

*Econometrics Journal* (2014), volume 17, pp. S1–S19.  
doi: 10.1111/ectj.12018

## An instrumental variable random-coefficients model for binary outcomes

ANDREW CHESHER<sup>†,‡</sup> AND ADAM M. ROSEN<sup>†,‡</sup>

<sup>†</sup>*Centre for Microdata Methods and Practice, Institute for Fiscal Studies,  
7 Ridgmount Street, London WC1E 7AE, UK.*

<sup>‡</sup>*Department of Economics, University College London, Gower Street,  
London WC1E 6BT, UK.*

E-mail: [andrew.chesher@ucl.ac.uk](mailto:andrew.chesher@ucl.ac.uk), [adam.rosen@ucl.ac.uk](mailto:adam.rosen@ucl.ac.uk)

First version received: October 2012; final version accepted: August 2013

**Summary** In this paper, we study a random-coefficients model for a binary outcome. We allow for the possibility that some or even all of the explanatory variables 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, and we thus apply identification results from our previous studies of such models 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:** *Endogeneity, Incomplete models, Instrumental variables, Partial identification, Random coefficients, Random sets, Set identification.*

### 1. INTRODUCTION

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

$$Y = 1[\beta_0 + X\beta_1 + W\beta_2 > 0], \quad (1.1)$$

where  $\beta \equiv (\beta_0, \beta_1', \beta_2')$  are random coefficients. Although 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  $\beta$  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  $\beta$  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

Correction Note: This article was first published online on the 2nd of September 2013, under a subscription publication licence. The article has since been made OnlineOpen, and the copyright line and licence statement was therefore updated in June 2014.

coefficients  $\beta$ , denoted  $F_\beta$ , with the joint distribution of random vectors  $X$  and  $\beta$  left unrestricted. As a special case, we also allow for the possibility there are no exogenous regressors  $W$ .<sup>1</sup> 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  $F_\beta$ , we carry out our identification analysis along the lines of Chesher et al. (2013), hereafter CRS, and Chesher and Rosen (2013). Like CRS, we consider a single-equation model for a discrete outcome, but here we restrict the outcome to be binary. However, the model (1.1) used in this paper 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 (2013), where we provide identification analysis for a broad class of instrumental variable (IV) models. Like those models, the random-coefficients model (1.1) allows for multiple sources of unobserved heterogeneity whereas, traditionally, IV methods have been employed in models admitting a single source of unobserved heterogeneity. Thus, in this paper, we investigate, and illustrate by way of example, the identifying power of IV restrictions with multivariate unobserved heterogeneity in the determination of a binary outcome. The characterizations we employ rely on results from random set theory. These and related results have been used for identification analysis in various ways and in a variety of contexts by Beresteanu et al. (2011, 2012), Galichon and Henry (2011), CRS, and Chesher and Rosen (2012, 2013). As in CRS and Chesher and Rosen (2012, 2013), 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 IV models for binary outcomes considered in Chesher (2010, 2013), 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. 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 might achieve greater identifying power.

Binary response specifications that model  $\beta$  in (1.1) as a random vector include, for example, 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  $\beta$  are independently distributed, and where the distribution of  $\beta$  is parametrically specified, enabling estimation via maximum likelihood. Ichimura and Thompson (1998) and Gautier and Kitamura (2013) focus on the binary outcome model (1.1), again with covariates and random coefficients independently distributed, but with  $F_\beta$  non-parametrically specified. Ichimura and Thompson (1998) provide sufficient conditions for point identification of  $F_\beta$  in this case, and prove that  $F_\beta$  can be consistently estimated via non-parametric maximum likelihood. Gautier and Kitamura (2013) introduce a computationally simple estimator for the density of  $\beta$ , and derive its rate of convergence and pointwise asymptotic normality. 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

<sup>1</sup> Similarly, the random intercept  $\beta_0$  can be easily removed from the analysis by restricting  $\beta_0 = 0$  throughout.

is assumed for the determination of  $X$  as a function of  $Z$ . Hoderlein and Sherman (2011) study identification and estimation of a trimmed mean of random coefficient  $\beta$  when again endogenous variables can be written as a function of mutually independent instruments  $Z$  and control variables  $V$ , additionally employing some conditional median restrictions. However, our model does not require one to specify the form of the stochastic relation between  $X$  and  $Z$ , and is thus incomplete.<sup>2</sup>

The random-coefficients logit model of Berry et al. (1995), hereafter BLP, now a bedrock of the empirical industrial organization literature, allows for endogeneity of prices using insight from Berry (1994) to handle endogeneity. Yet, the endogeneity problem in that and related models in industrial organization 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.<sup>3</sup> Recent papers that give identification results for micro-level discrete choice models with exogenous covariates and high-dimensional unobserved heterogeneity include Briesch et al. (2010), Bajari et al. (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.

The paper is organized as follows. 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 numerical illustrations of identified sets for subsets of parameters in a parametric version of our model for four different data-generation processes. We conclude in Section 5. The proof of the main identification result, which adapts theorems from CRS, is provided in Appendix 1. 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.

Throughout the paper, we use the following notation. We use upper-case Roman letters  $A$  to denote random variables and lower-case letters  $a$  to denote particular realizations. For the probability measure  $\mathbb{P}$ ,  $\mathbb{P}(\cdot|a)$  is used to denote the conditional probability measure given  $A = a$ . The 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-coefficients vector  $\beta$ , and  $\mathcal{S}$  denotes a closed set on  $\mathcal{B}$ . For any pair of random vectors  $A_1, A_2$ ,  $A_1 \perp A_2$  denotes stochastic independence,  $\text{Supp}(A_1, A_2, \dots, A_n)$  denotes the joint support of the collection of random vectors  $A_1, A_2, \dots, A_n$ , and  $\text{Supp}(A_1, A_2, \dots, A_n|b_1, \dots, b_m)$  denotes the conditional support

<sup>2</sup> The model is incomplete because there is no specification for the determination of  $X$  given exogenous variables  $Z$  and unobserved heterogeneity  $\beta$ . Thus, for any realization of  $(Z, \beta)$ , each  $x$  on the support of  $X$  is a feasible realization of  $X$ . However, the triangular structure used in the control function approach implies a unique value of  $X$  for any realization of exogenous variables and unobservables.

