cemmap
centre for microdata methods and practice

An instrumental variable random coefficients model for binary outcomes

Andrew Chesher
Adam M. Rosen

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP34/12

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL
An ESRC Research Centre

An Instrumental Variable Random Coefficients Model for Binary Outcomes*

Andrew Chesher† Adam M. Rosen‡

UCL and CeMMAP UCL and CeMMAP

October 23, 2012

Abstract

In this paper we study a random coefficient model for a binary outcome. We allow for the possibility that some or even all of the regressors are arbitrarily correlated with the random coefficients, thus permitting endogeneity. We assume the existence of observed instrumental variables Z that are jointly independent with the random coefficients, although we place no structure on the joint determination of the endogenous variable X and instruments Z, as would be required for a control function approach. The model fits within the spectrum of generalized instrumental variable models studied in Chesher and Rosen (2012a), and we thus apply identification results from that and related studies to the present context, demonstrating their use. Specifically, we characterize the identified set for the distribution of random coefficients in the binary response model with endogeneity via a collection of conditional moment inequalities, and we investigate the structure of these sets by way of numerical illustration.

Keywords: random coefficients, instrumental variables, endogeneity, incomplete models, set identification, partial identification, random sets.

*We thank participants at the June 2011 joint Northwestern Econometrics and CeMMAP conference in honor of Joel Horowitz for comments, and Konrad Smolinski for assistance with computations in the initial stages of this work. Financial support from the UK Economic and Social Research Council through a grant (RES-589-28-0001) to the ESRC Centre for Microdata Methods and Practice (CeMMAP) and through the funding of the "Programme Evaluation for Policy Analysis" node of the UK National Centre for Research Methods, and from the European Research Council (ERC) grant ERC-2009-StG-240910-ROMETA is gratefully acknowledged.

†Address: Andrew Chesher, Department of Economics, University College London, Gower Street, London WC1E 6BT, andrew.chesher@ucl.ac.uk.

‡Address: Adam Rosen, Department of Economics, University College London, Gower Street, London WC1E 6BT, adam.rosen@ucl.ac.uk.

JEL classification: C10, C14, C50, C51.

1 Introduction

In this paper we analyze a random coefficients model for a binary outcome,

$$Y = 1\left[\beta_0 + X\beta_1 + W\beta_2 > 0\right], \tag{1.1}$$

where $\beta \equiv \left(\beta_0, \beta_1', \beta_2'\right)'$ are random coefficients. While covariates W are restricted to be exogenous, covariates X are permitted to be endogenous in the sense that the joint distribution of X and random coefficients β is *not* restricted. We assume that in addition to the variables (Y, X, W), the researcher observes realizations of a random vector of instrumental variables Z such that (W, Z) and β are independently distributed. Thus our goal is to use knowledge of the joint distribution of (Y, X, W, Z) to set identify the marginal distribution of the random coefficients β, denoted F_β, with the joint distribution of random vectors X and β left unrestricted. As a special case we also allow for the possibility there are no exogenous regressors W.[1] As shorthand we use the notation $\tilde{Z} \equiv (W, Z)$ to denote the composite vector of all exogenous variables.

In order to characterize the identified set for F_β we carry out our identification analysis along the lines of Chesher, Rosen, and Smolinski (forthcoming) (CRS) and Chesher and Rosen (2012a). Like CRS we consider a single equation model for a discrete outcome, but here we restrict the outcome to be binary. The model (1.1) used in this paper however features random coefficients, which are not present in CRS. The model *is* a special case of the general class of models considered in Chesher and Rosen (2012a), where we provide identification analysis for a broad class of instrumental variable models. Like those models, the random coefficient model (1.1) allows for multiple sources of unobserved heterogeneity, whereas traditionally instrumental variable methods have been employed in models admitting a single source of unobserved heterogeneity. This paper thus investigates and illustrates by way of example the identifying power of instrumental variable restrictions with multivariate unobserved heterogeneity in the determination of a binary outcome.

[1] Similarly, the random intercept β_0 can be easily removed from the analysis by restricting $\beta_0 = 0$ throughout.

The characterizations we employ rely on results from random set theory, and these and related results have been used for identification analysis in various ways and in a variety of contexts by Beresteanu, Molchanov, and Molinari (2011), Galichon and Henry (2011), Beresteanu, Molchanov, and Molinari (2012), CRS, and Chesher and Rosen (2012a, 2012b). As in CRS and Chesher and Rosen (2012a, 2012b), our characterizations make use of properties of conditional distributions of certain random sets in the space of unobserved heterogeneity.

The model also builds on the instrumental variable models for binary outcomes considered in Chesher (2010) and Chesher (forthcoming), where a single source of unobserved heterogeneity was permitted. There it was found that even if parametric restrictions were brought to bear, the models were in general not point-identifying, and so with the addition of further sources of unobserved heterogeneity, point identification should not generally be expected. The paper thus serves to illustrate in part the effect of additional sources of heterogeneity from the perspective of identification. The case of a binary outcome variable is convenient for illustration, but models that permit more variation in outcome variables may achieve greater identifying power.

Binary response specifications that model β in (1.1) as a random vector include e.g. those of Quandt (1966) and McFadden (1976), and can be viewed as special cases of the discrete choice models of Hausman and Wise (1978) and Lerman and Manski (1981). These papers focus on specifications where all covariates and β are independently distributed, and where the distribution of β is parametrically specified, enabling estimation via maximum likelihood. Ichimura and Thompson (1998) and Gautier and Kitamura (forthcoming) focus on the binary outcome model (1.1), again with covariates and random coefficients independently distributed, but with F_β nonparametrically specified. Ichimura and Thompson (1998) provide sufficient conditions for point identification of F_β in this case, and prove that F_β can be consistently estimated via nonparametric maximum likelihood. Gautier and Kitamura (forthcoming) introduce a computationally simple estimator for the density of β, and derive its rate of convergence and pointwise asymptotic normality, while Gautier and LePennec (2011) propose an adaptive estimation method.

In contrast, we do not require that $X \perp\!\!\!\perp \beta$ and we employ instrumental variables Z. The use of an IV approach in a random coefficients binary response model with endogeneity is new. A

control function approach is employed by Hoderlein (2009) to provide identification results for marginal effects and local average structural derivatives when a triangular structure is assumed for the determination of X as a function of Z. He shows that the additional structure for the relation between the potentially endogenous variable X and the instrument Z then allows estimation via a control function approach. Our model does not require one to specify the form of the stochastic relation between X and Z, and is thus incomplete for the endogenous variables X.[2]

The random coefficients logit model of Berry, Levinsohn, and Pakes (1995) (BLP), now a bedrock of the empirical IO literature, allows for endogeneity of prices using insight from Berry (1994) to handle endogeneity. Yet the endogeneity problem in that and related models in IO is fundamentally different from the one in this paper. Their approach deals with correlation between alternative-specific unobservables with prices at the market level, both of which are assumed independent of random coefficients that allow for consumer-specific heterogeneity. Important identification results in such models are provided by Berry and Haile (2009, 2010), and a general treatment of the literature on such models and their relation to other models of demand is given by Nevo (2011). Here we focus on binary response models at a micro-level, rather than across separate markets, absent alternative-specific unobservables, and we allow random coefficients to be correlated with regressors.[3] Recent papers that give identification results for micro-level discrete choice models with exogenous covariates and high-dimensional unobserved heterogeneity include Briesch, Chintagunta, and Matzkin (2010), Bajari, Fox, Kim, and Ryan (2012), and Fox and Gandhi (2012). The latter also allows for endogeneity with alternative-specific special regressors and further structure on the determination of endogenous regressors as a function of the instruments.

Outline of the Paper: In Section 2 we formally present our model and key restrictions, and we introduce a simple example in which there is one endogenous regressor and no exogenous regressors. In Section 3 we characterize the identified set for the distribution of random coefficients in the general model set out in Section 2 and we provide two further examples. In Section 4 we provide

[2] The model is incomplete because there is no specification for the determination of X given exogenous variables Z and unobserved heterogeneity β. Thus, for any realization of (Z, β), each x on the support of X is a feasible realization of X. On the other hand, the triangular structure used in the control function approach implies a unique value of X for any realization of exogenous variables and unobservables.

