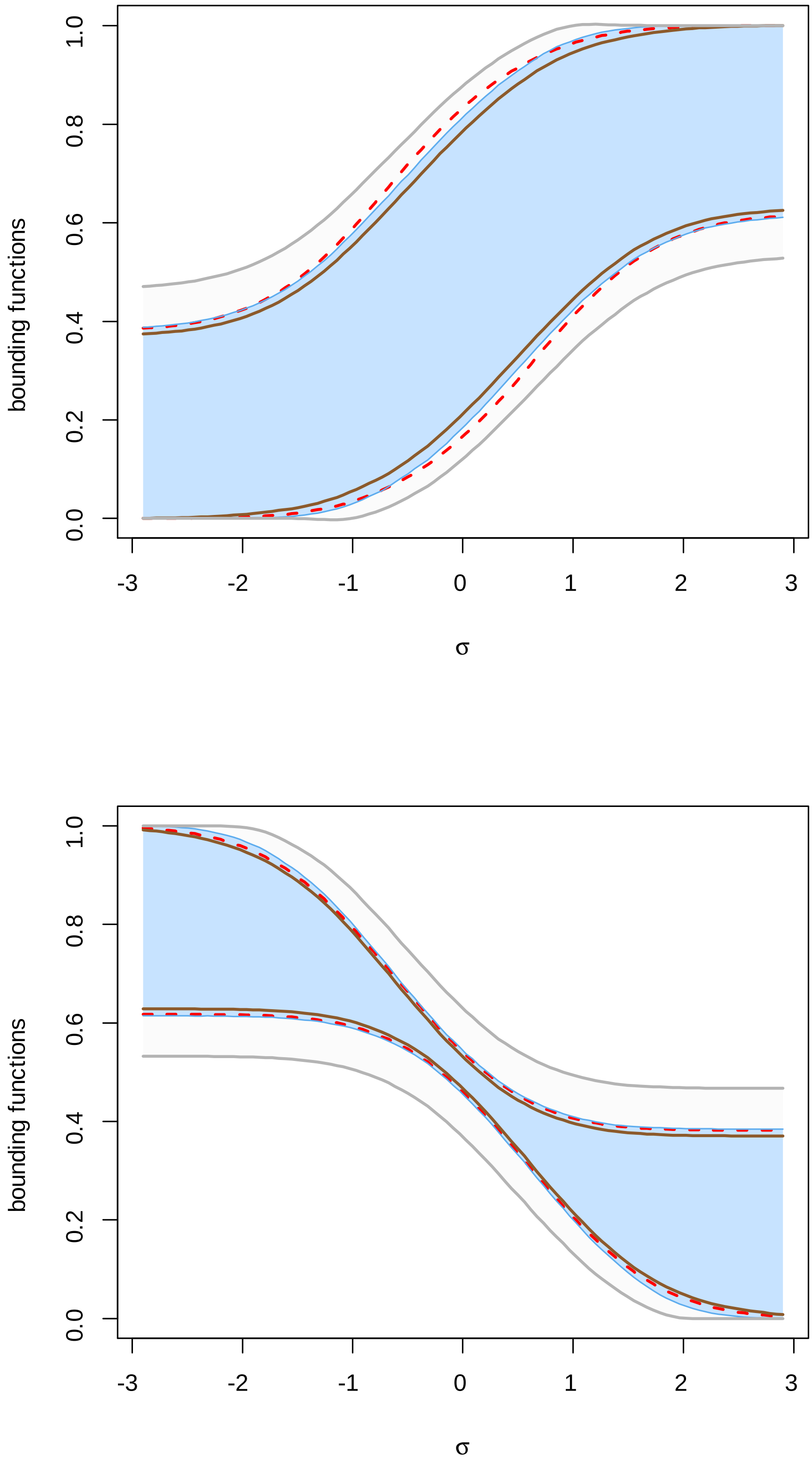


Figure 14: Monte Carlo experiment MC3. Bounding functions (dashed red), medians of naive estimates (brown), medians of corrected estimates (enclosing the shaded blue areas) and medians of boundaries of upper and lower 95% confidence regions.