<sup>3</sup> In a binary choice model, the presence of unobserved, additively separable, alternative-specific utility shifters can be subsumed into the threshold-crossing specification, and so is unnecessary.

of  $(A_1, A_2, \dots, A_n)$  given realizations  $(b_1, \dots, b_m)$  of random vectors  $(B_1, \dots, B_m)$ . The empty set is denoted by  $\emptyset$ . We use  $F_\beta$  to denote the probability distribution of  $\beta$ , mapping from subsets of  $\mathcal{B}$  to the unit interval.  $\mathcal{F}$  is used to denote the admissible parameter space for  $F_\beta$ ,  $F$  is used to denote a generic element of  $\mathcal{F}$ , and  $\mathcal{F}^*$  denotes the identified set for  $F_\beta$ . We use  $\text{cl}(\mathcal{A})$  to denote the closure of a set  $\mathcal{A}$ . Finally,  $\tilde{Z} \equiv (W, Z)$  with support denoted  $\text{Supp}(\tilde{Z})$  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 2.1.**  $Y \in \{0, 1\}$ ,  $X \in \mathcal{X} \subseteq \mathbb{R}^{k_x}$ , 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  $Z \in \mathcal{Z} \subseteq \mathbb{R}^{k_z}$ .  $(\beta, W, X, Y, Z)$  belong to a probability space  $(\Omega, \mathfrak{S}, \mathbb{P})$  endowed with the Borel sets on  $\Omega$  and the joint distribution of  $(X, W, Y, Z)$ , denoted  $F_{XWYZ}^0$ , is identified. For all  $(x, w, z) \in \text{Supp}(X, W, Z)$ ,  $0 < \mathbb{P}[Y = 1|x, w, z] < 1$ .

**RESTRICTION 2.2.** For any  $(w, x, z)$  on the support of  $(W, X, Z)$ , the conditional distribution of random vector  $\beta$  given  $W = w$ ,  $X = x$ , and  $Z = z$  is absolutely continuous with respect to Lebesgue measure on  $\mathcal{B}$ .  $\beta$  is marginally distributed according to the probability measure  $F_\beta$  mapping from subsets of  $\mathcal{B}$  to the unit interval, with associated density  $f_\beta$ .  $F_\beta$  is known to belong to some class of probability measures  $\mathcal{F}$ .<sup>4</sup>

**RESTRICTION 2.3.**  $(W, Z)$  and  $\beta$  are independently distributed.

Restriction 2.1 invokes the random-coefficients 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 2.3 is our IV restriction, requiring independence of  $(W, Z)$  and  $\beta$ . Restriction 2.2 restricts  $F_\beta$  to some known class of distribution functions. In principle, this class could be parametrically, semi-parametrically, or non-parametrically specified. Of course, greater identifying power will be afforded when  $\mathcal{F}$  is parametrically specified. In our numerical illustrations in Section 4,  $\beta$  is restricted to be normally distributed, which is a common restriction in random-coefficients models.

As is always the case in models of binary response, it will be prudent to impose a scale normalization because  $\tilde{x}\beta > 0$  holds if and only if  $c \cdot \tilde{x}\beta > 0$  for all scalars  $c > 0$ , where  $\tilde{x} \equiv (1, x, w)$ .<sup>5</sup> This can be done by imposing, for example, that  $\mathcal{B} = \{b \in \mathbb{R}^k : \|b\| = 1\}$  if  $\mathcal{F}$  is non-parametrically specified, or by imposing that the first component of  $\beta$  has unit variance (e.g., when  $\mathcal{F}$  is parametrically specified as in the following example, and as also employed in the numerical illustrations of Section 4).

<sup>4</sup> If  $\mathcal{B}$  is bounded, then the absolute continuity condition should be understood to be required to hold with respect to the uniform measure on  $\mathcal{B}$ .

<sup>5</sup> Such normalizations are not strictly required when allowing for set identification, but it is wise to impose them in order to enable comparison of set and point identifying models.

EXAMPLE 2.1 (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

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

with  $\beta = (\beta_0, \beta_1)'$ . Suppose that  $\mathcal{F}$  is the class of bivariate normal distributions whose first component has unit variance. Then, defining  $\alpha_0, \alpha_1$  as the means of  $\beta_0, \beta_1$ , respectively, we have the representation

$$Y = 1[U_0 + U_1 X > -\alpha_0 - \alpha_1 X], \tag{2.1}$$

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 2.3 that  $U \perp\!\!\!\perp Z$ , and we can parametrize the distribution  $U \equiv (U_0, U_1)$  as

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

equivalently

$$U \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \gamma_0 \\ \gamma_0 & \gamma_1 + \gamma_0^2 \end{pmatrix}\right). \tag{2.2}$$