[3] In a binary choice model the presence of unobserved, additively-separable, alternative-specific utility shifters may be subsumed into the threshold-crossing specification, and so is unnecessary.

numerical illustrations of identified sets for subsets of parameters in a parametric version of our model for four different data generation processes. Section 5 concludes. The proof of the main identification result, which adapts theorems from CRS, is provided in Appendix A. Appendix B provides computational details absent from the main text, and Appendix C verifies that there would be point identification in the example considered in the numerical illustrations of Section 4 if exogeneity restrictions were imposed.

Notation: We use capital Roman letters A to denote random variables and lower case letters a to denote particular realizations. For probability measure $\mathbb{P}$, $\mathbb{P}(\cdot|a)$ is used to denote the conditional probability measure given $A = a$. Calligraphic font $\mathcal{A}$ is used to denote the support of A for any well-defined random variable A in our model. $\mathcal{B}$ denotes the support of the random coefficient vector β, and $\mathcal{S}$ denotes a random closed set on $\mathcal{B}$. For any pair of random vectors A_1, A_2, $A_1 \perp\!\!\!\perp A_2$ denotes stochastic independence, $\text{Supp}(A_1, A_2, ..., A_n)$ denotes the joint support of the collection of random vectors $A_1, A_2, ..., A_n$, and $\text{Supp}(A_1, A_2, ..., A_n|b_1, ..., b_m)$ denotes the conditional support of $(A_1, A_2, ..., A_n)$ given realizations random vectors $(B_1, ..., B_m) = (b_1, ..., b_m)$. $\emptyset$ denotes the empty set. We use F_β to denote the probability distribution of β, mapping from Borel sets on $\mathcal{B}$ to the unit interval. $\mathcal{F}$ is used to denote the admissible "parameter" space for F_β, F is used to denote a generic element of $\mathcal{F}$, and $\mathcal{F}^*$ denotes the identified set for F_β. We use $\text{cl}(\mathcal{A})$ to denote the closure of a set $\mathcal{A}$. Finally, $\tilde{\mathcal{Z}} \equiv (W, Z)$ is used to denote the vector of all exogenous variables, and $\tilde{z} = (w, z)$ for particular realizations.

2 The Model

We now formally set out the restrictions of our model.

Restriction A1: $Y \in \{0, 1\}$, $X \in \mathcal{X} \subseteq \mathbb{R}^{k_x}$, and $W \in \mathcal{W} \subseteq \mathbb{R}^{k_w}$ obey (1.1) for some unobserved $\beta \in \mathcal{B} \subseteq \mathbb{R}^k$ with $k = k_x + k_w + 1$, and $Z \in \mathcal{Z} \subseteq \mathbb{R}^{k_z}$. (β, W, X, Y, Z) belong to a probability space $(\Omega, \Im, \mathbb{P})$ endowed with the Borel sets on Ω and the joint distribution of (X, W, Y, Z), denoted F^0_{XWYZ}, is identified. For all $(x, w, z) \in \text{Supp}(X, W, Z)$, $0 < \mathbb{P}[Y = 1|x, w, z] < 1$.

Restriction A2: For any (w, x, z) on the support of (W, X, Z), the conditional distribution of

random vector β given $W = w$, $X = x$, and $Z = z$ is absolutely continuous with respect to Lebesgue measure on $\mathcal{B}$. β is marginally distributed according to the probability measure F_β mapping from the Borel sets on $\mathcal{B}$ to the unit interval, with associated density f_β. F_β is known to belong to some class of probability measures $\mathcal{F}$.[4]

Restriction A3: (W, Z) and β are independently distributed.

Restriction A1 invokes the random coefficient model for the binary outcome Y and defines the support of random vectors X, W and Z. The restriction further requires that for all (x, w, z) both $Y = 1$ and $Y = 0$ have positive probability $\mathbb{P}(\cdot|x, w, z)$. This simplifies the exposition of some of the developments that follow, but is not essential. We do not otherwise restrict the joint support of (W, X, Y, Z). We require that the joint distribution of (W, X, Y, Z) is identified, as would be the case under random sampling, for instance. Restriction A3 is our instrumental variable restriction, requiring independence of (W, Z) and β. Restriction A2 restricts F_β to some known class of distribution functions. In principle, this class could be parametrically, semiparametrically, or nonparametrically specified. Of course greater identifying power will be afforded when $\mathcal{F}$ is parametrically specified, as is the case in our illustrations in Section 4, where β is restricted to be normally distributed, a common restriction in random coefficient models.

As is always the case in models of binary response, it will be prudent to impose a scale normalization since $x\beta > 0$ holds if and only if $c \cdot x\beta > 0$ for all scalars $c > 0$.[5] This may be imposed by imposing for example that $\mathcal{B} = \left\{b \in \mathbb{R}^k : \|b\| = 1\right\}$ if $\mathcal{F}$ is nonparametrically-specified, or by imposing that the first component of β has unit variance, e.g. when $\mathcal{F}$ is parametrically-specified as in the following example, also employed in the numerical illustrations of Section 4.

Example 1 (One endogenous variable, no exogenous variables): Suppose $X \in \mathbb{R}$ and that there are no exogenous covariates W. Then we can write (1.1) as

$$Y = 1\left[\beta_0 + \beta_1 X > 0\right],$$

[4] If $\mathcal{B}$ is bounded the absolute continuity condition should be understood to be required to hold with respect to the uniform measure on $\mathcal{B}$.

[5] Such normalizations are not strictly required when allowing for set identification, but are wise to impose in order to enable comparison of set and point-identifying models.

with $\beta = \left(\beta_0, \beta_1'\right)'$. Suppose that $\mathcal{F}$ is the class of bivariate normal distributions whose first component has unit variance. Then defining α_0, α_1 as the means of β_0, β_1, respectively, we have the representation

$$Y = 1\left[\alpha_0 + \alpha_1 X > -U_0 - U_1 X\right],$$

where $U_0 \equiv \beta_0 - \alpha_0$ and $U_1 \equiv \beta_1 - \alpha_1$ are mean-zero bivariate normally distributed with the same variance as $\beta = (\beta_0, \beta_1)$. We then have from Restriction A3 that $U \perp\!\!\!\perp Z$, and we can parameterize the distribution $U \equiv (U_0, U_1)$ as

$$U_0 \sim N(0,1) \qquad U_1 | U_0 = u_0 \sim N(\gamma_0 u_0, \gamma_1),$$

equivalently:

$$U \sim N\left(\left(\begin{array}{c} 0 \\ 0 \end{array}\right), \left(\begin{array}{cc} 1 & \gamma_0 \\ \gamma_0 & \gamma_1 + \gamma_0^2 \end{array}\right)\right).$$

Knowledge of the parameter vector $(\alpha_0, \alpha_1, \gamma_0, \gamma_1)$ would then suffice for the determination of F_β, so the identified set for F_β can be succinctly expressed as the identified set for $(\alpha_0, \alpha_1, \gamma_0, \gamma_1)$. ■

3 Identification

For identification analysis it will be useful to consider the correspondence

$$\mathcal{T}(w, x, y) \equiv \text{cl}\left\{\left(b_0, b_1', b_2'\right)' \in \mathcal{B} : y = 1\left[b_0 + x b_1 + w b_2 > 0\right]\right\}, \tag{3.1}$$

which is the closure of the halfspace of $\mathcal{B}$ on which $2y - 1$ and $b_0 + x b_1 + w b_2$ have the same sign. Application of this correspondence to random elements (W, X, Y) yields a *random closed set* $\mathcal{T}(W, X, Y)$. For any realization of the exogenous variables $\tilde{z} \in \tilde{\mathcal{Z}} \equiv \text{Supp}(W, Z)$, the conditional distribution of this random set given $\tilde{Z} = \tilde{z}$ is completely determined by the distribution of (W, X, Y) given $\tilde{Z} = \tilde{z}$, which is identified given knowledge of F^0_{WXYZ} under Restriction A1. The identified set for F_β, denoted $\mathcal{F}^*$, is then the set of measures $F \in \mathcal{F}$ that are *selectionable* from the conditional distribution of $\mathcal{T}(W, X, Y)$ given $\tilde{Z} = \tilde{z}$ for almost every $\tilde{z} \in \tilde{\mathcal{Z}}$. That is, $F \in \mathcal{F}^*$

if and only if $F \in \mathcal{F}$ and there exists a random variable $\tilde{\beta}$ realized on $(\Omega, \Im, \mathbb{P})$ and distributed F such that $\mathbb{P}\left(\tilde{\beta} \in \mathcal{T}(W, X, Y) | \tilde{z}\right) = 1$, a.e. $\tilde{z} \in \tilde{\mathcal{Z}}$.[6]

