cemmap
centre for microdata methods and practice

An instrumental variable model of multiple discrete choice

Andrew Chesher
Adam Rosen
Konrad Smolinski

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP39/11

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

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

An Instrumental Variable Model of Multiple Discrete Choice*

Andrew Chesher[†] UCL and CeMMAP

Adam M. Rosen[‡] UCL and CeMMAP

Konrad Smolinski[§] CeMMAP and IFS

December 15, 2011

Abstract

This paper studies identification of latent utility functions in multiple discrete choice models in which there may be endogenous explanatory variables, that is explanatory variables that are not restricted to be distributed independently of the unobserved determinants of latent utilities. The model does not employ large support, special regressor or control function restrictions, indeed it is silent about the process delivering values of endogenous explanatory variables and in this respect it is incomplete. Instead the model employs instrumental variable restrictions requiring the existence of instrumental variables which are excluded from latent utilities and distributed independently of the unobserved components of utilities.

We show that the model delivers set identification of the latent utility functions and we characterize sharp bounds on those functions. We develop easy-to-compute outer regions which

*This paper is a revised version of the February 2011 CeMMAP working paper CWP06/11. We thank seminar participants at Brunel University, CeMMAP, CREST, Harvard/MIT, The Institute for Advanced Studies (Vienna), UC Berkeley, UCLA, USC, and the University of Manchester, as well as audiences at the September 2010 CIREQ conference on revealed preferences and partial identification, the 21^{st} EC^2 conference held in Toulouse in December 2010, and the December 2011 workshop on consumer behavior and welfare measurement held at the IFS in London for comments and discussion. We especially thank Francesca Molinari for very helpful and detailed discussion on our use of random set theory. Financial support from the Economic and Social Research Council through the ESRC Centre for Microdata Methods and Practice grant RES-589-28-0001, 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.

§Address: Konrad Smolinski, Institute for Fiscal Studies, 7 Ridgmount Street, London WC1E 7AE, konrad_s@ifs.org.uk.

in parametric models require little more calculation than what is involved in a conventional maximum likelihood analysis. The results are illustrated using a model which is essentially the parametric conditional logit model of McFadden (1974) but with potentially endogenous explanatory variables and instrumental variable restrictions.

The method employed has wide applicability and for the first time brings instrumental variable methods to bear on structural models in which there are multiple unobservables in a structural equation.

Keywords: Partial identification, random sets, multiple discrete choice, endogeneity, instrumental variables, incomplete models.

1 Introduction

This paper develops results on the identification of features of models of choice amongst multiple, discrete, unordered alternatives. The model we employ allows for the possibility that explanatory variables are endogenous.

Our model uses the random utility maximizing framework set down in the groundbreaking work of McFadden (1974). Individuals choose one of $y = 1, \ldots, M$ alternatives, achieving utility $U_y = u_y(X, V_y)$ if choice y is made. Individuals observe the utility achieved from all choices and select the alternative delivering maximum utility. The econometrician observes the choice made, a realization of a discrete random variable Y, and the explanatory variables, X. There is interest in the functions $u \equiv (u_1, \ldots, u_M)$ and the distribution of $V \equiv (V_1, \ldots, V_M)$ and functionals of these features.

In the setup considered by McFadden the explanatory variables X and unobservable utility shifters V are independently distributed. Our model relaxes this restriction, permitting components of X to be endogenous. For example in a travel demand context one of the explanatory variables might be distance to work. This could be endogenous if individuals choose where to live based in part on unobserved tastes for varieties of transport, for instance because they dislike driving through rush-hour traffic and prefer public transit. We bring a classical instrumental variable (IV) restriction on board, requiring that there exist observed variables Z such that Z and V are independently distributed. Components of Z may either correspond to components of X thought to be exogenous, or may be excluded from the utility functions $u_1, \ldots, u_M$. In the travel demand setting excluded components of Z may be variables that influence choice of residential location but have no other role in determining propensities to travel by alternative transport modes. We show that this model is set identifying and we characterize the identified set of utility functions and distributions of unobservable utility shifters.

In McFadden (1974) the distribution of V is fully specified. The elements of V are independently and identically distributed Type 1 extreme value variates leading to the conditional logit model. Since that seminal contribution there have been many less restrictive, parametric specifications, as in for example the conditional probit model of Hausman and Wise (1978) which gives V a multivariate normal distribution, and the nested logit model of Domencich and McFadden (1975)[1] in which V has a Generalized Extreme Value distribution. Our results apply in all these cases and our development is quite general, delivering characterizations of the identified set even in the absence of parametric restrictions. In some illustrative calculations we work with McFadden's specification which produces a conditional logit model when the explanatory variables are restricted to be exogenous.

A novel feature of our results is that they demonstrate that instrumental variable models can have identifying power in cases in which there are multiple unobservables appearing in structural functions. Hitherto IV models have required unobservables to be scalar - see for example Newey and Powell (2003) Chernozhukov and Hansen (2005), and Chesher (2010). A general approach to identification in models with multiple unobservables is set out in Chesher, Rosen, and Smolinski (2011).

The IV model studied here is unrestrictive relative to many other models of multiple discrete choice permitting endogeneity that have been used till now. In our IV model there is no restriction placed on the process generating the potentially endogenous explanatory variables. In this sense the model is incomplete. Because of this incompleteness the model is generally not point identifying. The model does not employ large support conditions or special regressors and there need not be alternative-specific covariates. Explanatory variables and instrumental variables can be continuous or discrete. Because our model's restrictions are weak the model can be credibly applied in a wide variety of situations.

Here is a brief outline of the main results of the paper.

1.1 The main results

The set of utility functions and distributions of latent variables identified by our IV multiple discrete choice model is characterized by a system of inequalities which it is convenient to express in terms of a conditional containment functional associated with a set-valued

[1] See also Ben-Akiva (1973) and McFadden (1978).

random variable, or random set, $\mathcal{T}_v(Y, X; u)$. A realization of one of these random sets, $\mathcal{T}_v(y, x; u)$, is the set of values of unobserved utility shifters, $V = (V_1, ..., V_M)$ that leads to a particular realization y of the choice variable Y when the explanatory variables X take the value x and the utility functions u govern choices. The conditional containment functional $\Pr[\mathcal{T}_v(Y, X; u) \subseteq \mathcal{S}|z]$ gives the probability conditional on instrumental variable $Z = z$ that $\mathcal{T}_v(Y, X; u)$ is a subset of the set $\mathcal{S}$.

We show that a utility function u and a distribution P_V of unobservable utility shifters lies in the identified set associated with conditional distributions of Y and X given Z, $F^0_{YX|Z}$, if and only if

$$P_V(\mathcal{S}) \geq \Pr_0[\mathcal{T}_v(Y, X; u) \subseteq \mathcal{S}|z]$$

for almost every z in the support of Z and all closed sets $\mathcal{S}$ on the support of V. Here $\Pr_0$ indicates probabilities taken with respect to $F^0_{YX|Z}$ and $P_V(\mathcal{S})$ is the probability mass the distribution P_V assigns to the set $\mathcal{S}$. By the "identified set" we mean the set comprising all and only admissible duples (u, P_V) which deliver the distributions $F^0_{YX|Z}$ for almost every z in the support of Z.[2]

We show that the only sets $\mathcal{S}$ that need to be considered when judging whether a particular pair (u, P_V) are in the identified set are unions of sets on the support of $\mathcal{T}_v(Y, X; u)$, with the property that the union of the interiors of these sets is a connected set. When X is discrete this implies that the identified set is characterized by a finite number of inequalities, and an algorithm is provided enabling computation of the collection of such sets and their corresponding moment inequalities.

We also develop characterizations of two outer regions within which the identified set is guaranteed to lie. Even if interest ultimately lies in the identified set, computation of these outer regions is generally a simpler task and may therefore be a useful first step in computation of the identified set. Alternatively, an outer region may be sufficiently informative in the context of any particular model to address the question at hand.

Consider a model which specifies P^*_V as the distribution of V and utility functions u^* for which $p(y, x; u^*, P^*_V)$ is the probability that $Y = y$ given $X = x$ when V and X are independently distributed. In the classical conditional logit model with utility functions

$$u^*_y(x) = x'\beta^*_y$$

[2] Some authors term this the "sharp identified set".

the probabilities involved are the following well known expressions.

$$p(y, x; u^*, P_V^*) = \frac{\exp(x'\beta_y^*)}{1 + \sum_{y'=1}^{M-1} \exp(x'\beta_{y'}^*)}.$$

Our first outer region associated with conditional distributions of Y and X given Z, $F_{YX|Z}^0$, in the case of discrete X, contains all utility functions u^* and distributions P_V^* such that the inequalities:

$$p(y, x; u^*, P_V^*) \geq \max_{z \in \mathcal{Z}} \{\Pr_0[Y = y \wedge X = x | Z = z]\} \tag{1.1}$$

hold for all y and x in the support of Y and X. Here $\mathcal{Z}$ denotes the support of the instrumental variables. Any researcher in a position to calculate a parametric likelihood function when explanatory variables X are assumed exogenous is able to calculate our outer regions directly. In the conditional logit case this outer region is convex which simplifies computation. Our second outer region provides a refinement of this region that can be informative with discrete and continuous X.

1.2 Related results

The prior literature on multinomial choice models is substantial. Only a small subset of this literature has allowed for endogeneity. An important early contribution is in Matzkin (1993) where it is shown that, if the unobservable components of utility from the different alternatives are identically distributed and conditionally independent of one another, and if there is an alternative-specific regressor with large support, then the latent utility functions can be nonparametrically identified. Lewbel (2000) shows how a special regressor can be used to achieve point-identification in various qualitative response models, including multinomial choice models where the joint distribution of the error and regressors is independent of the special regressors conditional on the instrument. Some recent papers have provided sufficient conditions for point-identification under alternative assumptions. This includes the use of triangular structures as in Petrin and Train (2010), who provide a control function approach, and Fox and Gandhi (2009), who provide sufficient conditions for identification in a fully nonparametric recursive setting. Chiappori, Komunjer, and Kristensen (2011) provide an alternative route to nonparametric identification, relying on conditional independence and completeness conditions that differ from the marginal independence restrictions imposed here. In limited dependent variables models with simultaneity, Matzkin (2012) builds on the results of Matzkin (2008) to provide conditions for the nonparametric identification of structural functions and the distribution of unobserved heterogeneity when there are exogenous regressors with large support.

Also related is the recent literature on the estimation of demand for differentiated products by means of random coefficient discrete-choice models pioneered by Berry, Levinsohn, and Pakes (1995). This approach uses the insight of Berry (1994) to allow for the endogeneity of prices. The setting in which this method is applied differs from ours in that demand estimation is carried out on market-level data that consists of a large number of markets. Berry and Haile (2010) and Berry and Haile (2009) establish conditions for nonparametric identification, the latter when micro-level data is also available, as in Berry, Levinsohn, and Pakes (2004). The endogenous variable in these models is product price, which varies across alternatives and markets, but not across individuals. Our model allows endogenous variables to differ across individuals, and does not require either variables that differ across alternatives or covariates with large support.

There are antecedents to our work that partially identify quantities of interest in other models of discrete choice. Chesher (2010) and Chesher and Smolinski (2010) study *ordered* discrete outcome models with endogeneity. Those papers provide set identification results for a single equation specification for an ordered choice, which includes endogenous covariates. In this paper we focus on choices from unordered sets of alternatives. This differs fundamentally by requiring a utility specification for each of the alternatives. Each utility function admits an unobservable, and as a consequence the present context is one in which there are multiple sources of unobserved heterogeneity, rather than a single source. Other research on partially-identifying models of multinomial response includes Manski (2007) and Beresteanu, Molchanov, and Molinari (2011), although the models studied and the mechanisms by which partial identification is obtained in these papers are quite distinct. Manski (2007) provides bounds on predicted choice probabilities from counterfactual choice sets using variation in choices made by individuals who previously faced heterogeneous choice sets. Beresteanu, Molchanov, and Molinari (2011) provide sharp bounds on the parameters of multinomial response model with interval data on regressors, demonstrating general identification results derived from random set theory. Papers with set identifying results for parameters of binary choice models include Manski and Tamer (2002), Magnac and Maurin (2008), and Komarova (2007).

To establish that our bounds are sharp we make use of important results from random set theory, in particular Artstein's inequality (Artstein (1983)). Such methods have been previously used to establish set identification in other contexts by Beresteanu, Molchanov, and Molinari (2011), Galichon and Henry (2011), and Beresteanu, Molchanov, and Molinari (2012). Beresteanu, Molchanov, and Molinari (2011) use the Aumann expectation of set-valued random variables to tractably characterize the identified set in models with convex moment predictions. Their characterization is shown to apply rather generally, covering as examples models of games with multiple equilibria, and best linear prediction and multinomial choice models with interval data. Galichon and Henry (2011) characterize the identified set of structural features in econometric models of normal form games through the use of inequalities generated by the Choquet capacity functional. They provide several approaches to facilitate the computational tractability of this approach, with further results pertaining to optimal transportation given in Ekeland, Galichon, and Henry (2010). Beresteanu, Molchanov, and Molinari (2012) illustrate how random set theory can be employed across a variety of models, paying particular attention to the selection problem in the analysis of treatment effects and best linear prediction, and discussing the relative merits of the capacity functional and Aumann expectation approaches in different contexts.

Our use of random set theory for identification analysis of an instrumental variable model of multiple discrete choice is novel, though the main device employed, Artstein's inequality, has been used in the above papers. Unlike previous approaches, our construction makes use of random sets defined on the space of unobservables, rather than on the outcome space. In models of games with strategic interactions among agents that can yield multiple mixed or pure-strategy equilibria, and that have been the focus of much of the previous research, exogenous variation is obtained from agents' observed payoff shifters. In our setup the choice problem entails a single decision maker, and exogenous variation is provided by instruments that are excluded from agents' utility functions and independent of unobserved heterogeneity. Moreover, our use of random set theory provides a characterization of the identified set that applies in fully nonparametric, semi-parametric, and parametric models. We employ the notion of core-determining classes defined in Galichon and Henry (2011) to refine our characterization of the identified set. They show how this can be done in econometric models of games under a monotonicity condition which is not satisfied in our model. Thus, we provide a novel algorithm for the construction of core-determining classes in our setup.