Knowledge of the parameter vector  $(\alpha_0, \alpha_1, \gamma_0, \gamma_1)$  would then suffice for the determination of  $F_\beta$ , so the identified set for  $F_\beta$  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

$$\mathcal{T}(w, x, y) \equiv \text{cl}\{(b_0, b'_1, b'_2)' \in \mathcal{B} : y = 1[b_0 + xb_1 + wb_2 > 0]\}, \tag{3.1}$$

which is the closure of the halfspace of  $\mathcal{B}$  on which  $2y - 1$  and  $b_0 + xb_1 + wb_2$  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  $F_{WXYZ}^0$  under Restriction 2.1. The identified set for  $F_\beta$ , denoted  $\mathcal{F}^*$ , is then the set of measures  $F \in \mathcal{F}$  that are selectable from the conditional distribution of  $\mathcal{T}(W, X, Y)$  given  $\tilde{Z} = \tilde{z}$  for almost every  $\tilde{z} \in \tilde{\mathcal{Z}}$ .<sup>6</sup> Intuitively, this holds because selectionability guarantees the existence of a random variable  $\tilde{\beta}$  realized on  $(\Omega, \mathfrak{S}, \mathbb{P})$  and distributed  $F$ , such that  $\mathbb{P}(\tilde{\beta} \in \mathcal{T}(W, X, Y) | \tilde{z}) = 1$ , a.e.  $\tilde{z} \in \tilde{\mathcal{Z}}$ .<sup>7</sup> Thus, there exists a random variable  $\tilde{\beta}$  distributed  $F$  that delivers the conditional distribution

$$F_{XWY|\tilde{Z}}^0(\cdot | \tilde{z}), \text{ a.e. } \tilde{z} \in \tilde{\mathcal{Z}},$$

and all such  $F$  are observationally equivalent.

<sup>6</sup> For further details and results regarding selections of random sets, see, e.g., Section 1.2 of Molchanov (2005).

<sup>7</sup> The requirement that  $\tilde{\beta}$  lives on  $(\Omega, \mathfrak{S}, \mathbb{P})$  is innocuous. If this were not the case, then one could simply redefine the initial probability space as the product of  $(\Omega, \mathfrak{S}, \mathbb{P})$  and the space on which  $\tilde{\beta}$  lives.

As done in CRS for utility-maximizing discrete choice models without random coefficients and in Chesher and Rosen (2013) for single-equation IV models more generally, we can exploit Artstein’s Inequality (Artstein, 1983, see also Norberg, 1992, and Molchanov, 2005, Section 1.4.8.) to characterize the identified set through the use of conditional containment functional inequalities. Using the same steps taken in Theorem 1 of CRS, Artstein’s Inequality guarantees that 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  $\mathcal{S} \subseteq \mathcal{B}$ ,

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

The use of the conditional containment inequality (3.2) reduces the problem of determining which  $F$  are selectionable from  $\mathcal{T}(W, X, Y)$  to a collection of conditional moment inequalities. In CRS and Chesher and Rosen (2012, 2013), 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  $(W, Z)$ . For any such realization  $(w, z)$ , the support of  $\mathcal{T}(W, X, Y)$  is the collection of sets

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

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  $\mathbb{T}(w, z)$  does not vary with  $(w, z)$ . The larger the conditional support  $\text{Supp}(X|w, z)$ , the larger the core-determining collection of test sets will be.

Given any  $(w, z)$ , each element of  $\mathbb{T}(w, z)$  is a halfspace in  $\mathcal{B}$ , so the required test sets  $\mathcal{S}$  take the form of unions of such halfspaces.<sup>8</sup> Alternatively, each such test set can be written as the complement of intersections of sets, each of which are complements of elements of  $\mathbb{T}(w, z)$ . This is convenient because the complement of each  $\mathcal{T} \in \mathbb{T}(w, z)$ , denoted  $\mathcal{T}^c$ , is also a halfspace, and the intersection of halfspaces is a convex polytope. Thus, the collection of core-determining test sets  $\mathcal{S}$  contains sets that are complements of intersections of halfspaces, equivalently complements of convex polytopes. The formal result follows.

**THEOREM 3.1.** *Let Restrictions 2.1–2.3 hold. Then, the identified set for  $F_\beta$  is*

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

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

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

<sup>8</sup> If  $\text{Supp}(X|w, z)$  is uncountable, then the collection of all such test sets will include uncountable unions. If  $\text{Supp}(X|w, z)$  is discrete and finite, then there are only finitely many possible unions of sets  $\mathcal{T} \in \mathbb{T}(w, z)$ , and only finite unions are required.

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

$$T^c(w, z) \equiv \{T^c(w, x, y) : y \in \{0, 1\} \wedge x \in \text{Supp}(X|w, z)\},$$

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

The theorem follows from consideration of Theorems 1 and 2 of CRS, adapted to the random set  $T(W, X, Y)$  defined in (3.1), which make use of Artstein's Inequality (Artstein, 1983) to prove sharpness; see also Norberg (1992) and Molchanov (2005), Section 1.4.8. The characterization of test sets for the containment functional characterization (3.4) of Theorem 2 in CRS stipulates that a core-determining collection of test sets  $\mathcal{S}$  is given by those that are (i) unions of elements of  $T(w, z)$ , and (ii) such that the union of the interiors of component sets is a connected set. In this paper, condition (ii) can be ignored because the sets  $T(w, x, y)$  and  $T(w', x', y')$  are all halfspaces through the origin, ensuring that  $T(w, x, y) \cap T(w', x', y')$  has open interior except in the special case  $(x, w) = (x', w')$  and  $y' = 1 - y$ , in which case  $T(w, x, y) \cup T(w', x', y') = \mathcal{B}$ . The test set  $\mathcal{B}$  can indeed be safely discarded from consideration because from  $F(\mathcal{B}) = 1$ , (3.4) is trivially satisfied. The equivalence of the containment functional characterization (3.4) and the capacity functional characterization (3.5) follow from the fact that, for any sets  $\mathcal{T}, \mathcal{S}$ , the events  $\mathcal{T} \subseteq \mathcal{S}$  and  $\mathcal{T} \cap \mathcal{S}^c = \emptyset$  are identical.