As done in CRS for utility-maximizing discrete choice models without random coefficients and in Chesher and Rosen (2012a) for single equation IV models more generally, we can characterize the identified set through the use of *conditional containment functional inequalities*. By the same steps taken in Theorem 1 of CRS, a distribution F is selectionable from the conditional distribution of $\mathcal{T}(W, X, Y)$ given $\tilde{Z} = \tilde{z}$, if and only if for all closed sets $\mathcal{S} \subseteq \mathcal{B}$,

$$F(\mathcal{S}) \geq \mathbb{P}\left[\mathcal{T}(W, X, Y) \subseteq \mathcal{S} | \tilde{z}\right]. \tag{3.2}$$

Using the conditional containment inequality (3.2) reduces the problem of determining which F are selectionable from $\mathcal{T}(W, X, Y)$ to the verification of a collection of conditional moment inequalities. In Chesher, Rosen, and Smolinski (forthcoming), Chesher and Rosen (2012a), and Chesher and Rosen (2012b) we devised algorithms to determine which test sets $\mathcal{S}$ are sufficient in the contexts of the models in those papers to imply (3.2) for *all* possible test sets $\mathcal{S}$. The collection of such sets, referred to as *core-determining* sets, is crucially dependent on the support of the random set under consideration. By the same reasoning as in those papers, it is sufficient to focus on test sets that are unions of sets that belong to the support of $\mathcal{T}(W, X, Y)$ conditional on the realization of exogenous variables $\tilde{Z}$. For any realization (w, z) this is given by the collection of test sets

$$\mathsf{T}(w, z) \equiv \{\mathcal{T}(w, x, y) : y \in \{0, 1\} \wedge x \in \text{Supp}(X | w, z)\}. \tag{3.3}$$

We do not require that the conditional support of X given (w, z) coincide with its unconditional support, but in that case $\text{Supp}(X|w, z)$ in (3.3) can be replaced with $\mathcal{X}$, and the collection of sets $\mathsf{T}(w, z)$ does not vary with (w, z). The larger the conditional support $\text{Supp}(X|w, z)$, the larger will be the core-determining collection of test sets.

Given any (w, z), each element of $\mathsf{T}(w, z)$ is a half-space in $\mathcal{B}$, so the required test sets $\mathcal{S}$ take

[6] The requirement that $\tilde{\beta}$ lives on $(\Omega, \Im, \mathbb{P})$ is innocuous. If this were not the case, then one could simply redefine the initial probability space as the product of $(\Omega, \Im, \mathbb{P})$ and the space on which $\tilde{\beta}$ lives.

the form of unions of these halfspaces:

$$\mathcal{S} = \mathcal{T}_1 \cup \cdots \cup \mathcal{T}_J,$$

where each $\mathcal{T}_j \in \mathsf{T}(w,z)$, $j = 1, ..., J$. Alternatively, we can write

$$\mathcal{S} = (\mathcal{T}_1^c \cap \cdots \cap \mathcal{T}_J^c)^c,$$

where for any set $\mathcal{A} \subseteq \mathcal{B}$, $\mathcal{A}^c$ denotes the complement of $\mathcal{A}$ in $\mathcal{B}$. This is convenient because the complement of each $\mathcal{T}_j$, $\mathcal{T}_j^c$, is also a halfspace, and the intersection of halfspaces is a convex polytope. Thus the collection of core-determining test sets $\mathcal{S}$ are all complements of intersections of halfspaces, equivalently complements of convex polytopes. The formal result follows.

Theorem 1 *Let Restrictions A1-A3 hold. Then the identified set for F_β is*

$$\mathcal{F}^* = \left\{F \in \mathcal{F} : \forall \mathcal{S} \in \mathsf{T}^{\cup}(w,z), \quad F(\mathcal{S}) \geq \mathbb{P}\left[\mathcal{T}(W,X,Y) \subseteq \mathcal{S} | w,z\right], \text{ a.e. } (W,Z)\right\} \tag{3.4}$$

where $\mathsf{T}^{\cup}(w,z)$ denotes the collection of sets that are unions of members of $\mathsf{T}(w,z)$. Equivalently,

$$\mathcal{F}^* = \left\{F \in \mathcal{F} : \forall \mathcal{S} \in \mathsf{T}^{\cap}(w,z), \quad F(\mathcal{S}) \leq \mathbb{P}\left[\mathcal{T}(W,X,Y) \cap \mathcal{S} \neq \emptyset | w,z\right], \text{ a.e. } (W,Z)\right\}, \tag{3.5}$$

where $\mathsf{T}^{\cap}(w,z)$ denotes the collection of sets that are intersections of members of $\mathsf{T}^{\mathbf{c}}(w,z)$, where

$$\mathsf{T}^{\mathbf{c}}(w,z) \equiv \left\{\mathcal{T}^c(w,x,y) : y \in \{0,1\} \wedge x \in Supp(X|w,z)\right\},$$

which is the collection of sets that are complements of those in $\mathsf{T}(w,z)$.

The theorem follows from consideration of Theorems 1 and 2 of CRS, adapted to the random set $\mathcal{T}(W,X,Y)$ defined in (3.1), which make use of Artstein's inequality (Artstein (1983)) to prove sharpness.[7] The characterization of test sets for the containment functional characterization (3.4)

[7] See also Norberg (1992) and Molchanov (2005) Section 1.4.8.

of CRS Theorem 2 stipulates that a core determining collection of test sets $\mathcal{S}$ is given by those which are (i) unions of elements of $\mathsf{T}(w,z)$ (ii) such that the union of the interiors of component sets is a connected set. In this paper condition (ii) may be ignored because the sets $\mathcal{T}(w,x,y)$ and $\mathcal{T}(w',x',y')$ are all halfspaces through the origin, ensuring that $\mathcal{T}(w,x,y) \cap \mathcal{T}(w',x',y')$ has open interior except in the special case $(x,w) = (x',w')$ and $y' = 1-y$, in which case $\mathcal{T}(w,x,y) \cup \mathcal{T}(w',x',y') = \mathcal{B}$. The test set $\mathcal{B}$ can indeed be safely discarded from consideration because from $F(\mathcal{B}) = 1$, (3.4) is trivially satisfied. The containment functional characterization (3.4) and capacity functional characterization (3.5) are equivalent, although the latter form may prove convenient from a computational standpoint since the collection of associated test sets $\mathsf{T}^{\cap}$ are convex polytopes.

In general Theorem 1 delivers a collection of conditional moment inequalities characterizing the identified set, with one such inequality conditional on the realization of exogenous variables (w,z) for each element of $\mathsf{T}^{\cup}(w,z)$ in (3.4), equivalently one conditional moment inequality for each element of $\mathsf{T}^{\cap}(w,z)$ in (3.5). In some important special cases, considered in the examples below, characterization of the identified set can be further simplified.

Example 2 (No endogenous covariates): A leading and well-studied example is the case where there are no endogenous variables X. Then for each (w,z) we have that

$$\mathsf{T}(w,z) = \left\{ \left\{ b \in \mathcal{B} : b_0 + w b_2 \geq 0 \right\}, \left\{ b \in \mathcal{B} : b_0 + w b_2 \leq 0 \right\} \right\},$$

where b is of the form $b = (b_0, b_2')'$. The intersection of these sets is $\{b \in \mathcal{B} : b_0 + w b_2 = 0\}$, which has zero measure F_β under restriction A2, and their union is $\mathcal{B}$, which has measure 1. It follows from similar reasoning as in Theorem 6 of Chesher and Rosen (2012b) that for any (w,z) the inequalities of the characterizations of Theorem 1 produce moment *equalities*. Consider for example the containment functional inequalities of (3.4) delivered by $\mathcal{S} \in \mathsf{T}^{\cup}(w,z)$:

$$\begin{aligned}
F\left(\left\{b \in \mathcal{B} : b_0 + w b_2 \geq 0\right\}\right) &\geq \mathbb{P}\left[\mathcal{T}(W,Y) \subseteq \left\{b \in \mathcal{B} : b_0 + w b_2 \geq 0\right\} | w, z\right] = \mathbb{P}\left[Y = 1 | w, z\right], \\
F\left(\left\{b \in \mathcal{B} : b_0 + w b_2 \leq 0\right\}\right) &\geq \mathbb{P}\left[\mathcal{T}(W,Y) \subseteq \left\{b \in \mathcal{B} : b_0 + w b_2 \leq 0\right\} | w, z\right] = \mathbb{P}\left[Y = 0 | w, z\right], \\
F(\mathcal{B}) &\geq \mathbb{P}\left[\mathcal{T}(W,Y) \subseteq \mathcal{B} | w, z\right] = 1.
\end{aligned}$$