There are now a variety of methods for estimation and inference available when model parameters are set identified. We show in this paper that the identified set delivered by our model, and the outer regions we provide, can be represented by a set of conditional moment inequalities. Papers that provide methods for estimation and inference on parameters characterized by conditional moment inequalities are therefore applicable. For instance, when covariates and instruments are discrete the identified set is characterized by a finite number of moment inequalities, and one may apply the methods proposed by Chernozhukov, Hong, and Tamer (2007), Beresteanu and Molinari (2008), Romano and Shaikh (2008), Rosen (2008), Galichon and Henry (2009), Bugni (2010), or Canay (2010), among others. When covariates or instruments are continuous, there are infinitely many moment inequalities to incorporate, and one may employ for example the methods of Andrews and Shi (2009), Chernozhukov, Lee, and Rosen (2009), Kim (2009), or Menzel (2009) for estimation and inference.

1.3 Plan of the paper

The paper proceeds as follows. Section 2 defines the instrumental variable multiple discrete choice model with which we work throughout.

Section 3 develops our main identification results. In Section 3.1 we provide a theorem that characterizes the identified set of structural functions applicable in both parametric and nonparametric models. In Section 3.2 we show that when X and V are independent, equivalently if $Z = X$, our characterization reduces to a system of equalities for the conditional probabilities $\Pr_0[Y = y|X = x]$ for all $(y, x) \in \text{Supp}(Y, X)$, which are precisely likelihood contributions if the model is parametrically specified. In Section 3.3 we provide a theorem that defines a minimal system of "core determining" inequalities that are all that need to be considered when calculating the identified set. In Section 3.4 we provide two easy-to-compute outer regions.

In Section 4 the results are illustrated for three-choice models, core determining inequalities are listed for the binary explanatory variable case and identified sets and outer regions are calculated and displayed for an instrumental variable version of the conditional logit model studied by McFadden (1974). Section 5 concludes.

2 The Instrumental Variable Model

We begin with a model that allows utility functions to be nonseparable in components of unobserved heterogeneity, and then specialize our results to the separable case, on which much of the previous literature on models of multiple discrete choice has focused.

2.1 Nonseparable Utility

An individual makes one choice from M alternatives obtaining utility U_y from alternative y as follows.

$$U_y = u_y(X, V_y) \qquad y \in \mathcal{Y} \equiv \{1, 2, \ldots, M\}, \tag{2.1}$$

where for each $y \in \mathcal{Y}$, $U_y : \text{Supp}(X, V_y) \to \mathbb{R}$, where $\text{Supp}(A, B)$ denotes the joint support of any two random vectors A, B. The elements of X are observed variables and the elements of V are unobservable variables that capture heterogeneity in tastes across individuals. Thus the specification of utility from each alternative $y \in \mathcal{Y}$ is dependent upon an alternative-specific unobservable V_y. Each utility function $u_y(\cdot, \cdot)$, is assumed monotone in its second argument, with strict monotonicity imposed for all $y < M$, as we formalize in Restriction A5 below. In Section 2.2 we consider the common special case where the utility functions are additively separable in unobservables.

The elements of Z are observable variables which are required to be jointly independently distributed with $V \equiv (V_1, ..., V_M)$.

Individuals are utility maximizers, observing the value of U and choosing an alternative that gives the highest utility, so that

$$Y \in h_v(X, V; u) \equiv \underset{y \in \mathcal{Y}}{\arg\max}\ u_y(X, V_y), \tag{2.2}$$

where $U \equiv (U_1, \ldots U_M)$. This formulation allows for the possibility of multiple utility-maximizing choices, and in this case remains agnostic as to the determination of Y among these. However, due to monotonicity of the utility functions $u_y(\cdot, \cdot)$ in their second argument coupled with Restriction A4 below, ties in the value delivered by any two alternatives occur with probability zero, and the utility-maximizing alternative is unique with probability one conditional on any realization of (X, Z). We impose sufficient conditions for this both for convenience and because it is common in models of multiple discrete choice, but the tools of random set theory we employ can be applied to models where outcome variables are not uniquely determined, see e.g. Beresteanu, Molchanov, and Molinari (2011) and Galichon and Henry (2011) and the present setup can be easily modified to accommodate ties in utility-maximizing choices.[3] The model is comprised of the following restrictions.

[3] Specifically Theorem 1 goes through without modification if the conditional distribution of $V|X, Z$ is not absolutely continuous with respect to Lebesgue measure, while the results on core-determining class in Section 3.3 would require some modification. Chesher and Rosen (2011) consider simultaneous equations models of discrete choice for which multiple or indeed no solutions are feasible. This raises further issues of

Restriction A1: (Y, X, Z, V) are defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, where $\mathcal{F}$ contains the Borel sets. The support of Y is a finite set $\mathcal{Y} \equiv \{1, 2, ..., M\}$, and the supports of X and Z are $\mathcal{X}$ and $\mathcal{Z}$, respectively. The joint support of (Y, X, Z) is a (possibly non-strict) subset of $\mathcal{Y} \times \mathcal{X} \times \mathcal{Z}$. For any (x, z) on the support of (X, Z) the support of V conditional on $X = x$ and $Z = z$, denoted $\text{Supp}(V|X = x, Z = z)$ is an open subset of $\mathbb{R}^M$ with strictly positive Lebesgue measure. Likewise the support of the marginal distribution of V, denoted $\mathcal{V}$, is an open, positive Lebesgue measure subset of $\mathbb{R}^M$.
Restriction A2: For each value $z \in \mathcal{Z}$ there is a conditional distribution of (Y, X) given $Z = z$, $F^0_{YX|Z}(y, x|z)$. The associated conditional distribution of X given $Z = z$ is denoted by $F^0_{X|Z}(x|z)$. The conditional distributions $F^0_{YX|Z}(y, x|z)$ and $F^0_{X|Z}(x|z)$ are identified by the sampling process. The marginal distribution of Z is either identified by the sampling process or known a priori.
Restriction A3: Given (V, X, Z), Y is determined by (2.1) and (2.2).
Restriction A4: For any (x, z) on the support of (X, Z), the conditional distribution of $V|\,(X = x, Z = z)$ is absolutely continuous with respect to Lebesgue measure with everywhere positive density on its support, $\text{Supp}(V|X = x, Z = z) \subseteq \mathbb{R}^M$. The marginal distribution of V belongs to a specified family of distributions $\mathcal{P}_V$.
Restriction A5: The utility functions $u = \{u_1, ..., u_M\}$ belong to a specified family of functions $\mathcal{U}$ such that for all $x \in \mathcal{X}$, $u_y(x, \cdot)$ is continuous for all $y \in \mathcal{Y}$, is strictly monotone increasing for all $y < M$, and $u_M(x, \cdot)$ is weakly monotone increasing.
Restriction A6: V and Z are stochastically independent.

Restriction A1 formally defines the probability space on which (Y, X, Z, V) lives. It also provides some weak conditions on their support. The support of (Y, X, Z) is not required to be the product of their marginal supports. The support of unobservable V may vary when conditioning upon different realization of X and Z, but is required to be an open, positive Lebesgue measure subset of $\mathbb{R}^M$. This includes the typical case where $\text{Supp}(V|X = x, Z = z) = \mathbb{R}_M$ for all (x, z).

In our analysis of the identifying power of this model we determine the set of observationally equivalent structures which are admitted by the model and deliver the probability distribution $F^0_{YX|Z}(y, x|z)$ of Restriction A2. Throughout the notation "Pr_0" will indicate probabilities calculated using these distributions. Under Restriction A2 the distribution of Z is either identified or a priori known, for example if individual observations are intentionally drawn in accord with a particular distribution of Z. All statements regarding almost every

coherence and completeness that are logically distinct from the study of multiple discrete choice.

$z \in \mathcal{Z}$ are made with respect to this distribution.

Restriction A6 requires V and the variables Z to be independently distributed. Of course this restriction has no force unless Z has some role in the determination of X. The model employed here is silent about this role unlike other models used in the analysis of multiple discrete choice with potentially endogenous explanatory variables.

In Restriction A4 the family of distributions $\mathcal{P}_V$ can be more or less constrained in particular applications allowing consideration of nonparametric or parametric specifications. Restriction A5 similarly allows consideration of parametric and nonparametric specifications of utility functions. Note that although we do not assume the existence of alternative-specific covariates in our analysis, this restriction is fully compatible with these, as it allows for the possibility that only one of the utility functions $u_y(\cdot)$ varies with a particular subset of components of X. Moreover, we impose strict monotonicity of all but one of the utility functions in its corresponding unobservable, and weak monotonicity of the remaining utility function in its unobservable. Combined with Restriction A4 this guarantees that conditional on any realization of (X, Z) there is a unique utility maximizing choice of Y almost surely.

2.2 Separable Utility

A common restriction in analyses of multiple discrete choice is additive separability of the utility functions in unobservable components. This entails a restriction on the class of utility functions $\mathcal{U}$, formally expressed below as Restriction A5*. Since the optimal selection of alternatives is entirely determined by utility differences it is convenient here to impose the normalization that $u_M(x) = 0$ for all $x \in \mathcal{X}$.

Restriction A5* (Additive Separability): Restriction A5 holds with the added restriction that for any $u \in \mathcal{U}$, $u_y(X, V_y) \equiv u_y(X) + V_y$ where for each $y \in \mathcal{Y}$, $u_y : \mathcal{X} \to \mathbb{R}$, and where the normalization $u_M(X) = 0$ is imposed.

Two popular examples of models that satisfy additive separability, each placing different sets of restrictions on the family of distributions $\mathcal{P}_V$ are the following.

1. In an instrumental variable (IV) extension of McFadden's (1974) conditional logit model there is just one distribution in the family $\mathcal{P}_V$, namely the distribution in which the elements of V are *mutually independently* distributed with *common* extreme value distribution function as follows.

$$\Pr[\bigwedge_{y \in \mathcal{Y}} (V_y \leq v_y)] = \prod_{y \in \mathcal{Y}} \exp(-\exp(-v_y)) \tag{2.3}$$

In McFadden's (1974) model the class of utility functions $\mathcal{U}$ is restricted to the parametric family in which $u_y(X) \equiv X'\beta_y$ for $y \in \mathcal{Y}$ and each vector β_y is nonstochastic.

2. The same restriction on $\mathcal{U}$ applies in an IV generalization of the conditional probit model studied in Hausman and Wise (1978) which specifies $\mathcal{P}_V$ as a parametric family of multivariate normal, $N(0, \Sigma)$, distributions with a suitable normalization of Σ.

Note that unlike the classical conditional logit and multinomial probit models, the specifications above do not require X and V to be independently distributed. The specification of $\mathcal{P}_V$ restricts the *unconditional* distribution of V, P_V, to be i.i.d. Type I Extreme Value or multivariate normal, respectively. Due to the independence Restriction A6 the conditional distribution of V given $Z = z$ is also P_V for any instrument value $z \in \mathcal{Z}$, but the conditional distributions of $V|X = x$ or $V|\left(X = x, Z = z\right)$ can differ. An implication is that in the conditional logit model above the components of V need not be independently distributed conditional on either the realization of X or that of (X, Z). Thus the model need not adhere to independence of irrelevant alternatives once we condition upon these variables.

Note that with the additively separable specification of utility, utility-maximizing choices can be deduced from knowledge of utility functions u, covariates X, and $W \equiv (W_1, ..., W_{M-1}) \in \mathbb{R}^{M-1}$, where for each $y \in \mathcal{Y}$,

$$W_y \equiv V_y - V_M.$$

To see why define the utility differences

$$\Delta U_y\left(X, W\right) \equiv U_y - U_M = u_y\left(X\right) + W_y.$$

Then there is a convenient representation for the selection of alternatives equivalent to (2.2) given by

$$Y \in h_w\left(X, W; u\right)$$

with h_w defined as follows.

$$h_w\left(x, w; u\right) \equiv \left\{ y \in \mathcal{Y} : \min_{y' \in \mathcal{Y}} \left(\Delta U_y(x, w) - \Delta U_{y'}(x, w)\right) \geq 0 \right\} \tag{2.4}$$

Because the dependence of the structural function $h_w(X, W; u)$ on the utility functions listed in u is crucial it is made explicit in the notation. Under restriction A4 it continues to hold that the set $h_w(x, W; u)$ is singleton with probability one for all $x \in \mathcal{X}$.[4]

The model requires the random components of utility, V, to have a distribution in the family $\mathcal{P}_V$. From the above we see that when Restriction A5* is imposed P_V is observationally equivalent to any P'_V that produces the same distribution of W, denoted P_W. Thus when additive separability is imposed we let $\mathcal{P}_W$ denote the family of probability distributions for the random utility *differences*, W, implied by $\mathcal{P}_V$. In this case our interest is in the identification of the utility functions listed in $u \in \mathcal{U}$ and the probability distribution $P_W \in \mathcal{P}_W$ that generate the distributions of Restriction A2.[5] This reduces by one the effective dimension of unobserved heterogeneity whose distribution we seek to set-identify. This will prove convenient for the illustration of three-choice models taken up in Section 4, permitting representation of sets of unobservables in $\mathbb{R}^2$.