Theorem 3.1 provides a characterization of the identified set of distributions of random coefficients for binary choice models with endogeneity and instrumental variables. In particular, the representation is given by a collection of conditional moment inequalities, with one such inequality conditional on the realization of exogenous variables  $(w, z)$  for each element of  $T^\cup(w, z)$  in (3.4), equivalently one conditional moment inequality for each element of  $T^\cap(w, z)$  in (3.5). These conditional moment inequalities can then be used as a basis for estimation and inference. To illustrate, suppose that the endogenous variable  $X$  is discrete, so that for any  $(w, z)$ ,  $T(w, z)$  is a finite collection of sets in  $\mathcal{B}$ . We can therefore enumerate the elements of  $T^\cup(w, z)$  as  $\mathcal{S}_1, \dots, \mathcal{S}_J$  for some  $J < \infty$ . Suppose further that  $\mathcal{F}^*$  is parametrically specified up to finite-dimensional parameter  $\theta$ , with typical element  $F(\cdot|\theta) \in \mathcal{F}^*$ . The characterization of the identified set in (3.4) can then be written as those  $F(\cdot|\theta) \in \mathcal{F}^*$  such that

$$\forall j = 1, \dots, J, \quad E[m_j(X, W, Z; \theta) | w, z] \geq 0, \text{ a.e. } (W, Z),$$

where

$$m_j(X, W, Z; \theta) \equiv F(\mathcal{S}_j) - 1[T(W, X, Y) \subseteq \mathcal{S}_j].$$

Inference can then be based on these conditional moment inequalities using, for example, methods from Andrews and Shi (2013) or Chernozhukov et al. (2013).

In some important special cases, considered in the following examples, characterization of the identified set can be further simplified.

**EXAMPLE 3.1. (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

$$T(w, z) = \{b \in \mathcal{B} : b_0 + wb_2 \geq 0\}, \{b \in \mathcal{B} : b_0 + wb_2 \leq 0\},$$

where  $b$  is of the form  $b = (b_0, b_2)'$ . The intersection of these sets is  $\{b \in \mathcal{B} : b_0 + wb_2 = 0\}$ , which has zero measure  $F_\beta$  under Restriction 2.2, and their union is  $\mathcal{B}$ , which has measure 1. It

follows from similar reasoning as in Theorem 6 of Chesher and Rosen (2012) that for any  $(w, z)$  the inequalities of the characterizations of Theorem 3.1 produce moment equalities. Consider, for example, the containment functional inequalities of (3.4) delivered by all  $\mathcal{S} \in \mathcal{T}^\cup(w, z)$ :

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

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 1, implying that these inequalities must, in fact, hold with equality, giving

$$F(\{b \in \mathcal{B} : b_0 + wb_2 \geq 0\}) = \mathbb{P}[Y = 1 | w, z], \quad (3.6)$$

$$F(\{b \in \mathcal{B} : b_0 + wb_2 \leq 0\}) = \mathbb{P}[Y = 0 | w, z]. \quad (3.7)$$

When there are no excluded exogenous variables  $z$  and  $F_\beta$  is not restricted to a parametric family, these equations coincide with the identifying equations in Ichimura and Thompson (1998) and Gautier and Kitamura (2013). Ichimura and Thompson (1998) provide sufficient conditions for point identification.<sup>9</sup> 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) and (3.7) a.e.  $(W, Z)$  nonetheless fully characterize the identified set.

**EXAMPLE 3.2. (ONE ENDOGENOUS COVARIATE WITH ARBITRARY EXOGENOUS COVARIATES).** Consider the common setting where there is a single endogenous explanatory variable,  $X \in \mathbb{R}$ , as well as some exogenous explanatory variables  $W$ , a random  $k_w$ -vector. Then, given any  $(w, z)$ , the collection of sets  $\mathcal{T}(w, z)$  is given by

$$\mathcal{T}(w, z) \equiv \{\{b \in \mathcal{B} : b_0 + xb_1 + wb_2 \geq 0\}, \{b \in \mathcal{B} : b_0 + xb_1 + wb_2 \leq 0\} : x \in \text{Supp}(X | w, z)\}.$$

Suppose, for simplicity, that  $\text{Supp}(X | w, z)$  is discrete. Consider now a test set  $\mathcal{S}$  which is one of the core-determining sets in  $\mathcal{T}^\cup(w, z)$  and hence an arbitrary union of sets in  $\mathcal{T}(w, z)$ .<sup>10</sup> Any such  $\mathcal{S}$  can be equivalently written as the set of  $b = (b_0, b_1, b_2)' \in \mathcal{B}$  that satisfy one of the inequalities

$$b_0 + wb_2 + \max_{x \in \mathcal{X}_1} \{xb_1\} \geq 0 \quad \text{or} \quad b_0 + wb_2 + \min_{x \in \mathcal{X}_0} \{xb_1\} \leq 0, \quad (3.8)$$

for some collections of values  $\mathcal{X}_0, \mathcal{X}_1 \subseteq \text{Supp}(X | w, z)$ .

<sup>9</sup> The restrictions used to ensure point identification include the requirements that for some fixed  $c \in \mathbb{R}^{k_w}$ ,  $F_\beta(\{b : c'b > 0\}) = 1$ , 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) and (3.7) in the case of only exogenous covariates, do not require these restrictions.

<sup>10</sup> The restriction to cases where  $\text{Supp}(X | w, z)$  is discrete is not essential but simplifies the exposition. An identical characterization of required test sets  $\mathcal{S}$  can be shown more generally by referring back to (3.2) appearing in (3.4) and making use of the absolute continuity of  $F_\beta$  from Restriction 2.2.