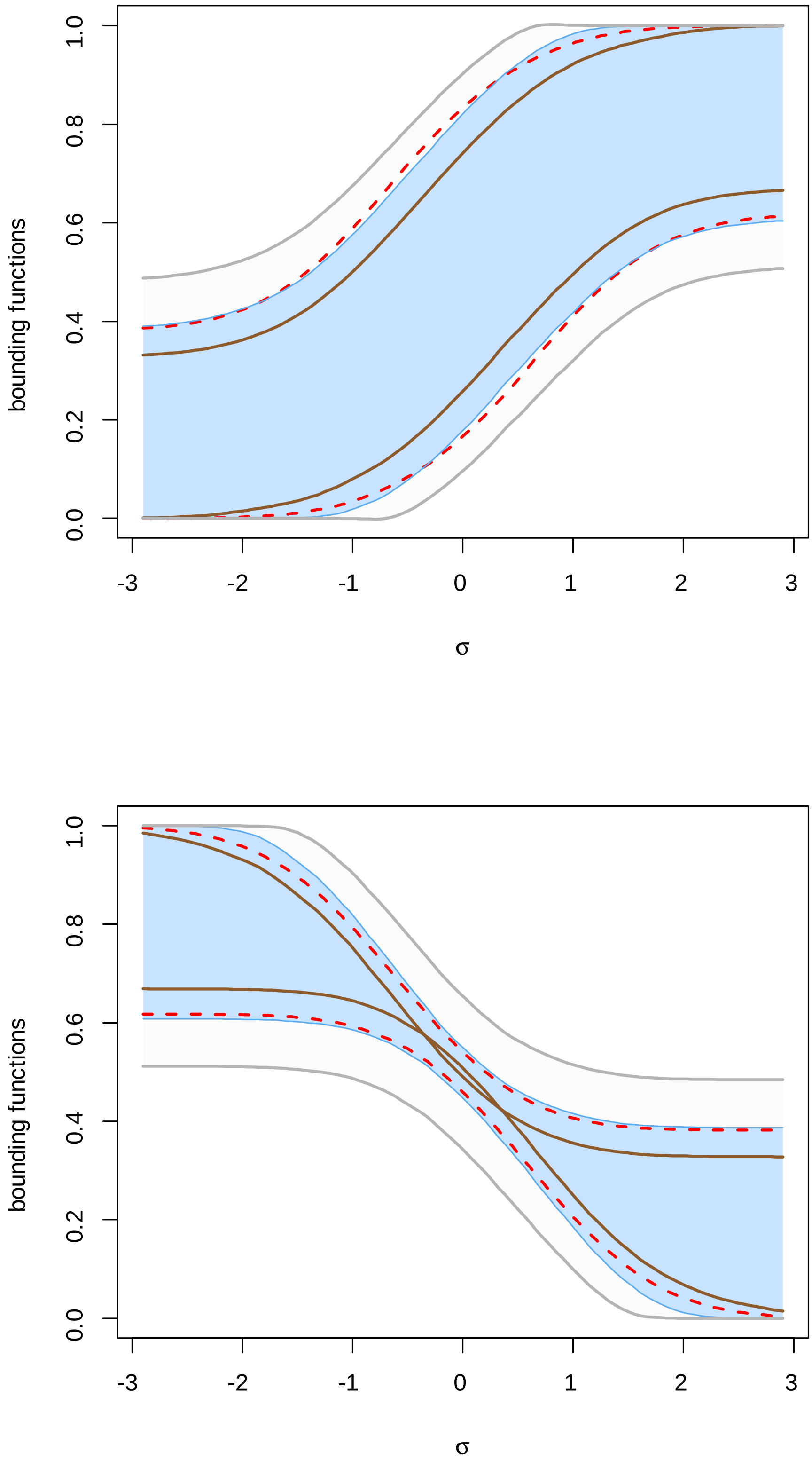


Figure 15: Monte Carlo experiment MC4. Bounding functions (dashed red), medians of naive estimates (brown), medians of corrected estimates (enclosing the shaded blue areas) and medians of boundaries of upper and lower 95% confidence regions.