3 Identification

3.1 The identified set

We now develop results on the identifying power of the IV model of multiple discrete choice. The task is to infer what structures are admitted by the model given knowledge of $F^0_{YX|Z}$. The structures admitted are characterized by a duple, $\mathcal{D} \equiv (u, P_V)$, comprising a list of utility functions, u, and a distribution of random utility shifters, P_V.[6] To characterize the identified set for (u, P_V), we consider for any candidate (u, P_V), the probability that the multivariate unobservable V lies in a collection of *test sets*. For any such test set $\mathcal{S}$ it is shown that the restrictions of the IV model and knowledge of $F^0_{YX|Z}$ combined with the candidate utility function u are compatible with a collection of upper and lower bounds on $P_V(\mathcal{S})$, the probability that P_V assigns to the event $\{V \in \mathcal{S}\}$. The set of (u, P_v) pairs that satisfy these inequality restrictions taken over any collection of test sets $\mathcal{S}$ comprise bounds on $\mathcal{D}$. We show that taken over a sufficiently rich collection of test sets $\mathcal{S}$ the implied bounds are sharp, delivering *the identified set*, which we denote $\mathcal{D}^0(\mathcal{Z})$. In general the collection of

[4]Note that Restriction A4 implies that the distribution of W conditional on X, Z is absolutely continuous with respect to Lebesgue measure.

[5]Note that due to additive separability any P_W with density f_W is observationally equivalent to any P_V that has density $f_V(v_1, ..., v_M) = f_W(v_1 - v_M, ..., v_{M-1} - v_M) \cdot f_{V_M}(v_m)$, for any density $f_{V_M}(\cdot)$ on the support of V_M. Thus when additive separability is imposed, knowledge of the identified set for (u, P_W) implies knowledge of the identified set for (u, P_V), and vice-versa, so there is no loss of generality in restricting attention to P_W.

[6]In additively separable models we can replace V with W defined above and P_V with P_W, and the subsequent derivations go through identically.

all closed sets in $\mathcal{V}$, denoted $\mathsf{F}(\mathcal{V})$, is sufficiently rich to characterize the sharp identified set. In Section 3.3 we show how in the context of any particular model one can characterize a smaller collection of test sets that are sufficient for characterization of the identified set. We refer to these collections of test sets as *core-determining classes* as in Galichon and Henry (2011).[7]

Key in what follows are the sets of values of the unobservable variables V that, for a particular list of utility functions, u, deliver the value y of Y as a utility-maximizing choice when $X = x$, defined as follows:

$$\mathcal{T}_v(y, x; u) \equiv \{v : y \in h_v(x, v; u)\} = \{v : \forall k \in \mathcal{Y},\ u_y(x, v_y) \geq u_k(x, v_k)\}.$$

Note that for any admissible u and each value x, the sets $\mathcal{T}_v(y, x; u)$, $y \in \mathcal{Y}$ form a partition of $\mathbb{R}^M$, ignoring shared boundaries which under Restriction A4 have measure zero according to P_V.

In the additively separable case with Restriction A5* imposed we can likewise define

$$\begin{aligned} \mathcal{T}_w(y, x; u) &\equiv \{w : \forall k \in \mathcal{Y},\ u_y(x) + w_y \geq u_k(x) + w_k\} \\ &= \{(v_1 - v_M, ..., v_{M-1} - v_M) : v \in \mathcal{T}_v(y, x; u)\}. \end{aligned}$$

Using this set, we can then replace V with W, P_V with P_W, and $\mathcal{V}$ with $\mathcal{W} \equiv \mathrm{Supp}(W)$, and the following derivations go through identically. These sets are illustrated for particular structural functions in Section 4. Because the derivations are otherwise identical we proceed in this section with the more general case where only Restriction A5 is imposed. Moreover, under restriction A5*, one can recover $\mathcal{T}_v(y, x; u)$ from knowledge of $\mathcal{T}_w(y, x; u)$ through the relation

$$\mathcal{T}_v(y, x; u) = \{(w_1 + c, ..., w_{M-1} + c, c) : w \in \mathcal{T}_w(y, x; u),\ c \in \mathbb{R}\}.$$

Consider now a family of conditional distributions $P_{V|XZ}$ for $(x, z) \in \mathrm{Supp}(X, Z)$ and for any test set $\mathcal{S} \subseteq \mathcal{V}$ let $P_{V|XZ}(\mathcal{S}|x, z)$ denote the associated conditional probability of the event $\{V \in \mathcal{S}\}$ given $X = x$ and $Z = z$. Recall that $F^0_{X|Z}$ denotes the conditional distribution functions of X given Z associated with the particular distributions $F^0_{YX|Z}$ of Restriction A2.

We first consider an implication of the IV model's *independence* restriction, Restriction

[7] Throughout we use a calligraphic font, e.g. $\mathcal{S}$, to denote a set and a sans serif font, e.g. K, to denote a collection of sets.

A6.

- **Independence**: The IV model requires V and Z to be independently distributed. It follows that for a choice $P_V \in \mathcal{P}_V$ all associated conditional distributions $P_{V|XZ}$ that (i) are admitted by the IV model and (ii) can generate the particular probability distributions of Restriction A2 must satisfy the condition

$$\int_{x \in \mathcal{X}} P_{V|XZ}(\mathcal{S}|x,z) dF^0_{X|Z}(x|z) = P_V(\mathcal{S}) \tag{3.1}$$

for all values $z \in \mathcal{Z}$ and test sets $\mathcal{S} \subseteq \mathcal{V}$. The left hand side of (3.1) is the conditional probability $P_{V|Z}(S|z)$ which the independence restriction requires to be invariant with respect to z.

Now consider *observational equivalence* conditions which all admissible utility functions $u \in \mathcal{U}$ and probability distributions $P_V \in \mathcal{P}_V$ must satisfy if they are to be capable of delivering the probability distributions of Restriction A2.

- **Observational equivalence**. Since for any value, x, of X, the utility functions u deliver $Y = y$ uniquely for almost every $V \in \mathcal{T}_v(y,x;u)$, and for no $V \notin \mathcal{T}_v(y,x;u)$, there is the requirement that, associated with P_V, there are conditional distributions $P_{V|XZ}$ such that for all $(y,x,z) \in \text{Supp}(Y,X,Z)$:

$$P_{V|XZ}(\mathcal{T}_v(y,x;u)|x,z) = \text{Pr}_0[Y = y|X = x, Z = z]. \tag{3.2}$$

These two implications of the IV model's restrictions lead to a system of inequalities which must be satisfied by all admissible duples that deliver the particular distributions of Restriction A2, that is all duples in the identified set associated with $F^0_{YX|Z}$ for $z \in \mathcal{Z}$. This system of inequalities is now derived.

Considering any test set $\mathcal{S} \subseteq \mathcal{V}$, equation (3.2) places restrictions on $P_{V|XZ}(\mathcal{S}|x,z)$ and the utility functions u associated with duples in $\mathcal{D}^0(\mathcal{Z})$.

First, if (3.2) is to be satisfied then the *smallest* value that $P_{V|XZ}(\mathcal{S}|x,z)$ can take is equal to the sum of the probabilities $\text{Pr}_0[Y = y|X = x, Z = z]$ associated with all sets $\mathcal{T}_v(y,x;u)$ contained *entirely within* $\mathcal{S}$. This is expressed in the inequality

$$P_{V|XZ}(\mathcal{S}|x,z) \geq \sum_{y \in \mathcal{Y}} 1[\mathcal{T}_v(y,x;u) \subseteq \mathcal{S}] \, \text{Pr}_0[Y = y|X = x, Z = z] \tag{3.3}$$

which holds for all $(x,z) \in \text{Supp}(X,Z)$.

Second, for any test set $\mathcal{S}$, the *largest* value that $P_{V|XZ}(\mathcal{S}|x,z)$ can take is equal to the sum of the probabilities $\text{Pr}_0[Y=y|X=x,Z=z]$ associated with all sets $\mathcal{T}_v(y,x;u)$ that have a non-null intersection with $\mathcal{S}$. This is expressed in the following inequality which holds for all $(x,z) \in \text{Supp}(X,Z)$. The symbol ϕ denotes the empty set.

$$P_{V|XZ}(\mathcal{S}|x,z) \leq \sum_{y\in\mathcal{Y}} 1[\mathcal{T}_v(y,x;u) \cap \mathcal{S} \neq \phi] \, \text{Pr}_0[Y=y|X=x,Z=z] \tag{3.4}$$

Marginalizing with respect to X given $Z=z$ on the left and right hand side of the inequalities (3.3) and (3.4) and simplifying using (3.1) there are the following inequalities.

$$P_V(\mathcal{S}) \geq \int\limits_{x\in\mathcal{X}} \left(\sum_{y\in\mathcal{Y}} 1[\mathcal{T}_v(y,x;u) \subseteq \mathcal{S}] \, \text{Pr}_0[Y=y|X=x,Z=z] \right) dF^0_{X|Z}(x|z) \tag{3.5}$$

$$P_V(\mathcal{S}) \leq \int\limits_{x\in\mathcal{X}} \left(\sum_{y\in\mathcal{Y}} 1[\mathcal{T}_v(y,x;u) \cap \mathcal{S} \neq \phi] \, \text{Pr}_0[Y=y|X=x,Z=z] \right) dF^0_{X|Z}(x|z) \tag{3.6}$$

All duples (u, P_V) in the identified set $\mathcal{D}^0(\mathcal{Z})$ satisfy these inequalities for all $z \in \mathcal{Z}$ and all $\mathcal{S} \subseteq \mathcal{V}$. So the inequalities (3.5) and (3.6) obtained as $\mathcal{S}$ passes across all test sets $\mathcal{S} \subseteq \mathcal{V}$ comprise a system of inequalities that defines at least an outer region for the identified set of duples. Note that given a choice of $u \in \mathcal{U}$ with knowledge of the distributions $F^0_{YX|Z}$ of Restriction A2 the right hand sides of these inequalities can be calculated for any test set $\mathcal{S}$, and for any such $\mathcal{S}$, given a choice $P_V \in \mathcal{P}_V$ the left hand sides of the inequalities can be calculated. We will shortly show that the system of inequalities taken over all $\mathcal{S}$ that are closed subsets of $\mathcal{V}$ define the identified set.

To facilitate that development it is convenient to express the inequalities (3.5) and (3.6) in terms of set valued random variables as in Beresteanu, Molchanov, and Molinari (2011) and Galichon and Henry (2011).

To this end, define random sets $\mathcal{T}_v(Y,x;u)$ and $\mathcal{T}_v(Y,X;u)$ as

$$\mathcal{T}_v(Y,x;u) \equiv \{v : h_v(x,v;u) = Y\},$$

and

$$\mathcal{T}_v(Y,X;u) \equiv \{v : h(X,v;u) = Y\},$$

which are random closed sets on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ of Restriction A1.[8]

Probability distributions of random closed sets are completely characterized either by containment functionals or by capacity functionals, see e.g. Molchanov (2005) Sections 1.1.2 and 1.1.6.[9] The containment and capacity functionals of $\mathcal{T}_v(Y, X; u)$ conditional on $X = x$ and $Z = z$ under the particular probability distributions of Restriction A2 are respectively

$$\Pr_0\left[\mathcal{T}_v(Y, X; u) \subseteq S | X = x, Z = z\right] = \sum_{y \in \mathcal{Y}} 1[\mathcal{T}_v(y, x; u) \subseteq \mathcal{S}] \Pr_0[Y = y | X = x, Z = z]$$

and

$$\Pr_0\left[\mathcal{T}_v(Y, X; u) \cap S \neq \phi | X = x, Z = z\right] = \sum_{y \in \mathcal{Y}} 1[\mathcal{T}_v(y, x; u) \cap \mathcal{S} \neq \phi] \Pr_0[Y = y | X = x, Z = z]$$

which are precisely the expressions on the right hand sides of respectively (3.3) and (3.4).

Similarly the containment and capacity functionals of $\mathcal{T}_v(Y, X; u)$ conditional on $Z = z$ *alone*, under the particular probability distributions of Restriction A2 are respectively

$$\Pr_0\left[\mathcal{T}_v(y, x; u) \subseteq S | Z = z\right] = \int_{x \in \mathcal{X}} \left(\sum_{y \in \mathcal{Y}} 1[\mathcal{T}_v(Y, X; u) \subseteq \mathcal{S}] \Pr_0[Y = y | X = x, Z = z] \right) dF^0_{X|Z}(x|z)$$

and

$$\Pr_0\left[\mathcal{T}_v(y, x; u) \cap S \neq \phi | Z = z\right] = \int_{x \in \mathcal{X}} \left(\sum_{y \in \mathcal{Y}} 1[\mathcal{T}_v(Y, X; u) \cap \mathcal{S} \neq \phi] \Pr_0[Y = y | X = x, Z = z] \right) dF^0_{X|Z}(x|z)$$

which are the expressions on the right hand sides of respectively (3.5) and (3.6).

It follows that all admissible duples (u, P_V) with probability distributions $P_V \in \mathcal{P}_V$ and utility functions $u \in \mathcal{U}$ that deliver the particular distributions in Restriction A2 satisfy the inequalities:

[8]These are random closed sets because the sigma-algebra $\mathcal{F}$ is endowed with the Borel sets. This guarantees that for any compact set $S \subseteq \mathbb{R}^{M-1}$, the events $\{\mathcal{T}_v(Y, x; u) \cap S \neq \phi\}$ and $\{\mathcal{T}_v(Y, X; u) \cap S \neq \phi\}$ are $\mathcal{F}$-measurable. For a formal definition of random closed sets see e.g. Molchanov (2005) or Beresteanu, Molchanov, and Molinari (2012) Appendix A.