Define now for each  $j = 0, 1$ ,

$$x_j^{\max} \equiv \max_{x \in \mathcal{X}_j} x, \quad x_j^{\min} \equiv \min_{x \in \mathcal{X}_j} x.$$

If  $b_1 \geq 0$ , (3.8) simplifies to

$$b_0 + wb_2 + x_1^{\max} b_1 \geq 0 \quad \text{or} \quad b_0 + wb_2 + x_0^{\min} b_1 \leq 0, \quad (3.9)$$

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

$$b_0 + wb_2 + x_1^{\min} b_1 \geq 0 \quad \text{or} \quad b_0 + wb_2 + x_0^{\max} b_1 \leq 0. \quad (3.10)$$

Furthermore, for any  $b \in \mathcal{B}$  with  $b_1 \geq 0$ , (3.10) implies (3.9), and for any  $b \in \mathcal{B}$  with  $b_1 < 0$ , (3.9) implies (3.10). Thus, for any  $b \in \mathcal{B}$ , (3.8) holds if and only if

$$b \in \mathcal{T}(w, x_1^{\min}, 1) \cup \mathcal{T}(w, x_1^{\max}, 1) \cup \mathcal{T}(w, x_0^{\min}, 0) \cup \mathcal{T}(w, x_0^{\max}, 0).$$

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

$$\mathcal{S} = \mathcal{T}(w, x_1, 1) \cup \mathcal{T}(w, x_2, 1) \cup \mathcal{T}(w, x'_1, 0) \cup \mathcal{T}(w, x'_2, 0),$$

where  $x_2 \geq x_1$  and  $x'_2 \geq x'_1$ .

EXAMPLE 2.1. (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  $\mathbb{T}(w, z) = \mathbb{T}(z)$  for any  $z$  reduces to

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

Each element of  $\mathbb{T}(z)$  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  $\mathbb{T}^\cup(w, z) = \mathbb{T}^\cup(z)$  is given by the collection of test sets that can be written as either elements of  $\mathbb{T}(z)$  or unions of a pair of elements in  $\mathbb{T}(z)$ ,

$$\mathbb{T}^\cup(z) = \{\mathcal{T}(x_1, y_1) \cup \mathcal{T}(x_2, y_2) : x_1, x_2 \in \text{Supp}(X|z), y_1, y_2 \in \{0, 1\}\}, \quad (3.11)$$

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

$$\mathcal{T}(x, y) = \text{cl} \{(b_0, b_1)' \in \mathcal{B} : y = 1 [b_0 + xb_1 > 0]\}.$$

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  $\mathbb{T}^\cup(z)$  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 2.1, where there are no exogenous covariates  $W$  and where  $F$  is restricted to a parametric (specifically Gaussian) family. In the illustrations, we investigate identified sets for averages of  $(\beta_0, \beta_1)$ , 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  $\mathbb{T}(w, z)$  in order to check whether such candidate values belong to the identified set.

**Table 1.** Parameter settings in the four calculations.

Parameter	Endogenous X		Exogenous X	
	Case N1	Case N2	Case X1	Case X2
$\delta_1$	1.000	1.500	1.000	1.500
$\delta_2$	0.577	0.577	0.000	0.000
$\delta_3$	-0.577	-0.577	0.000	0.000
$\delta_4$	0.577	0.462	1.414	1.371

#### 4. NUMERICAL ILLUSTRATIONS

To investigate the identifying power of the binary outcome random-coefficients IV model, we reconsider Example 2.1.  $Y$  is determined as in (2.1) with unobservable  $U = (U_0, U_1)$  bivariate normal with zero mean and variance as parametrized in (2.2). We define

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

as the probability that  $U$  belongs to the set  $\mathcal{U}$  where  $\theta = (\alpha_0, \alpha_1, \gamma_0, \gamma_1)$  and when  $\beta$  is distributed  $F_\beta$  with mean  $\alpha$  and variance governed by parameters  $(\gamma_0, \gamma_1)$ . Given the restriction that  $\beta = (\beta_0, \beta_1)'$  is bivariate normally distributed, knowledge of  $\theta$  implies knowledge of  $F_\beta$ . Thus, we consider the identified set for  $\theta$ , denoted  $\Theta^*$ , and focus attention on the identified set for  $(\alpha_0, \alpha_1)$ , the projection of the first two elements of  $\Theta^*$  on  $\mathbb{R}^2$ .

##### 4.1. Data-generating processes

Our examples employ data-generating processes in which  $X$  is determined as follows:

$$X = x_k \text{ iff } c_{k-1} < \delta_1 Z + \delta_2 U_0 + \delta_3 U_1 + \delta_4 V \leq c_k, \quad k \in \{1, \dots, K\} \quad (4.1)$$

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

We report four calculations, in all of which there are the following settings:

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

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

In two cases (N1 and N2), the parameters are set such that  $X$  is endogenous, and in another two cases (X1 and X2), they are set such that  $X$  is exogenous. We consider two possibilities for the coefficient  $\delta_1$  multiplying instrument  $Z$  in the determination of  $X$  in (4.1):  $\delta_1 = 1$  (N1 and X1) and  $\delta_1 = 1.5$  (N2 and X2). All parameter settings are shown in Table 1. Table 2 shows the

**Table 2.** Conditional probabilities  $\mathbb{P}[X = x|z]$ .

		$z = -2$	$z = -1$	$z = +1$	$z = +2$
$\delta_1 = 1.0$	$x = -1$	0.760	0.500	0.079	0.017
	$x = 0$	0.161	0.260	0.161	0.062
	$x = 1$	0.062	0.161	0.260	0.161
	$x = 2$	0.017	0.079	0.500	0.760
$\delta_1 = 1.5$	$x = -1$	0.928	0.642	0.034	0.002
	$x = 0$	0.058	0.221	0.103	0.013
	$x = 1$	0.013	0.103	0.221	0.058
	$x = 2$	0.002	0.034	0.642	0.928

two conditional distributions of  $X$  given  $Z$ . In all cases, the support of the instrumental variable is  $\mathcal{Z} = \{-2, -1, 1, 2\}$ .

If the exogeneity restriction  $X \perp\!\!\!\perp \beta$  is imposed then, as shown in Appendix C, the resulting model point identifies the full parameter vector  $\theta$ . In the structures delivering probability distributions in cases X1 and X2, it is the case that  $X \perp\!\!\!\perp \beta$  holds. However, we calculate identified sets for a model without the exogeneity restriction and thereby show the substantial loss in identifying power arising when exogeneity cannot be assumed to hold.