The last inequality is trivially satisfied for all $F \in \mathcal{F}$. Both the right-hand sides and the left-hand sides of the first two inequalities clearly sum to one, implying that these inequalities must in fact hold with equality, giving

$$\begin{aligned} F\left(\{b \in \mathcal{B} : b_0 + wb_2 \geq 0\}\right) &= \mathbb{P}\left[Y = 1 | w, z\right], \\ F\left(\{b \in \mathcal{B} : b_0 + wb_2 \leq 0\}\right) &= \mathbb{P}\left[Y = 0 | w, z\right]. \end{aligned} \tag{3.6}$$

When there are no excluded exogenous variables z and F_β is not restricted to a parametric family, these equations coincide with the identifying equations in Ichimura and Thompson (1998) and Gautier and Kitamura (forthcoming), and Ichimura and Thompson (1998) provide sufficient conditions for point identification.[8] When F is parametrically restricted these equalities are likelihood contributions, e.g. integrals with respect to the normal density in Hausman and Wise (1978) or Lerman and Manski (1981), and less stringent conditions are required for point identification. In the absence of sufficient conditions for point identification, the moment equalities (3.6) a.e. (W, Z) nonetheless fully characterize the identified set. ■

Example 3 (One endogenous covariate with arbitrary exogenous covariates): Consider the common setting where there is a single endogenous regressor, $X \in \mathbb{R}$, as well as some exogenous regressors W, a random k_w-vector. Then given any (w, z) the collection of sets $\mathsf{T}(w, z)$ is given by

$$\mathsf{T}(w, z) \equiv \bigcup_{x \in \text{Supp}(X|w,z)} \left\{ \left\{ \left(b_0, b_1, b_2'\right)' \in \mathcal{B} : b_0 + xb_1 + wb_2 \geq 0 \right\}, \left\{ \left(b_0, b_1, b_2'\right)' \in \mathcal{B} : b_0 + xb_1 + wb_2 \leq 0 \right\} \right\}.$$

Consider now a test set $\mathcal{S}$ which is one of the core-determining sets in $\mathsf{T}^{\cup}(w, z)$ and hence an arbitrary union of sets in $\mathsf{T}(w, z)$. Any such $\mathcal{S}$ can be written as the set of $(b_0, b_1, b_2')' \in \mathcal{B}$ that

[8] The restrictions used to ensure point identification include the requirements that for some fixed $c \in \mathbb{R}^{k_w}$, $F_\beta\left(\{b : c'b > 0\} = 1\right)$, and that the distribution of W has an absolutely continuous component with everywhere positive density. Our characterizations of the identified set, given by (3.6) in the case of only exogenous covariates, do not require these restrictions.

satisfy one of the inequalities

$$\begin{aligned} b_0 + wb_2 + \max_j \{x_{1j} b_1\} &\geq 0, \\ b_0 + wb_2 + \min_m \{x_{0m} b_1\} &\leq 0, \end{aligned}$$

for some collections of values for X, $\mathcal{X}_1 \equiv \{x_{11}, ..., x_{1J}\}$ and $\mathcal{X}_0 \equiv \{x_{01}, ..., x_{0M}\}$, with the maxima and minima taken over $j = 1, ..., J$ and $m = 1, ..., M$. If $b_1 \geq 0$ this simplifies to

$$\max_j x_{1j} b_1 \geq -b_0 - wb_2 \geq \min_m x_{0m} b_1,$$

while if $b_1 < 0$ the inequalities can be written

$$\min_j x_{1j} b_1 \geq -b_0 - wb_2 \geq \max_m x_{0m} b_1.$$

Without loss of generality, assume that the components of $\mathcal{X}_0$ and $\mathcal{X}_1$ are ordered from smallest to largest. It follows that we can write any $\mathcal{S} \in \mathsf{T}^{\cup}(w, z)$ as a union of no more than 4 elements of $\mathsf{T}(w, z)$ since by the above reasoning for any such $\mathcal{X}_0$ and $\mathcal{X}_1$ we have

$$\mathcal{S} = (\cup_j \mathcal{T}(w, x_j, 1)) \cup (\cup_m \mathcal{T}(w, x_m, 0)) = \mathcal{T}(w, x_{11}, 1) \cup \mathcal{T}(w, x_{1J}, 1) \cup \mathcal{T}(w, x_{01}, 0) \cup \mathcal{T}(w, x_{0M}, 0).$$

From this it follows that we need only consider for each (w, z) test sets $\mathcal{S}$ of the form

$$\mathcal{S} = \mathcal{T}(w, x_1, 1) \cup \mathcal{T}(w, x_2, 1) \cup \mathcal{T}\left(w, x_1', 0\right) \cup \left(w, x_2', 0\right),$$

where $x_2 \geq x_1$ and $x_2' \geq x_1'$

Example 1, continued: If we restrict attention to cases with no exogenous covariates W, there is in fact further simplification of the list of core-determining sets. To see why, note that in this

case the collection $\mathsf{T}(w,z) = \mathsf{T}(z)$ for any z reduces to

$$\mathsf{T}(z) \equiv \bigcup_{x \in \text{Supp}(X|z)} \left\{ \left\{ (b_0, b_1)' \in \mathcal{B} : b_0 + x b_1 \geq 0 \right\}, \left\{ (b_0, b_1)' \in \mathcal{B} : b_0 + x b_1 \leq 0 \right\} \right\}.$$

Each element of $\mathsf{T}(z)$ is thus a halfspace in $\mathbb{R}^2$ defined by a separating hyperplane through the origin intersected with $\mathcal{B}$. The union of an arbitrary number of such halfspaces can be equivalently written as the union of no more than two such halfspaces. Therefore the collection of core-determining sets $\mathsf{T}^{\cup}(w,z) = \mathsf{T}^{\cup}(z)$ is given by the collection of test sets that can be written as either elements of $\mathsf{T}(z)$ or unions of a pair of elements in $\mathsf{T}(z)$,

$$\mathsf{T}^{\cup}(z) = \bigcup_{\substack{x_1, x_2 \in \text{Supp}(X|z) \\ y_1, y_2 \in \{0,1\}}} \left\{ \mathcal{T}(x_1, y_1) \cup \mathcal{T}(x_2, y_2) \right\}, \tag{3.7}$$

where for any $x \in \mathcal{X}$ and $y \in \{0,1\}$,

$$\mathcal{T}(x,y) = \text{cl}\left\{ (b_0, b_1)' \in \mathcal{B} : y = 1\left[b_0 + x b_1 > 0\right] \right\}.$$

The characterization applies for either continuous or discrete X, but if X is discrete with K points of support there are no more than $2K^2$ sets in $\mathsf{T}^{\cup}(z)$ for any $z \in \mathcal{Z}$. This follows from noting there are $2K$ unique (x,y) pairs and the number of all pairwise unions (including the union of each set with itself) is $(2K)^2/2$, with division by two from the observation that for any (x_1, y_1) and (x_2, y_2), $\mathcal{T}(x_1, y_1) \cup \mathcal{T}(x_2, y_2) = \mathcal{T}(x_2, y_2) \cup \mathcal{T}(x_1, y_1)$. ■

In the numerical illustrations that follow we consider various instances of Example 1, where there are no exogenous covariates W and where F is restricted to a parametric (specifically Gaussian) family. In the illustration we investigate identified sets for averages of (β_0, β_1), and we show that this affords further computational simplification, in the sense that for any fixed candidate values of $(E\beta_0, E\beta_1)$, we need only consider test sets $\mathcal{S}$ that are unions of two elements of $\mathsf{T}(w,z)$ in order to check whether any $(E\beta_0, E\beta_1)$ belongs to the identified set.