[9]Specifically, the Choquet Theorem in Molchanov (2005), page 10, originally from Choquet (1954), implies that the capacity functional of a random closed set, taken over all compact sets of the relevant carrier space, uniquely determines its distribution. The same holds for the containment functional applied to all closed sets, see Molchanov (2005) page 22.

$$\Pr_0\left[\mathcal{T}_v(Y,X;u) \subseteq \mathcal{S} | Z = z\right] \leq P_V(\mathcal{S}) \leq \Pr_0\left[\mathcal{T}_v(Y,X;u) \cap \mathcal{S} \neq \phi | Z = z\right] \tag{3.7}$$

for all sets $\mathcal{S} \subseteq \mathcal{V}$ and instrumental values $z \in \mathcal{Z}$.

Capacity and containment functionals are equivalent characterizations of the distribution of a random set because for all $\mathcal{S} \subseteq \mathcal{V}$ and $z \in \mathcal{Z}$,

$$\Pr_0\left[\mathcal{T}_v(Y,X;u) \subseteq \mathcal{S} | Z = z\right] = 1 - \Pr_0\left[\mathcal{T}_v(Y,X;u) \cap \mathcal{S}^c \neq \phi | Z = z\right] \tag{3.8}$$

where $\mathcal{S}^c$ is the complement of $\mathcal{S}$. So the inequalities generated by the lower and upper bounds in (3.7) as $\mathcal{S}$ passes through *all* subsets of $\mathcal{V}$ are identical. It follows that only one of the bounds in (3.7) need be considered. We work henceforth with the lower bounding probability given by the conditional containment functional of $\mathcal{T}_v(Y,X;u)$.

The following theorem states that all and only duples (u, P_V) which satisfy the system of inequalities generated by the lower bound in (3.7) for all $z \in \mathcal{Z}$ and all $\mathcal{S}$ that are closed subsets of $\mathcal{V}$ deliver the distributions of Restriction A2, that is that the system of inequalities defines the identified set of duples.

Theorem 1 *Let restrictions A1-A6 hold. Then the identified set of admissible duples (u, P_V) associated with the conditional distributions $F^0_{YX|Z}$, $z \in \mathcal{Z}$, is*

$$\mathcal{D}^0(\mathcal{Z}) \equiv \left\{(u, P_V) \in \mathcal{U} \times \mathcal{P}_V : \Pr_0\left[\mathcal{T}_v(Y,X;u) \subseteq \mathcal{S} | Z = z\right] \leq P_V(\mathcal{S}),\ \forall \mathcal{S} \in \mathsf{F}(\mathcal{V}) \ \text{a.e. } z \in \mathcal{Z}\right\}, \tag{3.9}$$

where $\mathsf{F}(\mathcal{V})$ denotes the set of all closed subsets of $\mathcal{V}$.

Proof. *$\mathcal{D}^0(\mathcal{Z})$ contains all duples $(u, P_V) \in \mathcal{U} \times \mathcal{P}_V$ that satisfy for all $\mathcal{S} \in \mathsf{F}(\mathcal{V})$,*

$$\Pr_0\left[\mathcal{T}_v(Y,X;u) \subseteq \mathcal{S} | Z = z\right] \leq P_V(\mathcal{S})$$

for almost every $z \in \mathcal{Z}$. The preceding development shows that all admissible duples that deliver the conditional distributions $F^0_{YX|Z}$, $z \in \mathcal{Z}$ lie in this set. Further, a key result from random set theory, namely Artstein's inequality, provided by Artstein (1983) and Norberg (1992), see also Molchanov (2005) Section 1.4.8, guarantees sharpness, that is that all admissible duples in the set $\mathcal{D}^0(\mathcal{Z})$ can deliver the conditional distributions $F^0_{YX|Z}$, for almost every $z \in \mathcal{Z}$. To apply this result, we first proceed in similar fashion to that of the proof of Theorem 2.1 in Beresteanu, Molchanov, and Molinari (2012) to show that the containment functional inequalities of (3.9) are equivalent to Artstein's inequality. To do so consider any

$(u, P_V) \in \mathcal{D}^0(\mathcal{Z})$ and fix $z \in \mathcal{Z}$. Then with probability one we have that

$$\Pr_0 \left[\mathcal{T}_v(Y, X; u) \subseteq \mathcal{S} | Z = z \right] \leq P_V(\mathcal{S}), \; \forall \mathcal{S} \in \mathsf{F}(\mathcal{V}), \tag{3.10}$$

by definition of $\mathcal{D}^0(\mathcal{Z})$. Now using $P_V(\mathcal{S}) = 1 - P_V(\mathcal{S}^c)$ and

$$\Pr_0 \left[\mathcal{T}_v(Y, X; u) \subseteq \mathcal{S} | Z = z \right] = 1 - \Pr_0 \left[\mathcal{T}_v(Y, X; u) \cap \mathcal{S}^c \neq \phi | Z = z \right],$$

it follows that (3.10) holds if and only if

$$\Pr_0 \left[\mathcal{T}_v(Y, X; u) \cap \mathcal{S}^c \neq \phi | Z = z \right] \geq P_V(\mathcal{S}^c), \; \forall \mathcal{S} \in \mathsf{F}(\mathcal{V}),$$

or equivalently

$$\Pr_0 \left[\mathcal{T}_v(Y, X; u) \cap \mathcal{S} \neq \phi | Z = z \right] \geq P_V(\mathcal{S}), \; \forall \mathcal{S} \in \mathsf{G}(\mathcal{V}),$$

where $\mathsf{G}(\mathcal{V})$ is the collection of all open subsets of $\mathcal{V}$. By Corollary 1.4.44 of Molchanov (2005) this is in turn equivalent to the collection of inequalities

$$\Pr_0 \left[\mathcal{T}_v(Y, X; u) \cap \mathcal{S} \neq \phi | Z = z \right] \geq P_V(\mathcal{S}), \; \forall \mathcal{S} \in \mathsf{K}(\mathcal{V}),$$

where $\mathsf{K}(\mathcal{V})$ is the collection of all compact subsets of $\mathcal{V}$. This relation is Artstein's inequality. By Artstein (1983) and Norberg (1992) it follows that there exists a random variable $\tilde{V}$ and a random set $\tilde{\mathcal{T}}$ realized on the same probability space as $(V, \mathcal{T}_v(Y, X; u))$ such that conditional on $Z = z$, both $\tilde{V} \sim P_V$ and $\tilde{\mathcal{T}}$ is distributed identically to $\mathcal{T}_v(Y, X; u)$ when (Y, X) is distributed $F^0_{YX|Z}(\cdot | Z = z)$, with $\tilde{V} \in \tilde{\mathcal{T}}$ with probability one. This implies that conditional on $Z = z$ there exist random variables $\left(\tilde{Y}, \tilde{X}\right)$ defined on the same probability space with $\tilde{V} \in \mathcal{T}_v(\tilde{Y}, \tilde{X}; u)$ and $\left(\tilde{Y}, \tilde{X}\right)$ distributed $F^0_{YX|Z}(\cdot | Z = z)$. The choice of $z \in \mathcal{Z}$ is arbitrary and the inequality defining $\mathcal{D}^0(\mathcal{Z})$ holds for almost every $z \in \mathcal{Z}$. Thus the argument holds for almost every $z \in \mathcal{Z}$, implying there exist random variables $\left(\tilde{Y}, \tilde{X}\right)$ conditionally distributed $F^0_{YX|Z}$ a.e. $z \in \mathcal{Z}$ so that Restriction A2 is satisfied. ■

Corollary 1 *If Restriction A5 is replaced with the additive separability Restriction A5*, the* *identified set for* (u, P_W) *is*

$$\mathcal{D}_w^0(\mathcal{Z}) \equiv \{(u, P_V) \in \mathcal{U} \times \mathcal{P}_V : \Pr_0 [\mathcal{T}_w(Y, X; u) \subseteq \mathcal{S} | Z = z] \leq P_W(\mathcal{S}), \forall \mathcal{S} \in \mathsf{F}(\mathcal{W}) \ \ a.e. \ z \in \mathcal{Z}\}, \tag{3.11}$$

where $\mathsf{F}(\mathcal{V})$ *denotes the set of all closed subsets of* $\mathcal{W}$*.*

Proof. The proof is identical to the proof of Theorem 1 upon replacing V with W and P_V with P_W. ■

Remarks

1. Key to the proof of sharpness is Artstein's inequality, which states that for any random set $\mathcal{T}$ and any random variable $V \in \mathbb{R}^M$ such that

$$\Pr [\mathcal{T} \cap \mathcal{S} \neq \phi] \geq P_V(\mathcal{S}), \forall \mathcal{S} \in \mathsf{K}(\mathcal{V}),$$

   we can couple with V and $\mathcal{T}$ a random variable $\tilde{V}$ and a random set $\tilde{\mathcal{T}}$, respectively, living on the same probability space and with the same distributions as the original random variable V and random set $\mathcal{T}$, such that $\tilde{V} \in \tilde{\mathcal{T}}$ with probability one. Our proof makes use of the existence of such a coupling conditional on each instrumental value $z \in \mathcal{Z}$ to show that every duple (u, P_V) in $\mathcal{D}^0(Z)$ can produce the distributions $F^0_{YX|Z}$ of Restriction A2.

2. In the definition of the identified set $\mathcal{D}^0(\mathcal{Z})$ the containment functional inequality:

$$\Pr_0 [\mathcal{T}_v(Y, X; u) \subseteq \mathcal{S} | Z = z] \leq P_V(\mathcal{S}), \forall \mathcal{S} \in \mathsf{F}(\mathcal{V})$$

   can be replaced by the capacity functional inequality:

$$\Pr_0 [\mathcal{T}_v(Y, X; u) \cap \mathcal{S} \neq \phi | Z = z] \geq P_V(\mathcal{S}), \forall \mathcal{S} \in \mathsf{K}(\mathcal{V}).$$

3. The inequalities of Theorem 1 are required to hold for almost every $z \in \mathcal{Z}$ so for each $\mathcal{S} \in \mathsf{F}(\mathcal{V})$ only the maximum over $z \in \mathcal{Z}$ of the lower bounds is binding.

4. The development so far allows for the possibility that there are no parametric restrictions on the classes of utility functions $\mathcal{U}$ and probability distributions $\mathcal{P}_V$. When there are parametric restrictions these classes of functions are indexed by a finite dimensional parameter. It may be the case that only one of $\mathcal{U}$ and $\mathcal{P}_V$ are parametrically specified,

or that either are semiparametrically specified, in which case the model restrictions are semiparametric.

3.2 Relation to independent X and V

When X and V are stochastically independent the above characterization reduces to the usual maximum likelihood probabilities, and hence yields point identification under appropriate restrictions on the distribution of X. To show this is the case, we can apply the above analysis by taking $X = Z$, and considering the lower bound of (3.7),

$$\Pr_0\left[\mathcal{T}_v(Y, X; u) \subseteq \mathcal{S} | X = x\right] \leq P_V(S).$$

Setting $\mathcal{S} = \mathcal{T}_v(y, x; u)$ for each $x \in \mathcal{X}$ and any $u \in \mathcal{U}$, we have

$$\forall y \in \mathcal{Y}, \ \ \Pr_0\left[Y = y | X = x\right] \leq P_V(\mathcal{T}_v(y, x; u)),$$

where

$$\sum_{y \in \mathcal{Y}} \Pr_0\left[Y = y | X = x\right] = 1, \text{ and } \sum_{y \in \mathcal{Y}} P_V(\mathcal{T}_v(y, x; u)) = 1.$$

So it follows that

$$\forall y \in \mathcal{Y}, \ \ \Pr_0\left[Y = y | X = x\right] = P_V(\mathcal{T}_v(y, x; u)), \tag{3.12}$$

which holds for $(y, x) \in \text{Supp}(Y, X)$ and with sufficient restrictions on $\mathcal{U}$ and $\mathcal{P}_V$ there may be point identification of u and P_V. For instance, in the conditional logit example given in Section 2 with additive separability holding we have $u_y(x) = x\beta_y$ for $y < M$, and $u_M(x) = 0$, $P_V[\mathcal{T}_v(y, x; u)]$ takes the familiar form

$$P_V[\mathcal{T}_v(y, x; u)] = \frac{\exp\left(x\beta_y\right)}{1 + \sum_{y'=1}^{M-1} \exp\left(x\beta_{y'}\right)}.$$

In this case (3.12) provides precisely the conditional probabilities used in the construction of the classical maximum likelihood estimator, and under the usual rank condition there is point identification, as shown by McFadden (1974). This is easily satisfied in models with discrete regressors, but in semiparametric or nonparametric models with X and V independent, point identification additionally requires more restrictive rank and support conditions. These are not required for the characterization of the identified set provided by Theorem 1.

3.3 Core determining sets

It may not be feasible to consider the complete system of inequalities of Theorem 1 that are generated as $\mathcal{S}$ passes through *all* closed subsets of $\mathcal{V}$. However a system of inequalities based on only some of these sets will deliver at least an outer identification region and this may be useful in practice.

For some models it is possible to find a much smaller collection of the sets $\mathcal{S} \in \mathsf{F}(\mathcal{V})$ whose inequalities define $\mathcal{D}^0(\mathcal{Z})$. This is a *core-determining class* of sets as studied by Galichon and Henry (2011) in obtaining identified sets in models with multiple equilibria.