#### 4.2. Calculation of probabilities

To illustrate identified sets, we computed the conditional probabilities  $\mathbb{P}[X = x_k|z]$  and  $\mathbb{P}[Y = 0 \wedge X = x_k|z]$ .  $\mathbb{P}[X = x_k|z]$  is given by

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

where  $\Phi(\cdot)$  denotes the standard normal distribution function and

$$\lambda \equiv \delta_2^2 + 2\delta_2\delta_3\gamma_0 + \delta_3^2(\gamma_1 + \gamma_0^2) + \delta_4^2. \tag{4.2}$$

The conditional probability  $\mathbb{P}[Y = 0 \wedge X = x_k|z]$  can be calculated as the difference between two normal orthant probabilities because, when  $Z = z$ , we have

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

where

$$\tilde{V} \equiv \delta_2 U_0 + \delta_3 U_1 + \delta_4 V, \quad Q_k \equiv U_0 + x_k U_1.$$

Because  $(U_0, U_1, V) \perp\!\!\!\perp Z$ , conditional on  $Z = z$ , we have

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

from which we see that  $\mathbb{P}[Y = 0 \wedge X = x_k|z]$  is indeed the difference between two normal orthant probabilities. The conditional probability  $\mathbb{P}[Y = 1 \wedge X = x_k|z]$  can then be obtained by subtracting  $\mathbb{P}[Y = 0 \wedge X = x_k|z]$  from  $\mathbb{P}[X = x_k|z]$ .<sup>11</sup>

### 4.3. Calculation of projections

We calculate two-dimensional projections of the four-dimensional (4D) identified set for  $\theta_0$ , giving results for the projection on to the plane on which lie  $(\alpha_0, \alpha_1)$ . This is the identified set for the mean of the random coefficients  $(\beta_0, \beta_1)$ .

We calculate the projections as follows.<sup>12</sup> The full 4D identified set is

$$\Theta^* = \{\theta \in \Theta : \forall S \in \mathcal{S}, G_U(\mathcal{S}, \theta) \geq \max_{z \in \mathcal{Z}} \mathbb{P}[\mathcal{T}(X, Y) \subseteq S|z]\}, \tag{4.3}$$

where  $\mathcal{S} = \mathcal{T}^\cup(z)$  is a collection of 32 core-determining sets of the form described for Example 2.1 in Section 3, specifically (3.11), 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  $\theta$ . The probabilities  $\mathbb{P}[\mathcal{T}(X, Y) \subseteq S|Z = z]$ ,  $z \in \mathcal{Z}$ , are identified under Restriction 2.1.

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

$$D(\theta) \equiv \max_{z \in \mathcal{Z}, S \in \mathcal{S}} (\mathbb{P}[\mathcal{T}(X, Y) \subseteq S|z] - G_U(\mathcal{S}, \theta)), \tag{4.4}$$

which can be used to characterize the full 4D identified set as follows:

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

To compute identified sets for subvectors of parameters, let  $\theta_c$  denote a list of one or more elements of  $\theta$ , and let  $\theta_{-c}$  denote the remaining elements of  $\theta$ . The projection of the identified set on to the space in which  $\theta_c$  resides is the set of values of  $\theta_c$  for which there exists  $\theta_{-c}$  such that  $\theta = (\theta_c, \theta_{-c})$  lies in the identified set  $\Theta^*$ . We calculate this set,  $\Theta_c^*$ , as the set of values  $\theta_c$  for which the value of  $\min_{\theta_{-c}} D(\theta_c, \theta_{-c})$  is non-positive:

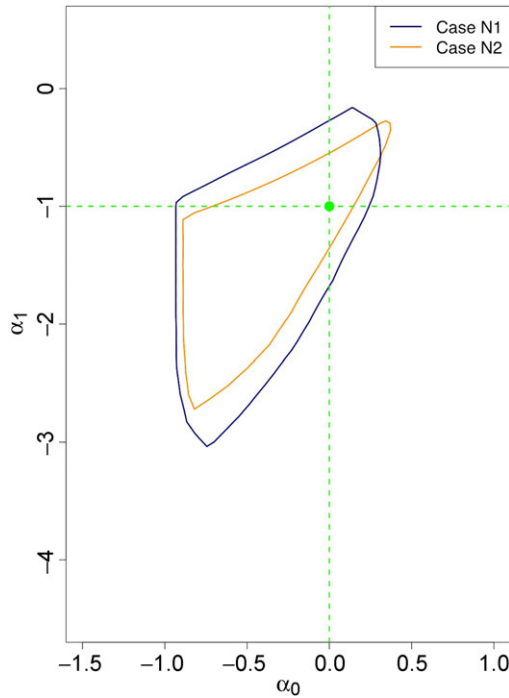
$$\Theta_c^* = \{\theta_c : \min_{\theta_{-c}} D(\theta_c, \theta_{-c}) \leq 0\}. \tag{4.5}$$

Here,  $D(\theta_c, \theta_{-c})$  is to be understood as the function defined in (4.4) applied to that value of  $\theta$  with subvectors equal to  $\theta_c$  and  $\theta_{-c}$ . We perform this minimization using the `optim` function in base R.

Figure 1 shows the projections of the identified set in cases N1 and N2 in which  $X$  is endogenously determined. The probability generating value  $(\alpha_0, \alpha_1) = (0, -1)$  is plotted. When the parameter  $\delta_1 = 1.5$  (drawn in beige, labelled Case N2), the area of the projection is smaller than when  $\delta_1 = 1.0$  (drawn in blue, labelled Case N1). Most values in the projection when  $\delta_1 = 1.5$  lie inside the projection obtained when  $\delta_1 = 1.0$ , but at high values of  $\alpha_0$  there is a very small region of the first projection that is not contained in the latter. Note that this can happen

<sup>11</sup> In our R programs (R Core Team, 2012), the bivariate normal orthant probabilities are calculated using the `pmvnorm` program provided in the `mvtnorm` package (Genz et al., 2012), which implements computation of multivariate normal and  $t$  probabilities from Genz and Bretz (2009).