4 Numerical Illustrations

To investigate the identifying power of our binary outcome random coefficient model with instruments, we consider Example 1, where

$$Y = 1\left[\beta_0 + \beta_1 X > 0\right].$$

with X a univariate random variable and $(\beta_0, \beta_1)'$ bivariate normally distributed with mean (α_0, α_1), $cov\left(\beta_0, \beta_1\right) = \gamma_0$, $var\left(\beta_0\right) = 1$, and $var\left(\beta_1\right) = \gamma_1 + \gamma_0^2$. We can then equivalently write the model as

$$Y = 1\left[U_0 + U_1 X > -\alpha_0 - \alpha_1 X\right],$$

where $U = (U_0, U_1)$ are bivariate normal with zero mean and the same variance as (β_0, β_1). We then define

$$G_U\left(\mathcal{U}, \theta\right) \equiv F_\beta\left(\left\{(u_0 + \alpha_0, u_1 + \alpha_1) : u \in \mathcal{U}\right\}\right)$$

as the probability that U belongs the the set $\mathcal{U}$ where $\theta = (\alpha_0, \alpha_1, \gamma_0, \gamma_1)$ and when β is distributed F_β with mean α and variance governed by parameters (γ_0, γ_1). Given the restriction that $\beta = (\beta_0, \beta_1)'$ is bivariate normally distributed, knowledge of θ implies knowledge of F_β. We thus consider the identified set for θ, denoted Θ^*, and focus our attention in particular on the identified set for (α_0, α_1), the projection of the first two elements of Θ^* on $\mathbb{R}^2$.

4.1 Data-Generating Processes

Our examples employ data-generating processes with a triangular structure for X as a function of instrument Z, as follows.

$$\begin{aligned} X &= x_k \text{ iff } c_{k-1} < X^* \leq c_k, \quad k \in \{1, \ldots, K\}, \\ X^* &= \delta_1 Z + \delta_2 U_0 + \delta_3 U_1 + \delta_4 V, \end{aligned}$$

where

$$\begin{bmatrix} U_0 \\ U_1 \\ V \end{bmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \gamma_0 & 0 \\ \gamma_0 & \gamma_1 + \gamma_0^2 & 0 \\ 0 & 0 & 1 \end{pmatrix}\right).$$

We do four calculations. In the first two calculations parameters are set such that X is endogenous, with the instrument having varying degrees of strength in terms of predictive ability for X. In both cases parameter values are set as follows:

$$\alpha_0 = 0, \quad \alpha_1 = -1, \quad \gamma_0 = -1 \quad \gamma_1 = 1,$$

$$K = 4, \quad (x_1, x_2, x_3, x_4) = (-1, 0, 1, 2), \quad (c_0, c_1, c_2, c_3, c_4) = (-\infty, -1, 0, 1, \infty).$$

The support of the instrument Z is specified as

$$\mathcal{Z} = \{-2, -1, 1, 2\}.$$

In the **stronger** instrument case there is:

$$\delta \equiv (\delta_1, \delta_2, \delta_3, \delta_4) = (1, 0.577, -0.577, 0.577),$$

and in the **weaker** instrument case:

$$\delta \equiv (\delta_1, \delta_2, \delta_3, \delta_4) = (1.5, 0.577, -0.577, 0.462).$$

In the stronger instrument case the coefficient on the instrumental variable in the ordered probit equation for X is larger and the variance of the unobservable variable V is slightly smaller. The result is that in the stronger instrument case the instrumental variable is a significantly better predictor of the value of the endogenous explanatory variable X. Table 1 shows the probability that X takes its four values conditional on the value of the instrumental variable at the two settings for the instrumental variable.

		$z=-2$	$z=-1$	$z=+1$	$z=+2$
weaker instrument	$x=-1$	.760	.500	.079	.017
	$x=0$	.161	.260	.161	.062
	$x=1$	.062	.161	.260	.161
	$x=2$	.017	.079	.500	.760
stronger instrument	$x=-1$	.928	.642	.034	.002
	$x=0$	.058	.221	.103	.013
	$x=1$	.013	.103	.221	.058
	$x=2$	.002	.034	.642	.928

Table 1: Conditional probabilities $\mathbb{P}[X=x|z]$.

For the third and fourth cases we calculate the identified set for probabilities generated by a structure in which X is exogenous. The identified sets are for (α_0, α_1) when this restriction is not imposed. We show in Appendix C that with $X \perp\!\!\!\perp \beta$ known, there is point-identification of the full parameter vector θ. The identified sets obtained without this restriction thus help to illustrate the identifying power of the exogeneity restriction for a pair of DGPs in which it does hold.

Everything is as in the first two DGPs where X is endogenous except for the following parameter settings.

$$\delta \equiv (\delta_1, \delta_2, \delta_3, \delta_4) = \left(1, 0, 0, 2^{1/2}\right).$$

The variance of the unobserved element in the equation for X is 2, which is the same as in the first two cases considered, but where the independence restriction $X \perp\!\!\!\perp \beta$ is false. The probabilities $\mathbb{P}[X=x_k|z]$ of Table 1 hold in both endogenous and exogenous X designs.

4.2 Calculation of Probabilities

There is

$$\mathbb{P}[X=x_k|z] = \Phi\left(\frac{c_k - \delta_1 z}{\lambda^{1/2}}\right) - \Phi\left(\frac{c_{k-1} - \delta_1 z}{\lambda^{1/2}}\right),$$

where $\Phi(\cdot)$ is the standard normal distribution function and λ is defined as

$$\lambda \equiv \delta_2^2 + 2\delta_2\delta_3\gamma_0 + \delta_3^2(\gamma_1 + \gamma_0^2) + \delta_4^2. \tag{4.1}$$

Now consider $\mathbb{P}[Y = 0 \wedge X = x_k|z]$. There is given $Z = z$:

$$\{Y = 0 \wedge X = x_k\} \Leftrightarrow \{\alpha_0 + U_0 + x_k(\alpha_1 + U_1) \leq 0\} \wedge \left\{c_{k-1} - \delta_1 z < \tilde{V} \leq c_k - \delta_1 z\right\},$$

where

$$\tilde{V} \equiv \delta_2 U_0 + \delta_3 U_1 + \delta_4 V.$$

It then follows that

$$(Y = 0 \wedge X = x_k) \Leftrightarrow (Q_k \leq -\alpha_0 - \alpha_1 x_k) \wedge \left(c_{k-1} - \delta_1 z < \tilde{V} \leq c_k - \delta_1 z\right),$$

where

$$Q_k \equiv U_0 + x_k U_1.$$

Since

$$\begin{pmatrix} \tilde{V} \\ Q_k \end{pmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \lambda & \delta_2 + \gamma_0(\delta_3 + \delta_2 x_k) - \delta_3 x_k(\gamma_1 + \gamma_0^2) \\ \delta_2 + \gamma_0(\delta_3 + \delta_2 x_k) - \delta_3 x_k(\gamma_1 + \gamma_0^2) & (1 + x_k\gamma_0)^2 + x_k^2\gamma_1 \end{bmatrix}\right),$$

$\mathbb{P}[Y = 0 \wedge X = x_k|z]$ can be calculated as the difference between two normal orthant probabilities. We then have

$$\mathbb{P}[Y = 1 \wedge X = x_k|z] = \mathbb{P}[X = x_k|z] - \mathbb{P}[Y = 0 \wedge X = x_k|z].$$

In our R programs the required bivariate normal orthant probabilities are calculated using the `pmvnorm` program provided in the `mvtnorm` package, Genz, Bretz, Miwa, Mi, Leisch, Scheipl, and Hothor (2012), which implements computation of multivariate normal and t probabilities from Genz and Bretz (2009).[9]

[9] R: A Language and Environment for Statistical Computing R Core Team (2012).

4.3 Calculation of projections

In this model with F_β bivariate normal, the distribution of random coefficients is fully conveyed by the value of the parameter $\theta_0 = (\alpha_0, \alpha_1, \gamma_0, \gamma_1)$. We calculate two dimensional projections of the four dimensional identified set for θ_0, giving results here for the projection onto the plane on which lie (α_0, α_1), which is equivalently the identified set for the mean of the random coefficients (β_0, β_1).

We calculate the projections of the identified set as follows. Let θ denote a conjectured value of the parameter vector $(\alpha_0, \alpha_1, \gamma_0, \gamma_1)$.