The result of Theorem 2 below is useful in producing collections of test sets that deliver core-determining classes of inequalities for the models considered in this paper. Unlike Galichon and Henry (2011) we allow these sets to be dependent upon the structural functions u, or, in parametric settings, model parameters. We call these sets *core-determining sets* in what follows. In the characterization of such collections we make use of the notation $int(\mathcal{S})$ and $cl(\mathcal{S})$ to denote the interior and closure, respectively, of any set $\mathcal{S}$. The proof of Theorem 2 makes use of the following lemma, which provides some properties of the sets $\mathcal{T}_v(y, x; u)$. In this lemma and the subsequent analysis we make use of the support of the random set $\mathcal{T}_v(Y, X; u)$,

$$\mathsf{T}_v(Y, X; u) \equiv \{\mathcal{T}_v(y, x; u) : \exists x \in \mathcal{X} \text{ s.t. } \mathbb{P}(Y = y|X = x) > 0\},$$

and likewise the support of $\mathcal{T}_w(Y, X; u)$,

$$\mathsf{T}_w(Y, X; u) \equiv \{\mathcal{T}_w(y, x; u) : \exists x \in \mathcal{X} \text{ s.t. } \mathbb{P}(Y = y|X = x) > 0\}.$$

Lemma 1 *Consider the model defined by Restrictions A1-A6. Under these restrictions, the following results hold: (i) The sets $\mathcal{T}_v(y, x; u)$ on the support of $\mathcal{T}_v(Y, X; u)$ are connected for any $u \in \mathcal{U}$ and $x \in \mathcal{X}$. (ii) If Restriction A5* holds the sets $\mathcal{T}_v(y, x; u)$ and $\mathcal{T}_w(y, x; u)$ are convex. (iii) If Restriction A5* and $\mathcal{V} = \mathbb{R}^M$ these sets are non-empty, with strictly positive Lebesgue measure whenever $u_{y'}(x) - u_y(x) < \infty$ for all $y' \in \mathcal{Y}$, $y' \neq y$.*

Proof. (i) *Consider any $v, v' \in \mathcal{T}_v(y, x; u)$. Define v^* such that* $v^*_y = \max\{v_y, v'_y\}$, and for all $k \neq y$, $v^*_k = \min\{v_k, v'_k\}$. *From the monotonicity Restriction A5 it follows that at the specified x the utility of choice y is weakly higher at $V = v^*$ than at either v or v', that is*

$$u_y(x, v^*_y) \geq u_y(x, v_y) \text{ and } u_y(x, v^*_y) \geq u_y(x, v'_y).$$

Likewise utility from any alternative $k \neq y$ is weakly lower at $V = v^*$ than at either of v, v'. Restriction A5 implies that indeed for any $\tilde{v}$ on the line from v to v^*, an individual with $X = x$ and $V = \tilde{v}$ is at least as disposed to y as an individual with $X = x$ and $V = v$. Thus any such $\tilde{v}$ is an element of $\mathcal{T}_v(y, x; u)$, so that the line from v to v^* constitutes a path in $\mathcal{T}_v(y, x; u)$ that connects these two point. By the same reasoning the line from v' to v^* constitutes a path in $\mathcal{T}_v(y, x; u)$ from v' to v^*. Thus there is a path in $\mathcal{T}_v(y, x; u)$ that connects any two points $v, v' \in \mathcal{T}_v(y, x; u)$, and thus $\mathcal{T}_v(y, x; u)$ is a connected set.[10]

(ii) *If Restriction A5* holds the sets* $\mathcal{T}_v(y, x; u)$ and $\mathcal{T}_w(y, x; u)$ *are convex because for any $u \in \mathcal{U}$ and $x \in \mathcal{X}$ these sets are an intersection of linear half spaces.*[11]

(iii) *If $u_{y'}(x) - u_y(x) = \infty$ for some $y' \neq y$, then the set $\mathcal{T}_w(y, x; u)$ is empty. Otherwise, for any $w_y = v_y - v_M \in \mathbb{R}$ there exists $w_{y'} = v'_{y'} - v_M$ small enough for each $y' \neq y$ such that $w_y - w_{y'} > u_{y'}(x) - u_y(x)$. Therefore the interior of $\mathcal{T}_w(y, x; u)$ is both open and non-empty. Since $\mathcal{T}_w(y, x; u)$ contains its interior and any non-empty open set has positive Lebesgue measure, $\mathcal{T}_w(y, x; u)$ also has positive Lebesgue measure. Note that $\mathcal{T}_v(y, x; u)$ is empty if and only if $\mathcal{T}_w(y, x; u)$ is empty, so the same conclusions hold for $\mathcal{T}_v(y, x; u)$.* ■

The following theorem characterizes core determining classes for the IV model of multiple discrete choice.

Theorem 2 *Let Restrictions A1-A6 hold. The identified set (3.9) of Theorem 1 is given by the inequalities generated by the collection of test sets $\mathcal{S}$ that (i) are unions of sets on the support of $\mathcal{T}_v(Y, X; u)$, and (ii) are such that the union of the interiors of the component sets is a connected set. The same statements hold applied to the characterization given by (3.11) in Corollary 1 if additionally Restriction A5* holds, replacing the support of $\mathcal{T}_v(Y, X; u)$ with that of $\mathcal{T}_w(Y, X; u)$.*

Proof. *We provide the proof for the more general case where Restrictions A1-A5 hold with regard to the characterization (3.9). We separate the proof into two cases, depending on whether or not the set*

$$\mathcal{Z}^{\phi} \equiv \{z \in \mathcal{Z} : \Pr_0 [\mathcal{T}_v(Y, X; u) = \phi | Z = z] > 0\}$$

has positive measure Z, equivalently on whether $\mathcal{T}_v(Y, X; u)$ is empty with positive probability. The proof for the characterization (3.11) where in addition Restriction A5 holds follows*

[10] See e.g. Sutherland (2009) Chapter 12 p.120 for the formal definition of a path and a formal proof that any set with the property that a path exists connecting any two elements is connected.

[11] They are convex polytopes if one uses a definition of "polytope" that does not exclude unbounded sets.

identical steps, replacing V with W.

Case 1: *Fix $(u, P_V) \in \mathcal{U} \times \mathcal{P}_V$ and suppose that $\mathcal{Z}^\phi$ has positive measure. Then ϕ is the union of all the sets $\mathcal{T}_v(y, x; u)$ with $(y, x) \in Supp(Y, X)$ for which $\mathcal{T}_v(y, x; u) = \phi$, i.e. the empty set can be written as a union of sets satisfying (i) and (ii). We now show that any $u \in \mathcal{U}$ for which $\mathcal{Z}^\phi$ has positive measure violates the containment functional inequality evaluated at $\mathcal{S} = \phi$ conditioning on $z \in \mathcal{Z}^\phi$, so that it indeed suffices to only use a test set satisfying conditions (i) and (ii). This is because if the containment functional inequality were satisfied with $\mathcal{S} = \phi$ it would follow that*

$$0 < \Pr_0 [\mathcal{T}_v(Y, X; u) \subseteq \phi | Z = z] \leq P_V(\phi) = 0,$$

which is a contradiction.

Case 2: *Again fix $(u, P_V) \in \mathcal{U} \times \mathcal{P}_V$ and now suppose that $\mathcal{Z}^\phi$ has zero measure. Then for almost every $z \in \mathcal{Z}$ the sets on the support of $\mathcal{T}_v(Y, X; u)$ are connected sets with positive Lebesgue measure. This follows from Restriction A1, which requires that the support of $V|(X = x, Z = z)$ is open, in conjunction with Restriction A5 requiring for all $(y, x) \in Supp(Y, X)$ and all $u \in \mathcal{U}$ that $u_y(x, v_y)$ is continuous in v_y. We now establish conditions (i) and (ii) in turn.*

(i) For any set S let $C_\mathcal{S}(u)$ denote the collection of sets on the support of $\mathcal{T}_v(Y, X; u)$ that are subsets of $\mathcal{S}$. Let

$$\mathsf{G}_\mathcal{S}(u) \equiv \bigcup_{\mathcal{T} \in \mathsf{C}_\mathcal{S}(u)} \mathcal{T},$$

be the union of sets on the support of $\mathcal{T}_v(Y, X; u)$ that are contained in $\mathcal{S}$. Then $G_\mathcal{S}(u) \subseteq \mathcal{S}$ and

$$\Pr_0 [\mathcal{T}_v(Y, X; u) \subseteq \mathcal{S} | Z = z] = \Pr_0 [\mathcal{T}_v(Y, X; u) \subseteq \mathsf{G}_\mathcal{S}(u) | Z = z].$$

It follows that if the inequalities of Theorem 1 hold for all unions of sets on the support of $\mathcal{T}_v(Y, X; u)$, then they hold for all sets $\mathcal{S} \subseteq \mathcal{V}$, since for any such $\mathcal{S}$,

$$\Pr_0 [\mathcal{T}_v(Y, X; u) \subseteq \mathsf{G}_\mathcal{S}(u) | Z = z] \leq P_V (\mathsf{G}_\mathcal{S}(u)) \leq P_V (\mathcal{S}),$$

where the final inequality follows by $\mathsf{G}_\mathcal{S}(u) \subseteq \mathcal{S}$.

(ii) We now show that the inequalities associated with those sets $\mathsf{G}_\mathcal{S}(u)$ such that (ii) does

not hold are redundant. Define

$$\mathsf{G}^0_{\mathcal{S}}(u) \equiv \bigcup_{\mathcal{T}\in\mathsf{C}_{\mathcal{S}}(u)} int\,(\mathcal{T})\,,$$

and suppose that $\mathsf{G}^0_{\mathcal{S}}(u)$ is not connected. Then $C_{\mathcal{S}}(u)$ can be divided into mutually exclusive and exhaustive sub-collections of sets each belonging to $C_{\mathcal{S}}(u)$, the union of whose interiors is connected. That is $C_{\mathcal{S}}(u)$ can be written

$$\mathsf{C}_{\mathcal{S}}(u) = \{\mathsf{C}_{\mathcal{S},1}(u), ..., \mathsf{C}_{\mathcal{S},J}(u)\}\,,$$

for some J, dependent upon $\mathcal{S}$, *such that for any $1 \leq j \leq J$, the sets*

$$\mathsf{G}^0_{\mathcal{S},j}(u) \equiv \bigcup_{\mathcal{T}\in\mathsf{C}_{\mathcal{S},j}(u)} int\,(\mathcal{T})$$

are connected, and for any $j \neq k$, $\mathsf{G}^0_{\mathcal{S},j}(u) \cap \mathsf{G}^0_{\mathcal{S},k}(u) = \phi$. Now define

$$\mathsf{G}_{\mathcal{S},j}(u) \equiv \bigcup_{\mathcal{T}\in\mathsf{C}_{\mathcal{S},j}(u)} \mathcal{T},$$

so that $\mathsf{G}_{\mathcal{S}}(u) = \cup_{j=1}^{J}\mathsf{G}_{\mathcal{S},j}(u)$. Consider any set $\mathcal{T}_v(y,x;u)$ on the support of $\mathcal{T}_v(Y,X;u)$. This set is connected by Lemma 1 and has positive Lebesgue measure, since $\mathcal{Z}^{\phi}$ has zero measure, by the above reasoning. It therefore cannot be contained in both $G_{\mathcal{S},j}(u)$ and $G_{\mathcal{S},k}(u)$ for any $j \neq k$ since $G^0_{\mathcal{S},j}(u) \cap G^0_{\mathcal{S},k}(u) = \phi$. Thus

$$\Pr_0\left[\mathcal{T}_v(Y,X;u) \subseteq \mathsf{G}_{\mathcal{S}}(u)|Z=z\right] = \sum_{j=1}^{J} \Pr_0\left[\mathcal{T}_v(Y,X;u) \subseteq \mathsf{G}_{\mathcal{S},j}(u)|Z=z\right], \tag{3.13}$$

and

$$P_V(\mathsf{G}_{\mathcal{S}}(u)) = \sum_{j=1}^{J} P_V(\mathsf{G}_{\mathcal{S},j}(u)). \tag{3.14}$$

Therefore:

$$\Pr_0\left[\mathcal{T}_v(Y,X;u) \subseteq \mathsf{G}_{\mathcal{S},j}(u)|Z=z\right] \leq P_V(\mathsf{G}_{\mathcal{S},j}(u)) \qquad \forall j \in \{1,\ldots,J\}$$

implies

$$\sum_{j=1}^{J} \Pr_0 \left[\mathcal{T}_v(Y,X;u) \subseteq \mathsf{G}_{\mathcal{S},j}(u) | Z=z\right] \leq \sum_{j=1}^{J} P_V(\mathsf{G}_{\mathcal{S},j}(u))$$

and so by (3.13) and (3.14):

$$\Pr_0 \left[\mathcal{T}_v(Y,X;u) \subseteq \mathsf{G}_{\mathcal{S}}(u) | Z=z\right] \leq P_V(\mathsf{G}_{\mathcal{S}}(u)).$$

■

The following algorithm delivers the collection of sets that define core determining inequalities for discrete X. This collection *varies with the specific utility functions u under consideration* but it is invariant with respect to changes in P_V. Let the support of discrete X be $\mathcal{X} \equiv \{x_1, \ldots, x_K\}$. X may be a finite dimensional vector. The algorithm may be applied to the sets on the support of $\mathcal{T}_v(Y,X;u)$ using the characterization of the identified set in Theorem 1 or in the separable case to sets on the support of $\mathcal{T}_w(Y,X;u)$ using the characterization of Corollary 1. We thus use $\mathcal{T}(y,x;u)$ in what follows to denote either of $\mathcal{T}_v(Y,X;u)$ or $\mathcal{T}_w(Y,X;u)$ throughout the remainder of this section.

For collections of sets C_1 and C_2 let $\mathsf{C}_1 \otimes \mathsf{C}_2$ be the collection of sets obtained when the union of each set in C_1 with each set in C_2 is formed.[12] Let $\mathsf{C}_1 \| \mathsf{C}_2$ denote the collection of the sets that appear either in C_1 or in C_2.[13] Let $\mathsf{C}(u)$ denote the collection of the interiors of the sets on the support of $\mathcal{T}(Y,X;u)$,

$$\mathsf{C}(u) \equiv \{int\,(\mathcal{T}(y,x;u)) : (x,y) \in \text{ Supp}\,(X,Y)\}\,.$$

Let $\mathsf{G}(u)$ denote the list of core determining sets to be produced by the algorithm.