<sup>12</sup> In this section, we write  $\mathcal{T}(X, Y)$  in place of  $\mathcal{T}(W, X, Y)$  because there are no exogenous covariates  $W$  in this model.



**Figure 1.** Projections of identified sets for cases N1 and N2.

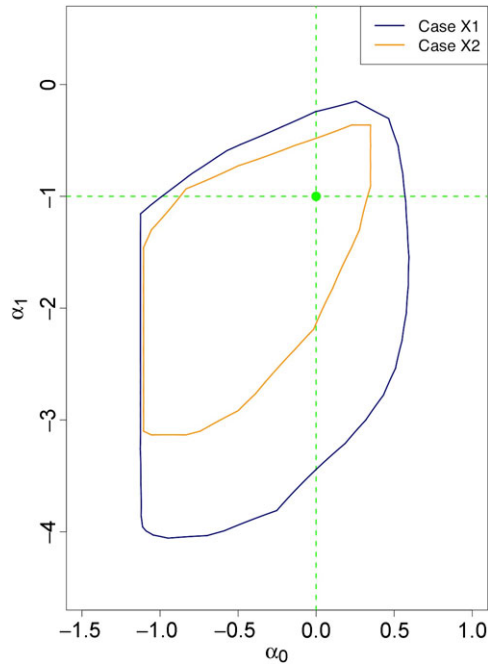
because even though the slope coefficient on  $Z$  in (4.1) is larger in the  $\delta_1 = 1.5$  case, this does not guarantee that the quantity  $\max_{z \in Z} \mathbb{P}[T(W, X, Y) \subset \mathcal{S}|z]$  providing the lower bound of the inequalities in (4.3) is larger than in the  $\delta_1 = 1.0$  case. Figure 2 similarly illustrates projections of the identified set for cases X1 and X2 in which  $X$  is exogenously determined in the probability generating process. In this case, the projection of the identified set when  $\delta_1 = 1.5$  is a subset of that when  $\delta_1 = 1.0$ . The identified sets are larger in the exogenous  $X$  cases, even though the predictive power of the instrument is the same as in the endogenous  $X$  cases. This occurs because the scale on which  $(\alpha_0, \alpha_1)$  is measured differs in the two cases.<sup>13</sup> Computations for both figures were implemented as described in Appendix B, with the `alphahull` parameter set to 5.

In all cases, the projections contain no positive values of  $\alpha_1$ , so the model allows one to sign  $\alpha_1$  and 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

<sup>13</sup> The scale difference arises because of the differential variability of the index  $U_0 + U_1 X$  in (2.1) as measured by the conditional variance given  $X$  and  $Z$ . Calculations using simulated values of the unobservables show that this is larger at every value of  $X$  and  $Z$  in the exogenous  $X$  case.



**Figure 2.** Projections of identified sets for cases X1 and X2.

restricted to be distributed independently of the random coefficients. We have shown that with an IV restriction we can apply analysis along the lines of that in CRS and Chesher and Rosen (2013) to characterize the identified set as those distributions that satisfy a collection of conditional moment inequalities. In our numerical illustrations of Section 4, there are 32 such 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 (2013) and Chernozhukov et al. (2013), are applicable. In some settings, the number of core-determining sets in the full characterization can be quite large, necessitating some care in choosing the number to employ in small samples. Issues that arise as a result of many moment inequalities have been investigated in an asymptotic paradigm by Menzel (2009). With discrete endogenous variables having finite support, the number of conditional moment inequalities can be large, but is necessarily finite, and future research on finite sample approximations for inference and computational issues is warranted.

We have provided numerical illustrations of identified sets under particular data-generation processes. We have given an overview of the computational approach we used for computing these identified sets, and details are set out in Appendix B.

Although our computational approaches are adequate for the examples considered, we have no doubt that they can be improved, either by developing more efficient implementations, or by devising new computational approaches altogether. Nonetheless, the illustrations serve to demonstrate the feasibility of computing identified sets in one particular setting in the general class of IV models studied in Chesher and Rosen (2013). These IV models can admit

high-dimensional unobserved heterogeneity, for example through a random-coefficients specification such as the one studied in this paper.

## ACKNOWLEDGEMENTS

We thank participants at the June 2011 joint Northwestern Econometrics and Centre for Microdata Methods and Practice (CeMMAP) conference in honour of Joel Horowitz for comments, and we are grateful to Konrad Smolinski for assistance with computations in the initial stages of this work. We gratefully acknowledge financial support from the UK Economic and Social Research Council (ESRC) through a grant (RES-589-28-0001) to the 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.

## REFERENCES

- Andrews, D. W. K. and X. Shi (2013). Inference based on conditional moment inequalities. *Econometrica* 81, 609–66.
- Artstein, Z. (1983). Distributions of random sets and random selections. *Israel Journal of Mathematics* 46, 313–24.
- Bajari, P., J. Fox, K-i. Kim and S. P. Ryan (2012). The random coefficients logit model is identified. *Journal of Econometrics* 166, 204–12.
- Beresteanu, A., I. Molchanov and F. Molinari (2011). Sharp identification regions in models with convex moment predictions. *Econometrica* 79, 1785–821.
- Beresteanu, A., I. Molchanov and F. Molinari (2012). Partial identification using random set theory. *Journal of Econometrics* 166, 17–32.
- Berry, S. T. (1994). Estimating discrete choice models of product differentiation. *Rand Journal of Economics* 25, 242–62.
- Berry, S. and P. Haile (2009). Nonparametric identification of multiple choice demand models with heterogeneous consumers. NBER Working Paper w15276, National Bureau of Economic Research.
- Berry, S. and P. Haile (2010). Identification in differentiated markets using market level data. NBER Working Paper w15641, National Bureau of Economic Research.
- Berry, S., J. Levinsohn and A. Pakes (1995). Automobile prices in market equilibrium. *Econometrica* 63, 841–90.
- 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, 291–307.
- Chernozhukov, V., S. Lee and A. Rosen (2013). Intersection bounds: estimation and inference. *Econometrica* 81, 667–737.
- Chesher, A. (2010). Instrumental variable models for discrete outcomes. *Econometrica* 78, 575–601.
- Chesher, A. (2013). Semiparametric structural models of binary response: shape restrictions and partial identification. *Econometric Theory* 29, 231–66.
- Chesher, A. and A. Rosen (2012). Simultaneous equations models for discrete outcomes: coherence, completeness, and identification. CWP21/12, Centre for Microdata Methods and Practice, Institute for Fiscal Studies and University College London.