The full 4D identified set is

$$\Theta^* = \left\{ \theta \in \Theta : \forall \mathcal{S} \in \mathsf{S}, \quad G_U(\mathcal{S}, \theta) \geq \max_{z \in \mathcal{Z}} \mathbb{P}[\mathcal{T}(W, X, Y) \subset \mathcal{S} | z] \right\}$$

where $\mathsf{S} = \mathsf{T}^{\cup}(z)$ is a collection of 32 core determining sets of the form described for Example 1 in Section 3, specifically (3.7), in the present case where X has four points of support. $G_U(\mathcal{S}, \theta)$ is the probability mass placed on the set $\mathcal{S}$ by a bivariate normal distribution with parameters θ and the probabilities $\mathbb{P}[\mathcal{T}(W, X, Y) \subset \mathcal{S} | Z = z]$, $z \in \mathcal{Z}$, which are identified under Restriction A1.

For computational purposes we make use of the following discrepancy measure:

$$D(\theta) \equiv \max_{z \in \mathcal{Z}, \mathcal{S} \in \mathsf{S}} \left(\mathbb{P}[\mathcal{T}(W, X, Y) \subset \mathcal{S} | z] - G_U(\mathcal{S}, \theta) \right). \tag{4.2}$$

For values of θ in the identified set $D(\theta) \leq 0$. For values of θ outside the identified set we have that for at least one set $\mathcal{S} \in \mathsf{S}$ and for some $z \in \mathcal{Z}$,

$$G_U(\mathcal{S}, \theta) - \mathbb{P}[\mathcal{T}(W, X, Y) \subset \mathcal{S} | z] < 0,$$

and so for such values $D(\theta) > 0$. The full four dimensional identified set can therefore be characterized as follows.

$$\Theta^* = \{ \theta \in \Theta : D(\theta) \leq 0 \}.$$

To compute identified sets for sub-vectors of parameters, let θ_c denote a sub-vector of θ, that

is one or more elements of θ, and let θ_{-c} denote that vector containing the remaining elements of θ. The projection of the identified set onto the space in which θ_c resides is the set of values of θ_c for which there exists θ_{-c} such that $\theta = (\theta_c, \theta_{-c})$ lies in the identified set Θ^*. We calculate this set as the set of values θ_c for which the value of $\min_{\theta_{-c}} D(\theta_c, \theta_{-c})$ is nonpositive:

$$\Theta_c^* = \left\{ \theta_c : \min_{\theta_{-c}} D(\theta_c, \theta_{-c}) \leq 0 \right\}, \tag{4.3}$$

where $D(\theta_c, \theta_{-c})$ is to be understood as the function defined in (4.2) applied to that value of θ with sub-vectors equal to θ_c and θ_{-c}. We perform this minimization using the `optim` function in base R.

Figure 1 shows the projections of the identified set in the two cases where X is endogenously determined. The data generating value $(\alpha_0, \alpha_1) = (0, -1)$ is plotted as well. In the stronger instrument case (drawn in red) the projection is smaller in area. Most values in the projection for the stronger instrument case lie inside the projection for the weaker instrument case, but at high values of α_0 there is a very small region of the stronger instrument projection which is not contained within the weaker instrument projection. Figure 2 similarly illustrates projections of the identified set for the exogenous X process. In this case the projection of the identified set for the stronger instrument case is a strict subset of that in the weaker instrument case.

In all cases, both with endogenous and exogenous X, the projections do not contain any positive values of α_1. That is, the model allows one to sign α_1, so that the hypothesis $H_0 : \alpha_1 \geq 0$ is falsifiable.

5 Conclusion

In this paper we have provided set identification analysis for a model of binary response featuring random coefficients and potentially endogenous regressors. The regressors in question are not restricted to be distributed independently of the random coefficients. We showed that with an instrumental variable restriction we can apply analysis along the lines of that in CRS and Chesher and Rosen (2012a) to characterize the identified set as the those distributions satisfying a collection of conditional moment inequalities. In our examples of Section 4 there are 32 conditional

moment inequalities, one for each core-determining set, which hold conditional on any value of the instrument. While our focus was on identification, recently developed approaches for estimation and inference based on such characterizations, such as those of Andrews and Shi (forthcoming) and Chernozhukov, Lee, and Rosen (forthcoming), are applicable. In some settings the number of core-determining sets in the full characterization may be quite large, necessitating some care in choosing the number to employ in small samples. Issues that arise due to many moment inequalities have been investigated in an asymptotic paradigm by Menzel (2009). Here the number of conditional moment inequalities may be quite large, but is necessarily finite, and future research on finite sample approximations for inference and computational issues is warranted.

We have further provided some numerical illustrations of identified sets under particular data generation processes. We gave an overview of the computational approach we used for computing these identified sets, and details of these approaches are described in Appendix B.

Although our computational approaches were adequate for the examples considered, we have little doubt that these approaches may be improved, either by developing more efficient implementations, or by devising new computational approaches altogether. Nonetheless, the illustrations serve to illustrate the feasibility of computing identified sets in one particular setting in the general class of instrumental variable models studied in Chesher and Rosen (2012a). These instrumental variable models can admit high-dimensional unobserved heterogeneity, for example through a random-coefficients specification such as the one studied in this paper.

References

Andrews, D. W. K., and X. Shi (forthcoming): "Inference for Parameters Defined by Conditional Moment Inequalities," *Econometrica*.

Artstein, Z. (1983): "Distributions of Random Sets and Random Selections," *Israel Journal of Mathematics*, 46(4), 313–324.

Bajari, P., J. Fox, K.-i. Kim, and S. P. Ryan (2012): "The Random Coefficients Logit Model is Identified," *Journal of Econometrics*, 166(2), 204–212.

Beresteanu, A., I. Molchanov, and F. Molinari (2011): "Sharp Identification Regions in Models with Convex Moment Predictions," *Econometrica*, 79(6), 1785–1821.

——— (2012): "Partial Identification Using Random Set Theory," *Journal of Econometrics*, 166(1), 17–32.

Berry, S., and P. Haile (2009): "Nonparametric Identification of Multiple Choice Demand Models with Heterogeneous Consumers," NBER working paper w15276.

——— (2010): "Identification in Differentiated Markets Using Market Level Data," NBER working paper w15641.

Berry, S., J. Levinsohn, and A. Pakes (1995): "Automobile Prices in Market Equilibrium," *Econometrica*, 63(4), 841–890.

Berry, S. T. (1994): "Estimating Discrete Choice Models of Product Differentiation," *Rand Journal of Economics*, 25(2), 242–262.

Briesch, R. A., P. K. Chintagunta, and R. L. Matzkin (2010): "Nonparametric Discrete Choice Models with Unobserved Heterogeneity," *Journal of Business and Economic Statistics*, 28(2), 291–307.

Chernozhukov, V., S. Lee, and A. Rosen (forthcoming): "Intersection Bounds, Estimation and Inference," *Econometrica*.

Chesher, A. (2010): "Instrumental Variable Models for Discrete Outcomes," *Econometrica*, 78(2), 575–601.

——— (forthcoming): "Semiparametric Structural Models of Binary Response: Shape Restrictions and Partial Identification," *Econometric Theory*.

Chesher, A., and A. Rosen (2012a): "Generalized Instrumental Variable Models," in preparation.

——— (2012b): "Simultaneous Equations Models for Discrete Outcomes: Coherence, Completeness, and Identification," CeMMAP working paper CWP21/12.

Chesher, A., A. Rosen, and K. Smolinski (forthcoming): "An Instrumental Variable Model of Multiple Discrete Choice," *Quantitative Economics*.

Fox, J. T., and A. Gandhi (2012): "Nonparametric Identification and Estimation of Random Coefficients in Multinomial Choice Models," working paper, University of Michigan.

Galichon, A., and M. Henry (2011): "Set Identification in Models with Multiple Equilibria," *Review of Economic Studies*, 78(4), 1264–1298.

Gautier, E., and Y. Kitamura (forthcoming): "Nonparametric Estimation in Random Coefficients Binary Choice Models," *Econometrica*.

Gautier, E., and E. LePennec (2011): "Adaptive Estimation in the Nonparametric Random Coefficients Binary Choice Model by Needlet Thresholding," working paper, CREST/ENSAE.

Genz, A., and F. Bretz (2009): *Computation of Multivariate Normal and t Probabilities*. Springer-Verlag, Heidelberg, Germany, Lecture Notes in Statistics, Vol. 195.