An algorithm for producing core determining sets when X is discrete

1. Initialization. Set $\mathsf{G}(u) = \mathsf{C}(u)$ and $\mathsf{G}^*(u) = \mathsf{C}(u)$.

2. Repeat steps (a)-(c) until the collection of sets $\mathsf{G}^*(u)$ is empty.

[12] This is a Kroneker-product-like operation hence our choice of symbol. For example if $\mathsf{C}_1 = \{\mathcal{C}_{11}, \mathcal{C}_{12}\}$ and $\mathsf{C}_2 = \{\mathcal{C}_{21}, \mathcal{C}_{22}\}$ then

$$\mathsf{C}_1 \otimes \mathsf{C}_2 = \{\mathcal{C}_{11} \cup \mathcal{C}_{21}, \mathcal{C}_{12} \cup \mathcal{C}_{21}, \mathcal{C}_{11} \cup \mathcal{C}_{22}, \mathcal{C}_{12} \cup \mathcal{C}_{22}\}.$$

[13] Thinking of collections of sets as sets of sets the concatenation $\mathsf{C}_1 \| \mathsf{C}_2$ is the union of the "sets" C_1 and C_2.

(a) Create the collection of sets $\mathsf{G}^*(u) \otimes \mathsf{C}(u)$ and place the *connected* sets in this collection that are not already present in $\mathsf{G}^*(u)$ into a collection of sets: $\mathsf{B}(u)$.

(b) Remove any duplicate sets from $\mathsf{B}(u)$.

(c) Let $\mathsf{G}^*(u) = \mathsf{B}(u)$ and replace $\mathsf{G}(u)$ by $\mathsf{G}(u) \| \mathsf{G}^*(u)$.

3. Set $\mathsf{G}(u)$ equal to the collection of closures of its component sets.

Let $\text{Con}(\cdot)$ applied to a list of sets select the connected sets in the list. Step two of the algorithm recursively creates the following list of sets.

$$\mathsf{C}(u) \| \, \text{Con}\,(\mathsf{C}(u) \otimes \mathsf{C}(u)) \, \| \, \text{Con}\,(\text{Con}\,(\mathsf{C}(u) \otimes \mathsf{C}(u)) \otimes \mathsf{C}(u)) \, \| \cdots$$

This is the same as the list

$$\text{Con}\,(\mathsf{C}(u) \| \mathsf{C}(u) \otimes \mathsf{C}(u) \| \mathsf{C}(u) \otimes \mathsf{C}(u) \otimes \mathsf{C}(u) \| \cdots)$$

which is evidently the list of all connected unions of sets on $\mathsf{C}(u)$, but is more efficient computationally. The closures of these sets provide the collection of sets required by Theorem 2, since the closure of a union of open sets is the same as the union of the closure of all the component sets. The algorithm terminates in at most $MK - 1$ iterations.

The algorithm we use to produce core-determining sets in the three-choice examples of Section 4 eliminates duplicates "from the left": first each element of $\mathsf{C}(u)$ is compared with every subsequent element in the list and elements in $\mathsf{C}(u)$ that arise further up the list are deleted, then each element of $\text{Con}\,(\mathsf{C}(u) \otimes \mathsf{C}(u))$ is compared with every subsequent element in the list and elements in $\text{Con}\,(\mathsf{C}(u) \otimes \mathsf{C}(u))$ that arise further up the list are deleted, and so on. The result is that where sets in $\mathsf{C}(u)$ are subsets of other sets in $\mathsf{C}(u)$ the latter (i.e. the "supersets") will appear later in the list than the other elements in $\mathsf{C}(u)$.

An advantage of this approach is that the lists of unions that are obtained reveal precisely which sets in $\mathsf{C}(u)$ lie in each of the unions that comprise the core determining sets. Thus, consider a member, $\mathcal{G}$, of a collection of core determining sets, $\mathsf{G}(u)$. Let $\mathsf{C}_{\mathcal{G}}(u)$ be the sets on the support of $\mathcal{T}(Y, X; u)$ that are subsets of $\mathcal{G}$. These are the lists produced by the algorithm. The lower bound in the inequality associated with the set $\mathcal{G}$ and the instrumental value $z \in \mathcal{Z}$ is:

$$\sum_{\{(y,x):\mathcal{T}(y,x;u) \in \mathsf{C}_{\mathcal{G}}(u)\}} \text{Pr}_0[Y = y \wedge X = x | Z = z].$$

Number of points of support of X	Number of core determining sets	Number of unions of sets in $\mathsf{T}(u)$
2	12	64
3	33	512
4	82	4096
5	188	32768
6	406	262144
7	842	2079152

Table 1: Number of core determining sets in the 3 choice model for each choice of u when (i) X is discrete having K points of support and (ii) utilities are linear in X.

The number of core determining sets is far smaller than the number of possible unions of sets on the support of $\mathcal{T}(Y, X; u)$. For example in a 3 choice model with a binary explanatory variable and separable utility, for any choice of u, there are at most 12 potentially informative core determining sets compared with $2^6 = 64$ possible unions of the 6 sets on the support of $\mathcal{T}(Y, X; u)$. In the three choice example studied in Section 4 in which a linear index restriction is imposed, when X takes just 7 values there are over 2 million unions of the 21 sets on the support of $\mathcal{T}(Y, X; u)$ but the number of potentially informative core determining sets for any choice of u is at most 842 - see Table 1.[14]

3.4 Two easy-to-compute outer regions

When X is discrete there is among the core determining inequalities always one associated with each set on the support of $\mathcal{T}(Y, X; u)$, equivalently, with each set in the collection $\mathsf{C}(u)$. These inequalities require that all duples (u, P_V) in the identified set be such that the inequalities:

$$P_V[\mathcal{T}_v(y, x; u)] \geq \Pr_0[Y = y \wedge X = x | Z = z]$$

hold for all $(y, x, z) \in \text{Supp}(Y, X, Z)$. It follows that:

$$P_V[\mathcal{T}_v(y, x; u)] \geq \max_{z \in \mathcal{Z}} \Pr_0[Y = y \wedge X = x | Z = z] \tag{3.15}$$

must hold for all $(y, x, z) \in \text{Supp}(Y, X, Z)$. These inequalities define an *outer region* within which lies the identified set of duples (u, P_V). This outer region is generally informative with discrete X, but not with continuous X as then the probabilities on the right-hand side of (3.15) are zero. Our second outer region, provided below, can be useful with either

[14]Note that with additive separability imposed the number of core-determining sets does not depend on whether $\mathcal{T}(Y, X; u) = \mathcal{T}_v(Y, X; u)$ or $\mathcal{T}(Y, X; u) = \mathcal{T}_w(Y, X; u)$ is used.

continuous or discrete X.

The probability $P_V[\mathcal{T}_v(y,x;u)]$ that appears on the left hand side is simply the probability assigned by the pair (u, P_V) to the event $Y = y$ when $X = x$. When X is exogenous this is the conditional probability that $Y = y$ given $X = x$. For example in the conditional logit model studied in Section 4 in which $\mathcal{P}_V$ admits only the distribution for V generated by the i.i.d. Type 1 Extreme Value distributions there is:

$$P_V[\mathcal{T}_v(y,x;u)] = \frac{\exp\left(u_y(x)\right)}{1 + \sum_{y'=1}^{M-1} \exp(u_{y'}(x))}, \qquad y \in \{1, \ldots, M\}. \tag{3.16}$$

In general the probability $P_V[\mathcal{T}_v(y,x;u)]$ is the probability that would appear in a classical discrete choice likelihood function (for independent realizations) constructed using (u, P_V) and defined by conditioning on observed values of the explanatory variables X as if they were exogenous. When X is endogenous $P_V[\mathcal{T}_v(y,x;u)]$ is the counterfactual choice probability for alternative y were all members of the population to have their covariates set to x, keeping each of their V fixed.

For all (u, P_V) in the identified set the inequalities (3.15) require that the probability $P_V[\mathcal{T}_v(y,x;u)]$ must exceed the maximal value over $z \in \mathcal{Z}$ of the *joint* probability that $Y = y$ and $X = x$ *conditional* on $Z = z$. Whenever a model is considered for which, under an exogeneity restriction, there is a well defined parametric likelihood function, the outer region defined by these inequalities is very easy and quick to compute.

This outer region can be tightened whenever there is (y, x) for which there exist values of $x' \neq x$ such that $\mathcal{T}_v(y,x';u) \subseteq \mathcal{T}_v(y,x;u)$ because in such cases the containment functional inequality requires:

$$P_V[\mathcal{T}_v(y,x;u)] \geq \int_{(x':\mathcal{T}_v(y,x';u)\subseteq\mathcal{T}_v(y,x;u))} \Pr{}_0[Y = y \wedge X = x'|Z = z] dF^0_{X|Z}(x'|z) .$$

In the three choice models with binary X considered in Section 4 this improvement is obtained for 2 of the 6 sets on the support of $\mathcal{T}_v(Y,X;u)$. In general there are many cases in which such improvements can be obtained. The lower bound in this inequality can be positive with discrete and with continuous X.

4 Illustration: Three choice models

4.1 Core determining sets

In this Section we provide illustrative examples of identified sets, focusing on models for choice among $M = 3$ alternatives in which the utility functions are assumed additively separable and in which X is discrete with finite support $\mathcal{X} \equiv \{x_1, \ldots, x_K\}$. Thus we work with W, P_w, and $\mathcal{T}(Y, X; u) \equiv \mathcal{T}_w(Y, X; u)$ throughout this section. In this case we can give a graphical display of the support of the set valued random variable $\mathcal{T}(Y, X; u)$ in $\mathbb{R}^2$. We provide the core determining inequalities for the case in which $K = 2$ and present numerical examples of identified sets for a variety of values of K.

In the 3 choice model utilities are determined as follows.

$$U_1 = u_1(X) + V_1, \qquad U_2 = u_2(X) + V_2, \qquad U_3 = V_3$$

With $W \equiv (W_1, W_2) = (V_1 - V_3, V_2 - V_3)$ the support of $\mathcal{T}(Y, X; u)$ is:

$$\begin{aligned}
\mathcal{T}(1, x; u) &= \{W : (W_1 \geq -u_1(x)) \wedge (W_1 \geq W_2 - u_1(x) + u_2(x))\} \\
\mathcal{T}(2, x; u) &= \{W : (W_2 \geq -u_2(x)) \wedge (W_1 \leq W_2 - u_1(x) + u_2(x))\} \\
\mathcal{T}(3, x; u) &= \{W : (W_1 \leq -u_1(x)) \wedge (W_2 \leq -u_2(x))\}
\end{aligned}$$

for $x \in \mathcal{X}$. The interior of these $3K$ sets comprise the collection of sets $\mathsf{C}(u)$.

For each value $x \in \mathcal{X}$, the collection of sets: $\mathcal{T}(y, x; u)$, $y \in \{1, 2, 3\}$, is a partition of $\mathbb{R}^2$ "centred" on a point denoted $w(x)$ with coordinates $W_1 = -u_1(x)$ and $W_2 = -u_2(x)$. The collection of sets $G(u)$ that generates the core determining inequalities varies with u, depending on the relative orientation of the points $w(x)$, $x \in \mathcal{X}$.

When $M = 3$ and $K = 2$ there are three such orientations, illustrated in Figure 1. Values of W_1 are measured vertically and values of W_2 are measured horizontally. Sets $\mathcal{T}(1, x; u)$, $\mathcal{T}_w(2, x; u)$ and $\mathcal{T}(3, x; u)$ lie respectively northwest, southeast and southwest of the point $w(x)$ for each of the two possible values of x.[15] The relative orientations of $w(x_1)$ and $w(x_2)$ are distinguished by the *slope* of the line that connects them: (1) in which the slope is negative, (2) in which the slope is positive and less than 1/2 and (3) in which the slope is positive and greater than 1/2. Within each of these cases there is one orientation in which

[15] Koning and Ridder (2003) consider these partitions in a paper studying the falsifiability of utility maximizing models of multiple discrete choice.

$w(x_1)$ lies higher (in the W_1 direction) than $w(x_2)$ and another in which these positions are reversed.

When K is much larger than 2 the number of orientations to be considered may be very large. There is substantial simplification in the case in which X is scalar and $u_1(x)$ and $u_2(x)$ are both linear functions of x. In this case the locus of points described by $w(x)$ as x varies in $\mathcal{X}$ is linear and there are only six orientations to be considered as in the case in which $K = 2$.

Tables 2 and 3 give the collections of sets $\mathsf{G}(u)$ that generate the core determining inequalities. There are 12 sets in each collection, substantially fewer than the $2^6 = 64$ possible unions of sets in the support of $\mathcal{T}(Y, X; u)$.

Table 2 gives the collections for three cases, 1a, 2a, 3a, in which $w(x_2)$ is above $w(x_1)$. Table 3 gives the collections for three cases, 1b, 2b, 3b, in which $w(x_2)$ is below $w(x_1)$. Table 3 is obtained from Table 2 by exchanging indexes identifying the points of support of X.

In these Tables, in each case, only 4 of the 6 sets in $\mathsf{C}(u)$ appear in the initial 4 columns of the Tables. The reason is that, as noted in Section 3.4, in each case two of the six sets in $\mathsf{C}(u)$ are subsets of others. For example, in Case 1a $\mathcal{T}_w(1, x_2; u) \subseteq \mathcal{T}(1, x_1; u)$ and $\mathcal{T}_w(2, x_1; u) \subseteq \mathcal{T}(2, x_2; u)$ (see Figure 1) and, as explained earlier, our algorithm includes the "supersets"

$$\mathcal{T}(1, x_2; u) \cup \mathcal{T}(1, x_1; u) = \mathcal{T}_w(1, x_1; u)$$

and

$$\mathcal{T}(2, x_1; u) \cup \mathcal{T}(2, x_2; u) = \mathcal{T}_w(2, x_2; u)$$

later in the list of core determining sets (in columns 5 and 6 in Case 1a in Table 2).