- Chesher, A. and A. Rosen (2013). Generalized instrumental variable models. CWP43/13, Centre for Microdata Methods and Practice, Institute for Fiscal Studies and University College London.
- Chesher, A., A. Rosen and K. Smolinski (2013). An instrumental variable model of multiple discrete choice. *Quantitative Economics* 4, 157–96.
- 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, 1264–98.
- Gautier, E. and Y. Kitamura (2013). Nonparametric estimation in random coefficients binary choice models. *Econometrica* 81, 581–607.
- Gautier, E. and E. LePennec (2011). Adaptive estimation in the nonparametric random coefficients binary choice model by needlet thresholding. Working Paper, CREST/ENSAE, Paris.
- Genz, A. and F. Bretz (2009). *Computation of Multivariate Normal and t Probabilities*. Lecture Notes in Statistics, Volume 195. Heidelberg: Springer.
- 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, 403–26.
- Hoderlein, S. (2009). Endogenous semiparametric binary choice models with heteroscedasticity. CWP34/09, Centre for Microdata Methods and Practice, Institute for Fiscal Studies and University College London.
- Hoderlein, S. and R. Sherman (2011). Identification and estimation in a correlated random coefficients binary response model. CWP42/12, Centre for Microdata Methods and Practice, Institute for Fiscal Studies and University College London.
- 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, 269–95.
- 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 frequencies to approximate choice probabilities. In C. F. Manski and D. L. McFadden (Eds.), *Structural Analysis of Discrete Data and Econometric Applications*, 305–19. Cambridge, MA: MIT Press.
- McFadden, D. L. (1976). Quantal choice analysis: a survey. *Annals of Economic and Social Measurement* 5, 363–90.
- Menzel, K. (2009). Estimation and inference with many weak moment inequalities. Working Paper, MIT.
- Molchanov, I. S. (2005). *Theory of Random Sets*. London: Springer.
- 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, 241–64.
- 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, 507–36.
- R Core Team (2012). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.



## APPENDIX A: PROOF OF THEOREM 3.1

**Proof of Theorem 3.1:** Following the same steps as in the proof of Theorem 1 of CRS 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, we obtain

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

where  $\mathcal{F}(\mathcal{B})$  denotes all closed subsets of  $\mathcal{B}$ . Then, the application of Theorem 2 of CRS, specifically part (i), further gives that  $\mathcal{F}(\mathcal{B})$  above can 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 (2012), it follows that when considering probabilities conditional on  $(W, Z) = (w, z)$ ,  $\mathcal{F}(\mathcal{B})$  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  $\mathbb{T}^{\cup}(w, z)$ . The representation

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

follows from the equivalence

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

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

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

## APPENDIX B: COMPUTATIONAL DETAILS

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

### B.1. Calculation of probabilities $G_U(\mathcal{S}, \theta)$

Each set  $\mathcal{S}$  in the collection  $\mathbb{T}^{\cup}(z) = \mathbb{T}^{\cup}$  is the union of one or more contiguous cones centred at the point  $(\alpha_0, \alpha_1)$ , 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 eight 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 eight cones by a bivariate normal density function with mean  $(0, 0)$  and variance matrix entirely determined by  $(\gamma_0, \gamma_1)$ . The probability mass supported by a particular set  $\mathcal{S}$  at the value of  $\theta$  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 eight 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 carried out by 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.

### B.2. Calculation of projections

First approximations to the  $(\alpha_0, \alpha_1)$ -projections of identified sets were obtained by evaluating over a coarse grid of values of  $(\alpha_0, \alpha_1)$ . 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  $(\alpha_0, \alpha_1)$  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. 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  $\alpha_0$  and  $\alpha_1$  axes to refine the boundary approximations in areas where it was relatively non-linear. 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 (4.5) when determining membership of the identified set is not very well behaved. There are points at which it is not differentiable and there appear to be some places in which there are 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 the inaccuracy in calculating these affects the computation of the minimum in (4.5). The effect is likely to be dependent on the parameter value  $(\alpha_0, \alpha_1)$  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  $\theta_{-c}$  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 (4.5) 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.

### B.3. Graphics

The projections 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 2.1 WITH EXOGENOUS $X$

Consider the setting of Example 2.1, 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

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

and thus

$$g(x) (1 + 2\gamma_0 x + \sigma x^2)^{1/2} = -\alpha_0 - \alpha_1 x, \tag{C.1}$$

where

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

is point identified under Restriction 2.1.

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 (C.1) when  $x$  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

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

Setting  $x = 0$  in (C.1) delivers

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

Using this and (C.2) and setting  $x = -1$  in (C.1) delivers

$$\alpha_1 = -\frac{1}{\sqrt{5}} (1 - 2\gamma_0 + \sigma)^{1/2}. \tag{C.3}$$

Setting  $x = 1$  and then  $x = 2$  in (C.1) gives the following pair of equations in  $(\gamma_0, \sigma)$ :

$$(1 + 2\gamma_0 + \sigma)^{1/2} = -\alpha_1; \tag{C.4}$$

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

Solving (C.3)–(C.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.