Genz, A., F. Bretz, T. Miwa, X. Mi, F. Leisch, F. Scheipl, and T. Hothor (2012): *mvtnorm: Multivariate Normal and t Distributions. R package version 0.9-9992.*

Hausman, J., and D. Wise (1978): "A Conditional Probit Model for Qualitative Choice: Discrete Decisions Recognizing Interdependence and Heterogeneous Preferences," *Econometrica*, 46(2), 403–426.

Hoderlein, S. (2009): "Endogenous Semiparametric Binary Choice Models with Heteroscedasticity," CeMMAP working paper CWP34/09.

Ichimura, H., and T. S. Thompson (1998): "Maximum Likelihood Estimation of a Binary Choice Model with Random Coefficients of Unknown Distribution," *Journal of Econometrics*, 86, 269–295.

Johnson, S. G. (2011): *Package 'cubature', R package version 1.1-1.*

Lerman, S. R., and C. F. Manski (1981): "On the Use of Simulated Fequencies to Approximat Choice Probabilities," in *Structural Analysis of Discrete Data and Econometric Applications*, ed. by C. F. Manski, and D. L. McFadden, pp. 305–319. MIT Press.

McFadden, D. L. (1976): "Quantal Choice Analysis: A Survey.," *Annals of Economic and Social Measurement*, 5, 363–390.

Menzel, K. (2009): "Estimation and Inference with Many Weak Moment Inequalities," working paper, MIT.

Molchanov, I. S. (2005): *Theory of Random Sets*. Springer Verlag, London.

Nevo, A. (2011): "Empirical Models of Consumer Behavior," *Annual Review of Economics*, 3, 51–75.

Norberg, T. (1992): "On the Existence of Ordered Couplings of Random Sets – with Applications," *Israel Journal of Mathematics*, 77, 241–264.

Pateiro-Lopez, B., and A. Rodriguez-Casal (2009): *alphahull: Generalization of the Convex Hull of a Sample of Points in the Plane. R package version 0.2-0.*

Quandt, R. E. (1966): "A Probabilistic Theory of Consumer Behavior," *Quarterly Journal of Economics*, 70, 507–536.

R Core Team (2012): *R: A Language and Environment for Statistical Computing*R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0.

Appendix A: Proof of Theorem 1

Following the same steps as in the proof of Theorem 1 of Chesher, Rosen, and Smolinski (forthcoming) applied to the random set $\mathcal{T}(W, X, Y)$ and exogenous variables $\tilde{Z} = (W, Z)$ in place of

$\mathcal{T}_v(Y, X; u)$ and instruments Z in the notation of that paper delivers

$$\mathcal{F}^* = \{F \in \mathcal{F} : \forall \mathcal{S} \in \mathsf{F}(\mathcal{B}), \quad F(\mathcal{S}) \geq \mathbb{P}[\mathcal{T}(W, X, Y) \subseteq \mathcal{S} | w, z], \text{ a.e. } (W, Z)\},$$

where $\mathsf{F}(\mathcal{B})$ denotes all closed subsets of $\mathcal{B}$. Application of Theorem 2 of Chesher, Rosen, and Smolinski (forthcoming), specifically part (i), then further gives that $\mathsf{F}(\mathcal{B})$ above may be replaced with unions of members of the support of $\mathcal{T}(W, X, Y)$. Then, using the same reasoning as in Lemma 1 of Chesher and Rosen (2012b), it follows that when considering probabilities conditional on $(W, Z) = (w, z)$, $\mathsf{F}(\mathcal{B})$ can be replaced by unions of elements of the conditional support of $\mathcal{T}(W, X, Y)$ given the realization of the exogenous variables, namely $\mathsf{T}^{\cup}(w, z)$. The representation

$$\mathcal{F}^* = \left\{F \in \mathcal{F} : \forall \mathcal{S} \in \mathsf{T}^{\cap}(w, z), \quad F(\mathcal{S}) \leq \mathbb{P}[\mathcal{T}(W, X, Y) \cap \mathcal{S} \neq \emptyset | w, z], \text{ a.e. } (W, Z)\right\},$$

follows from the equivalence

$$\mathcal{T}_1 \cup \cdots \cup \mathcal{T}_J = (\mathcal{T}_1^c \cap \cdots \cap \mathcal{T}_J^c)^c,$$

and that for all $\mathcal{S} \subseteq \mathcal{B}$, $F(\mathcal{S}^c) = 1 - F(\mathcal{S})$ and for all $z \in \mathcal{Z}$,

$$\mathbb{P}[\mathcal{T}(W, X, Y) \subseteq \mathcal{S} | Z = z] = 1 - \mathbb{P}[\mathcal{T}(W, X, Y) \cap \mathcal{S} \neq \emptyset | Z = z].$$

Appendix B: Computational Details

In this Section we provide computational details for the numerical illustrations of Section 4 not provided in the main text.

Calculation of probabilities $G_U(\mathcal{S}, \theta)$

Each set $\mathcal{S}$ in the collection $\mathsf{T}^{\cup}(z) = \mathsf{T}^{\cup}$ is the union of one or more contiguous cones centered at the point (α_0, α_1), which we refer to as elementary cones. The slopes of the rays defining the cones are determined entirely by the values of the points of support of X. In the case $K = 4$ there are 8 such cones. For each value of $\theta = (\alpha_0, \alpha_1, \gamma_0, \gamma_1)$ encountered we calculate the probability

mass supported on each of the 8 cones by a bivariate normal density function with mean $(0, 0)$ and variance matrix entirely determined by (γ_0, γ_1). The probability mass supported by a particular set $\mathcal{S}$ at the value of θ is obtained by adding the masses on the appropriate cones. Thus we are able to compute the probability mass $G_U(\mathcal{S}, \theta)$ allocated to each of the 32 core-determining sets by summing probabilities obtained for the 8 elementary cones.

The probability masses on each elementary cone are obtained by numerical integration after re-expressing the integrand in polar coordinates. In our R code the numerical integrations are done using the `adaptIntegrate` function provided in the `cubature` package, Johnson (2011). We have also programmed this calculation in Mathematica using the `NIntegrate` function and an integrand which is the appropriate bivariate normal density function with values outside the cone of interest set to zero using the `Boole` function. We obtained very close agreement.

The numerical integrations are necessarily computationally burdensome and some inaccuracy is inevitable which has a knock-on effect on the determination of membership of projections.

Calculation of Projections

First approximations to the (α_0, α_1)- projections of identified sets were obtained by evaluating over a coarse grid of values of (α_0, α_1). Refinements were then obtained by using a bisection procedure to search down a sequence of rays defined by angles $\gamma \in [0, 2\pi]$, each passing through the probability-generating value $(\alpha_0, \alpha_1) = (0, -1)$ which is known to lie in the projection. Each ray was stepped along until a value of (α_0, α_1) outside the projection was found. A value midway between this value and the last value found in the projection was then evaluated for membership of the projection and by repeated bisection a good approximation to the position of the boundary of the identified set along the ray under consideration was obtained. Sweeps were also made in directions parallel to the α_0 and α_1 axes to refine the boundary approximations in areas where it was relatively nonlinear. These were helpful in confirming the near convexity of the projections which is sufficient for our bisection-along-rays procedure to give a good view of the entire boundary.

The objective function minimized in (4.3) when determining membership of the identified set is not very well behaved. There are certainly points at which it is not differentiable and there

appeared to be some places in which there were small jump discontinuities. One difficulty is that the terms $G_U(\mathcal{S}, \theta)$ depend upon eight numerical integrals of bivariate normal density functions and inaccuracy in calculating these affects the computation of the minimum in (4.3). The effect is likely dependent on the parameter value (α_0, α_1) being considered.

There is plenty of scope for improvement in the numerical procedures employed here. In particular a very small further investment would deliver a much more efficient method of searching down a ray for an initial point outside the identified set. The method we use relies on the near convexity of the projection

There were a few cases in which isolated points appeared to be in the projections. These were examined individually and in most cases by choosing different starting points for the parameters θ_{-c} of the minimization the points were found on recalculation not to be in the projection. The remaining isolated points had a minimized value of the objective function in (4.3) that was very close to zero. The graphs of the identified set shown here were produced by assigning points with values of the minimized objective function less than 0.001 to the projection.

Graphics

The projections as calculated using our approximations are not convex although the departures from convexity are quite small. We do not know whether the projections are in fact convex with the non-convexity arising because of approximation errors. In this circumstance it seems unwise to draw boundaries of projections as the convex hulls of the points calculated to lie in the projections although in fact there is not so great an error produced by proceeding in this way. The projections drawn in Figures 1 and 2 are alpha-convex hulls calculated using the `ahull` function provided in the R package `alphahull`, Pateiro-Lopez and Rodriguez-Casal (2009), with the alphahull parameter set equal to 5. We experimented with different values of this parameter and found that the differences in the illustrations were minute.