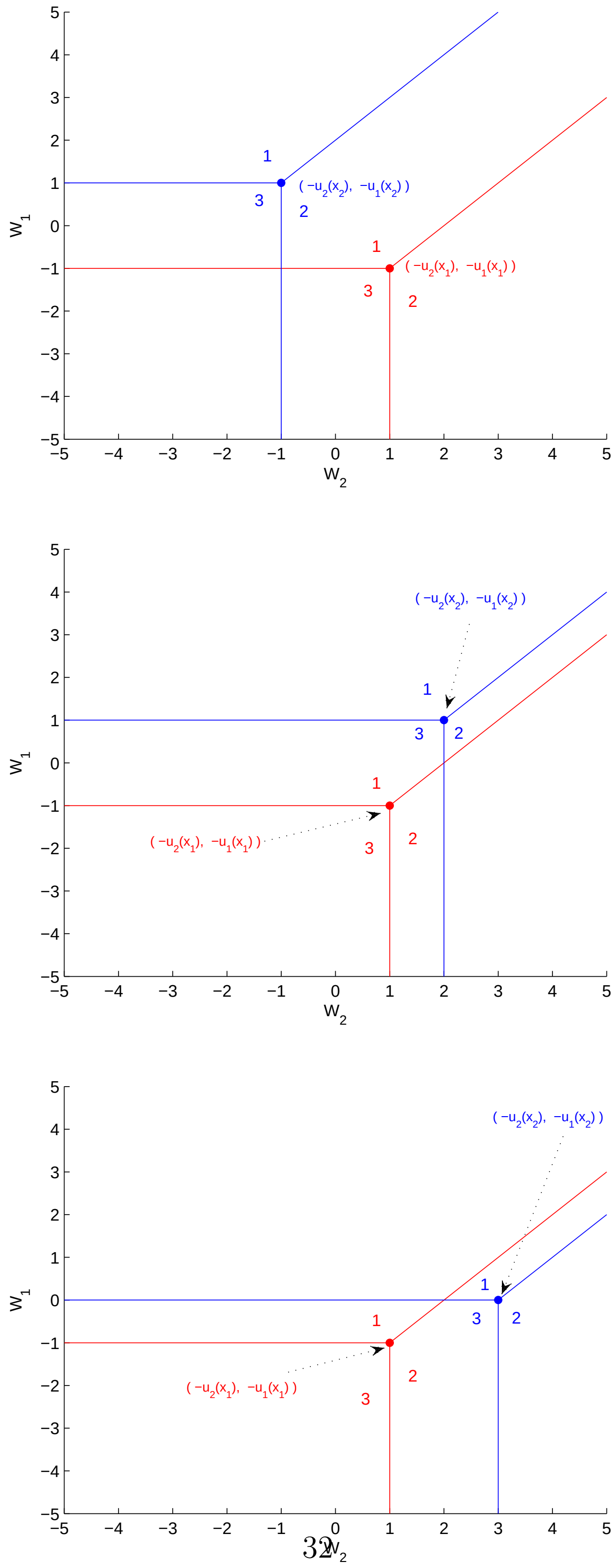
The 12 core determining sets for Case 1a are illustrated in Figures 2 and 3. The first six of these, shown in Figure 2 correspond to those sets on the support of $\mathcal{T}(Y, X; u)$. The remaining six, shown in Figure 3 are non-singleton unions of sets on the support of $\mathcal{T}(Y, X; u)$ obtained by following the algorithm provided above.

4.2 Some calculations

In this Section we give examples of identified sets for a particular probability distribution $F^0_{YX|Z}$. We study cases with $K = 2$ and $K = 4$ and to keep the dimensionality of the identified set small enough to allow a graphical display we impose a linear index restriction.

The model whose identifying power we study has X discrete with support $\mathcal{X} = \{x_1, \ldots, x_K\}$

Figure 1: Orientations of $w(x) = (-u_1(x), -u_2(x))$ when $M = 3$ and $K = 2$, cases 1a, 2a, and 3a.



Case	Support set	Unions											
		1	2	3	4	5	6	7	8	9	10	11	12
1a	$\mathcal{T}(1,x_1;u)$					■			■	■		■	
	$\mathcal{T}(2,x_1;u)$		■				■			■	■		■
	$\mathcal{T}(3,x_1;u)$			■				■			■	■	■
	$\mathcal{T}(1,x_2;u)$	■				■			■	■		■	
	$\mathcal{T}(2,x_2;u)$						■			■	■		■
	$\mathcal{T}(3,x_2;u)$				■			■	■			■	■
2a	$\mathcal{T}(1,x_1;u)$	■				■				■		■	■
	$\mathcal{T}(2,x_1;u)$						■		■		■	■	
	$\mathcal{T}(3,x_1;u)$				■			■		■	■		■
	$\mathcal{T}(1,x_2;u)$		■			■			■			■	■
	$\mathcal{T}(2,x_2;u)$			■			■		■		■	■	
	$\mathcal{T}(3,x_2;u)$							■		■	■		■
3a	$\mathcal{T}(1,x_1;u)$					■			■	■		■	
	$\mathcal{T}(2,x_1;u)$		■				■				■	■	■
	$\mathcal{T}(3,x_1;u)$				■			■		■	■		■
	$\mathcal{T}(1,x_2;u)$	■				■			■	■		■	
	$\mathcal{T}(2,x_2;u)$			■			■		■			■	■
	$\mathcal{T}(3,x_2;u)$							■		■	■		■

Table 2: Blocked cells indicate sets on the support of $\mathcal{T}(Y,X;u)$ that appear in the unions generating the 12 core determining inequalities, M=3, K=2, Case 1a, 2a and 3a.

Case	Support set	Unions											
		1	2	3	4	5	6	7	8	9	10	11	12
1b	$\mathcal{T}(1, x_1; u)$	■				■			■	■		■	
	$\mathcal{T}(2, x_1; u)$						■			■	■		■
	$\mathcal{T}(3, x_1; u)$			■				■	■			■	■
	$\mathcal{T}(1, x_2; u)$					■			■	■		■	
	$\mathcal{T}(2, x_2; u)$		■				■			■	■		■
	$\mathcal{T}(3, x_2; u)$				■			■			■	■	■
2b	$\mathcal{T}(1, x_1; u)$	■				■			■			■	■
	$\mathcal{T}(2, x_1; u)$			■			■		■		■	■	
	$\mathcal{T}(3, x_1; u)$							■		■	■		■
	$\mathcal{T}(1, x_2; u)$		■			■				■		■	■
	$\mathcal{T}(2, x_2; u)$						■		■		■	■	
	$\mathcal{T}(3, x_2; u)$				■			■		■	■		■
3b	$\mathcal{T}(1, x_1; u)$	■				■			■	■		■	
	$\mathcal{T}(2, x_1; u)$		■				■		■			■	■
	$\mathcal{T}(3, x_1; u)$							■		■	■		■
	$\mathcal{T}(1, x_2; u)$					■			■	■		■	
	$\mathcal{T}(2, x_2; u)$			■			■				■	■	■
	$\mathcal{T}(3, x_2; u)$				■			■		■	■		■

Table 3: Blocked cells indicate sets on the support of $\mathcal{T}(Y, X; u)$ that appear in the unions generating the 12 core determining inequalities, M=3, K=2, Case 1b, 2b and 3b.

Figure 2: Core-Determining Sets for Binary X: Sets on the Support of $\mathcal{T}(y, x; u)$

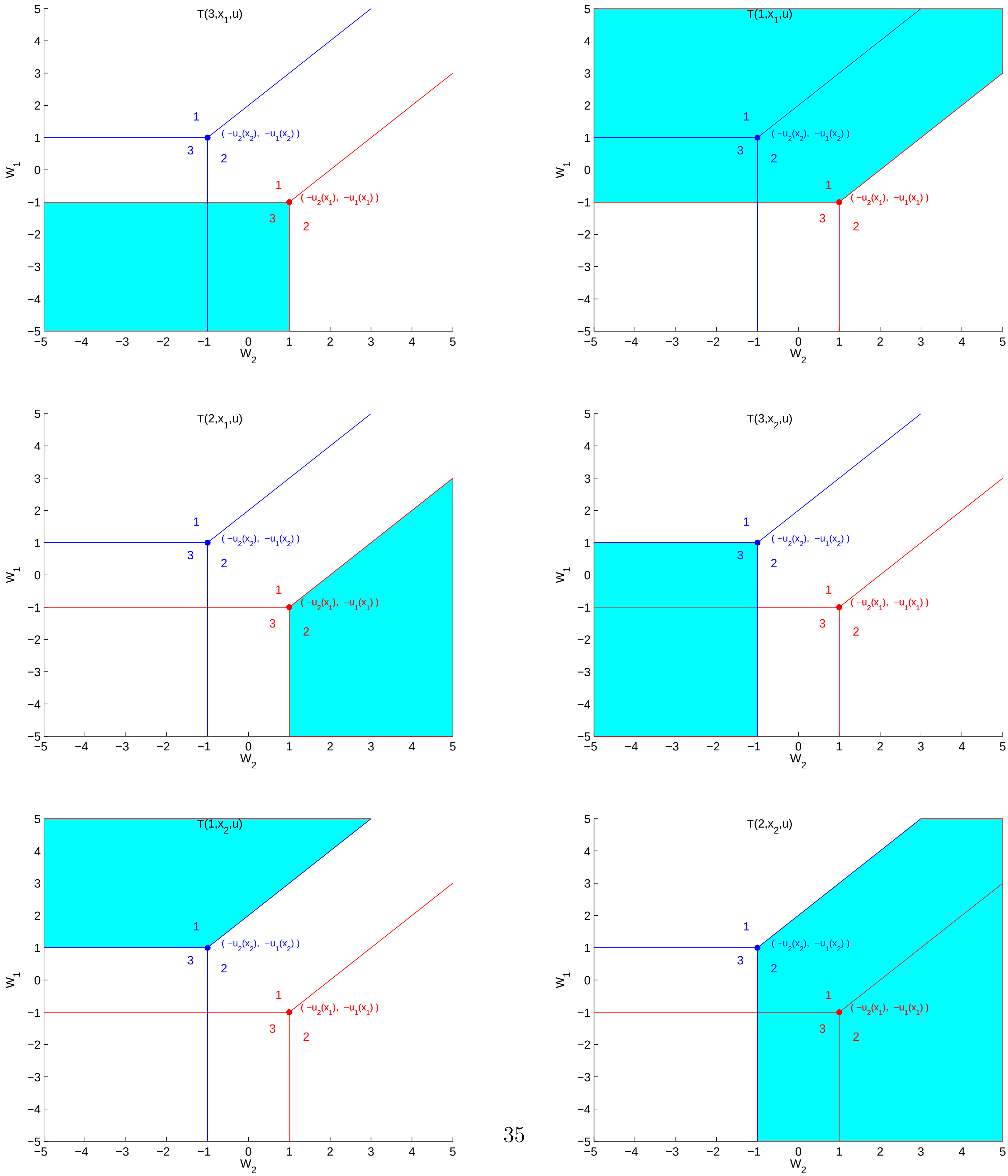
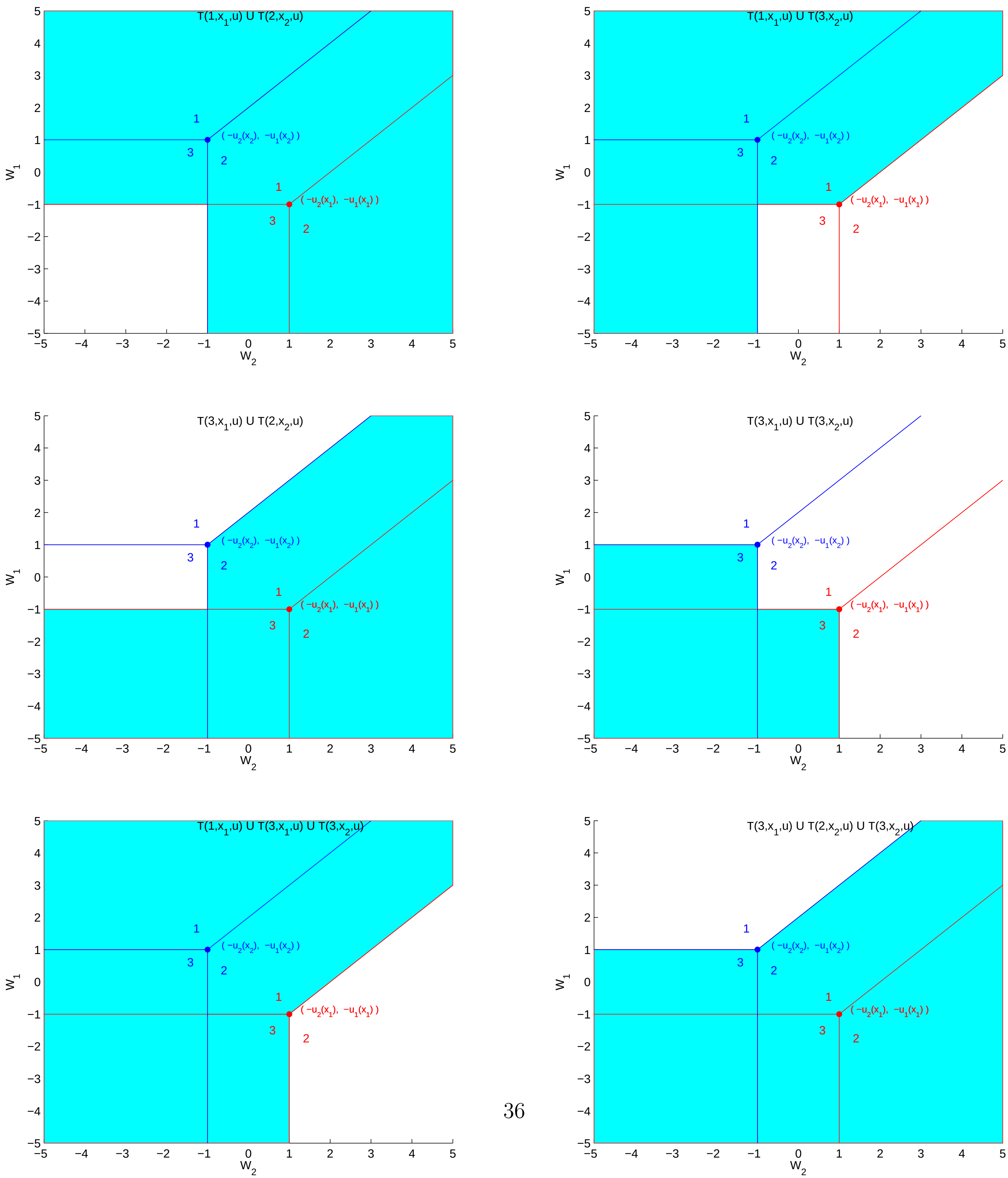