Appendix C: Identification in Example 1 With Exogenous X

Consider the setting of Example 1, but where in addition X is restricted to be exogenous. Here we show that the Gaussian random coefficients probit model is point identifying in this case.

The model stipulates that

$$Y = 0 \Leftrightarrow U_0 + U_1 X \leq -\alpha_0 - \alpha_1 X$$

and with $X \perp\!\!\!\perp U$:

$$(U_0 + U_1 X) \,|X = x \sim N(0, 1 + 2\gamma_0 x + \sigma x^2),$$

where $\sigma \equiv \gamma_0^2 + \gamma_1$ is the variance of U_1.

It follows that

$$P\left[Y = 0 | X = x\right] = \Phi\left(\frac{-\alpha_0 - \alpha_1 x}{(1 + 2\gamma_0 x + \sigma x^2)^{1/2}}\right)$$

and thus

$$g(x)\left(1 + 2\gamma_0 x + \sigma x^2\right)^{1/2} = -\alpha_0 - \alpha_1 x \tag{5.1}$$

where

$$g(x) \equiv \Phi^{-1}\left(P[Y = 0 | X = x]\right),$$

is point identified under Restriction A1.

The Gaussian random coefficients probit model with exogenous X is point identifying if there is a unique admissible solution for $\theta = (\alpha_0, \alpha_1, \gamma_0, \gamma_1)$ to the system of equations generated by (5.1) as x takes all values in the support of X. Admissible solutions are real-valued with $\gamma_1 + \gamma_0^2 \geq 0$.

In our numerical illustrations $\mathcal{X} = \{-1, 0, 1, 2\}$ and the parameter values employed are

$$(\alpha_0, \alpha_1, \gamma_0, \gamma_1) = (0, -1, -1, 1).$$

Thus $\sigma = 2$ and

$$g(-1) = -1/\sqrt{5}, \quad g(0) = 0, \quad g(1) = 1, \quad g(2) = 2/\sqrt{5}. \tag{5.2}$$

Setting $x = 0$ in (5.1) delivers

$$\alpha_0 = g(0) = 0.$$

Using this and (5.2) and setting $x = -1$ in (5.1) delivers

$$\alpha_1 = -\frac{1}{\sqrt{5}} \left(1 - 2\gamma_0 + \sigma\right)^{1/2}. \tag{5.3}$$

Setting $x = 1$ and then $x = 2$ in (5.1) gives the following pair of equations in (γ_0, σ):

$$\left(1 + 2\gamma_0 + \sigma\right)^{1/2} = -\alpha_1, \tag{5.4}$$

$$\frac{2}{\sqrt{5}} \left(1 + 4\gamma_0 + 4\sigma\right)^{1/2} = -2\alpha_1. \tag{5.5}$$

Solving (5.3), (5.4), and (5.5) we have the unique solution $(\alpha_1, \gamma_0, \sigma) = (-1, -1, 2)$, from which it follows that $\gamma_1 = 1$, and thus $\theta = (\alpha_0, \alpha_1, \gamma_0, \gamma_1)$ is point identified.

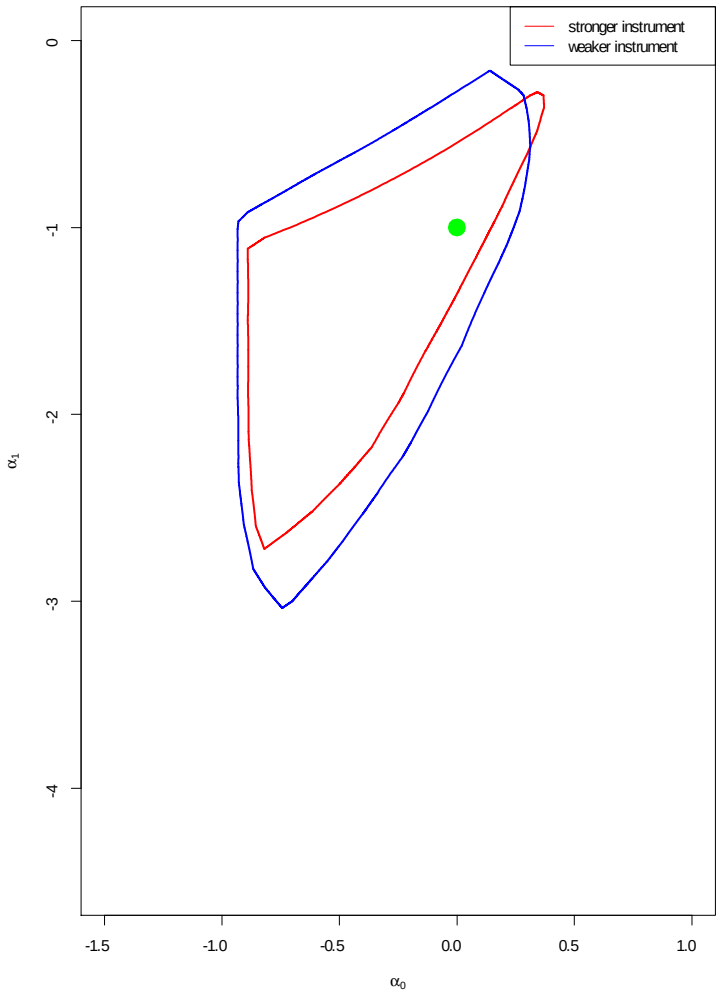

Figure 1: Projections of identified sets for (α_0, α_1) using weaker and stronger instruments for a process in which X is endogenous. The illustration is computed as described in Appendix B, with `alphahull` parameter set to 5.

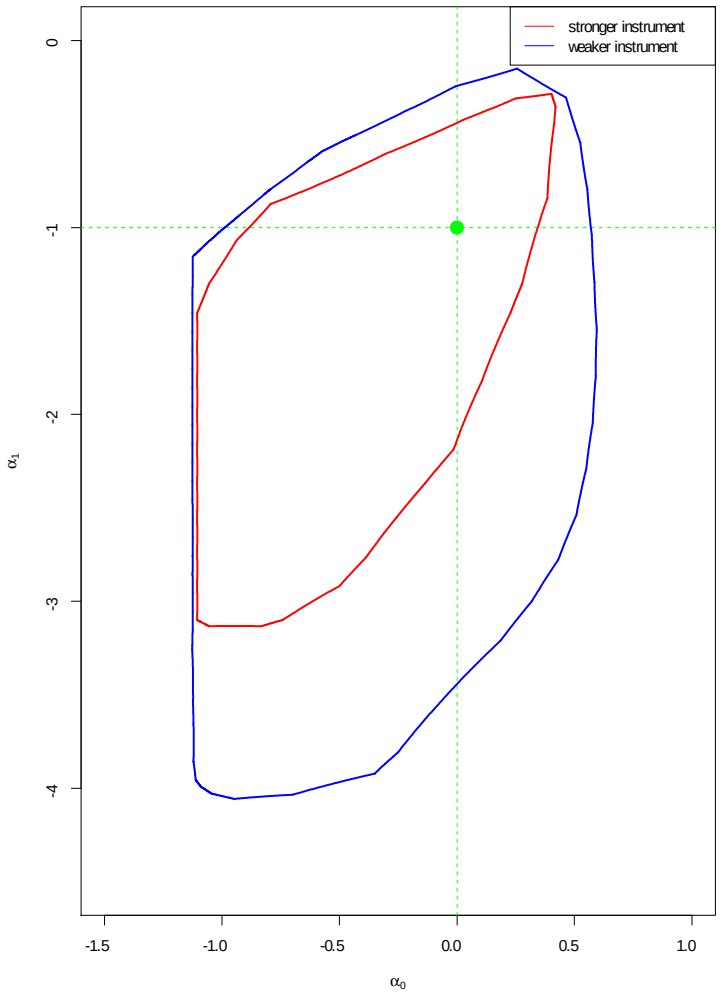

Figure 2: Projections of identified sets for (α_0, α_1) using weaker and stronger instruments for a process in which X is exogenous. The illustration is computed as described in Appendix B, with `alphahull` parameter set to 5.