Figure 3: Core-Determining Sets for Binary X: Non-singleton Unions of Sets on the Support of $\mathcal{T}(y, x; u)$



and utility functions determined by a parameter $\alpha = (\alpha_{01}, \alpha_{02}, \alpha_{11}, \alpha_{12})$ as follows.

$$\begin{aligned} u_1(x) &= \alpha_{01} + \alpha_{11} x \\ u_2(x) &= \alpha_{02} + \alpha_{12} x \end{aligned}$$

We generate probabilities from a structure in which a scalar explanatory variable is in fact exogenous. The joint distribution of Y and X given $Z = z$ is specified as ordered probit for X given Z and multinomial logit for Y given X with Y independent of Z given X. Probabilities are as follows.

$$\Pr_0[Y = 1 \wedge X = x_k | Z = z] = \frac{\exp(a_{01} + a_{11} x_k)}{1 + \exp(a_{01} + a_{11} x_k) + \exp(a_{02} + a_{12} x_k)} \left(\Phi\left(\frac{c_k - d_1 z}{d_2}\right) - \Phi\left(\frac{c_{k-1} - d_1 z}{d_2}\right)\right)$$

$$\Pr_0[Y = 2 \wedge X = x_k | Z = z] = \frac{\exp(a_{02} + a_{12} x_k)}{1 + \exp(a_{01} + a_{11} x_k) + \exp(a_{02} + a_{12} x_k)} \left(\Phi\left(\frac{c_k - d_1 z}{d_2}\right) - \Phi\left(\frac{c_{k-1} - d_1 z}{d_2}\right)\right)$$

$$\Pr_0[Y = 3 \wedge X = x_k | Z = z] = \frac{1}{1 + \exp(a_{01} + a_{11} x_k) + \exp(a_{02} + a_{12} x_k)} \left(\Phi\left(\frac{c_k - d_1 z}{d_2}\right) - \Phi\left(\frac{c_{k-1} - d_1 z}{d_2}\right)\right)$$

Here $k \in \{1, 2, \ldots, K\}$, the thresholds c_k are specified a priori, $c_0 \equiv -\infty$, $c_K = \infty$ and scalar z takes values in a set $\mathcal{Z}$, a set of instrumental values to be specified.

Structures like this are admitted by the instrumental variable multiple discrete choice model and in fact have $X \perp\!\!\!\perp V$ but of course this information is not embodied in the IV model whose identifying power we study. That model would be point identifying were that restriction to be imposed. Our calculations give a feel for the degree of ambiguity introduced when the exogeneity restriction is not imposed on X. A computational advantage of this choice of distribution is that probabilities can be calculated without using numerical integration methods.

In these calculations we study the IV extension of McFadden's (1974) model so the family of distributions $\mathcal{P}_V$ is permitted to have just one member which has the three elements of V identically and independently distributed with Type 1 extreme value distributions as in (2.3) with $M = 3$. The associated probability distribution function for the differences W is

$$F_W(w) = \frac{1}{1 + e^{-w_1} + e^{-w_2}}.$$

It is convenient to transform from W to $\tilde{W} = (\tilde{W}_1, \tilde{W}_2)$ using the transformations

$$\tilde{W}_y = \frac{1}{1+\exp(-W_y)}, \qquad W_y = -\log\left(\frac{1}{\tilde{W}_y} - 1\right), \quad y \in \{1,2\}.$$

The support of $(\tilde{W}_1, \tilde{W}_2)$ is the unit square. The joint distribution function of the random variables $\tilde{W}_1$ and $\tilde{W}_2$ is

$$c(\tilde{w}_1, \tilde{w}_2) = \frac{1}{\left(\tilde{w}_1^{-1} + \tilde{w}_2^{-1} - 1\right)}. \tag{4.1}$$

Probabilities $P_W(\mathcal{S})$ are approximated by evaluating the joint distribution function (4.1) over a dense grid of equally spaced values[16]

$$\tilde{w}_{ji} = \frac{i}{n}, \qquad j \in \{1,2\}, i \in \{1,\ldots,n\}$$

on the unit square and second differencing (once with respect to $\tilde{w}_1$ and once with respect to $\tilde{w}_2$) to obtain *exact* probability masses on each cell in the grid. Denote the mass in the cell whose north-east vertex has coordinates w_{1s} and w_{2t} by m_{st}. The probability mass placed by P_W on a set $S \subseteq [0,1]^2$ is approximated by

$$\hat{P}_W(\mathcal{S}) = \sum_{\{(s,t):\ (\tilde{w}_{1s}, \tilde{w}_{2t}) \in \mathcal{S}\}} m_{st}.$$

Define the transformation of the set $\mathcal{T}(y,x;u)$:

$$\tilde{\mathcal{T}}(y,x;u) \equiv \left\{(\tilde{w}_1, \tilde{w}_2) : \left(-\log\left(\frac{1}{\tilde{w}_1} - 1\right), -\log\left(\frac{1}{\tilde{w}_2} - 1\right)\right) \in \mathcal{T}(y,x;u)\right\}$$

which is a subset of the unit square.

The support of $\tilde{\mathcal{T}}(Y,X;u)$ is:

$$\tilde{\mathcal{T}}(1,x;u) = \left\{\tilde{W} : \left(\tilde{W}_1 \geq \frac{1}{1+\exp(u_1(x))}\right) \wedge \left(\tilde{W}_1 \geq \frac{1}{1+\exp\left(u_1(x) - u_2(x)\right)\left(\tilde{W}_2^{-1} - 1\right)}\right)\right\}$$

$$\tilde{\mathcal{T}}(2,x;u) = \left\{\tilde{W} : \left(\tilde{W}_2 \geq \frac{1}{1+\exp(u_2(x))}\right) \wedge \left(\tilde{W}_1 \leq \frac{1}{1+\exp\left(u_1(x) - u_2(x)\right)\left(\tilde{W}_2^{-1} - 1\right)}\right)\right\}$$

[16] A 500×500 grid is used in the calculations reported here.

$$\tilde{\mathcal{T}}(3,x;u)=\left\{\tilde{W}:\left(\tilde{W}_1\le\frac{1}{1+\exp(u_1(x))}\right)\wedge\left(\tilde{W}_2\le\frac{1}{1+\exp(u_2(x))}\right)\right\}$$

for $x\in\mathcal{X}$. These are connected sets which meet at the point

$$\tilde{W}_1=\frac{1}{1+\exp(u_1(x))}\qquad\tilde{W}_2=\frac{1}{1+\exp(u_2(x))},$$

the sets $\tilde{\mathcal{T}}(1,x;u)$, $\tilde{\mathcal{T}}(2,x;u)$ and $\tilde{\mathcal{T}}(3,x;u)$ lying respectively north-west, south-east and south-west of this point. The function separating $\tilde{\mathcal{T}}_w(1,x;u)$ and $\tilde{\mathcal{T}}(2,x;u)$:

$$\tilde{W}_1=\frac{1}{1+\exp\left(u_1(x)-u_2(x)\right)\left(\tilde{W}_2^{-1}-1\right)}$$

is monotone increasing, connecting the point

$$\tilde{W}_1=\frac{1}{1+\exp(u_1(x))}\qquad\tilde{W}_2=\frac{1}{1+\exp(u_2(x))}$$

to the point

$$\tilde{W}_1=1\qquad\tilde{W}_2=1$$

and is concave if $u_1(x)-u_2(x)<0$, linear if $u_1(x)-u_2(x)=0$ and convex if $u_1(x)-u_2(x)>0$.

In the illustrative calculations presented now, probability distributions, $F^0_{YX|Z}$ are generated for cases in which the coefficients in the utility functions are

$$a_{01}=0,\qquad a_{11}=1,\qquad a_{02}=0,\qquad a_{12}=-0.5.$$

The scalar instrumental variable takes two values, -1 and $+1$, the standard deviation parameter in the ordered probit model for X is $d_2=1$ and the slope coefficient is set to $d_1=1$ in one set of calculations (A) and $d_1=1.5$ in another (B). In the latter case the instrumental variable is a better predictor of the value of the variable X and in the discussion we describe this as the "strong instrument" case.

The explanatory variable has $K=2$ points of support in one pair of cases, $\mathcal{X}=\{-1,1\}$ (I) and values are generated using the single threshold $c_1=0$ in the ordered probit specification above. In another pair of cases (II) $K=4$, $\mathcal{X}=\{-1,-1/2,1/2,1\}$ and the thresholds are $c_1=-1/2$, $c_2=0$ and $c_3=1/2$.

Table 4 summarizes the settings for the four cases considered.

Figure 4 shows 2 dimensional projections of the 4 dimensional identified set and of two

Case	K	d_1	a_{01}	a_{11}	a_{02}	a_{12}
I.A	2	1	0	1	0	-1/2
I.B	2	1.5	0	1	0	-1/2
II.A	4	1	0	1	0	-1/2
II.B	4	1.5	0	1	0	-1/2

Table 4: Parameter values used in generating the probability distributions used in the illustrative examples

outer regions for each pair of parameters. Case I.A in which X is binary and the instrument is relatively weak is illustrated in Figure 4. Cases I.B, II.A and II.B are illustrated in Figures 5, 6 and 7.

In each case the results are obtained by calculating membership of identified sets and outer regions at each point on a grid of around $130,000$ values of the 4 parameters and plotting the boundary of the set or outer region for each pairing of parameters. For each pair of values in a 2-D projection of a 4-D set there exists a value of the other two parameters such that the quadruple thus obtained lies in the 4-D set.

In each case three sets are drawn.

1. The inner set (blue) is the identified set obtained using all the core determining inequalities of Theorem 2.

2. The outer set (green) is the outer region obtained using the $3K$ inequalities:

$$\frac{\exp(a_{0y} + a_{1y}x)}{1 + \sum_{y'=1}^{2} \exp(a_{0y'} + a_{1y'}x)} \geq \max_{z \in \mathcal{Z}} \Pr_0[Y = y \wedge X = x | Z = z], \qquad y \in \{1, 2, 3\}, \quad x \in \mathcal{X} \tag{4.2}$$

   implied by (3.15). Since, as shown in McFadden (1974), the logarithms of the choice probabilities on the left hand side of (4.2) are concave functions of the parameters $a \equiv (a_{01}, a_{11}, a_{02}, a_{12})$ these inequalities define a convex set.

3. The intermediate set (magenta) is the set obtained using $3K$ inequalities in which the left hand sides are as in (4.2) but the right hand sides take account of the existence of any x' such that $\mathcal{T}(y, x'; u) \subseteq \mathcal{T}(y, x; u)$. This intermediate set is a proper subset of the other outer region because allowing for the subset relationships leads to some increases in the values appearing on the right hand side of the inequalities (4.2) with no change in the values on the left hand sides. This set cannot be guaranteed convex because the identity of the values x' that are involved in subset relationships depends

on the relative signs and magnitudes of the parameters a_{11} and a_{12}. However in the cases considered here the values a_{11} and a_{12} in the outer region all have $a_{11} > 0$ and $a_{12} < 0$ which implies that the subset relationships do not vary within the set. This outer region is therefore an intersection of linear half spaces and so is convex.

In all four cases examined the calculations suggest that all the 2-D projections are convex. Accordingly the set boundaries we draw are the convex hulls of the points on the grids that are calculated to lie in the each of the projected 2-D sets. In each pane of the figures the red solid diamond locates the parameter value that generates the probability distributions used in this analysis.

The IV model is quite informative. For example the slope coefficients can be signed in the sense that all values of a_{11} and a_{12} in the identified set and the outer regions have $a_{11} > 0$ and $a_{12} < 0$. Comparing Figure 4 with Figure 5 ($K = 2$) and Figure 6 with Figure 7 ($K = 4$) it is clear that the identified set and the outer regions are much smaller in the stronger instrument case.

The sets in Figure 4 ($K = 2$) are substantially smaller than those in Figure 6 ($K = 4$). We believe this occurs because the predictive power of the binary instrumental variable for particular values of X decreases as the number of points of support of X rises. This result is sensitive to changes in the support of the instrumental variable and to changes in the specification of the relationship between potentially endogenous X and the instrumental variable Z.

The outer regions (green, magenta) are around 10 times faster to compute and they are quite informative, in some cases wrapping the identified set quite tightly. In case II.A the intermediate outer region (magenta) is substantially smaller than the extreme outer region. We think this happens because when K is large there are many more subset relationships and these bring substantial refinements of the inequalities defining the extreme outer region.

The probability distributions employed here are generated by structures in which the explanatory variable is exogenous. The model we use, with the addition of the exogeneity restriction, is point identifying, so the extent of the identified sets seen in these illustrations, relative to the solid red diamond demonstrates the identifying power of the exogeneity restriction.

Figure 4: Case I.A. 2-D projections of the identified set and two outer regions, $M = 3$, $K = 2$, weaker instrument.

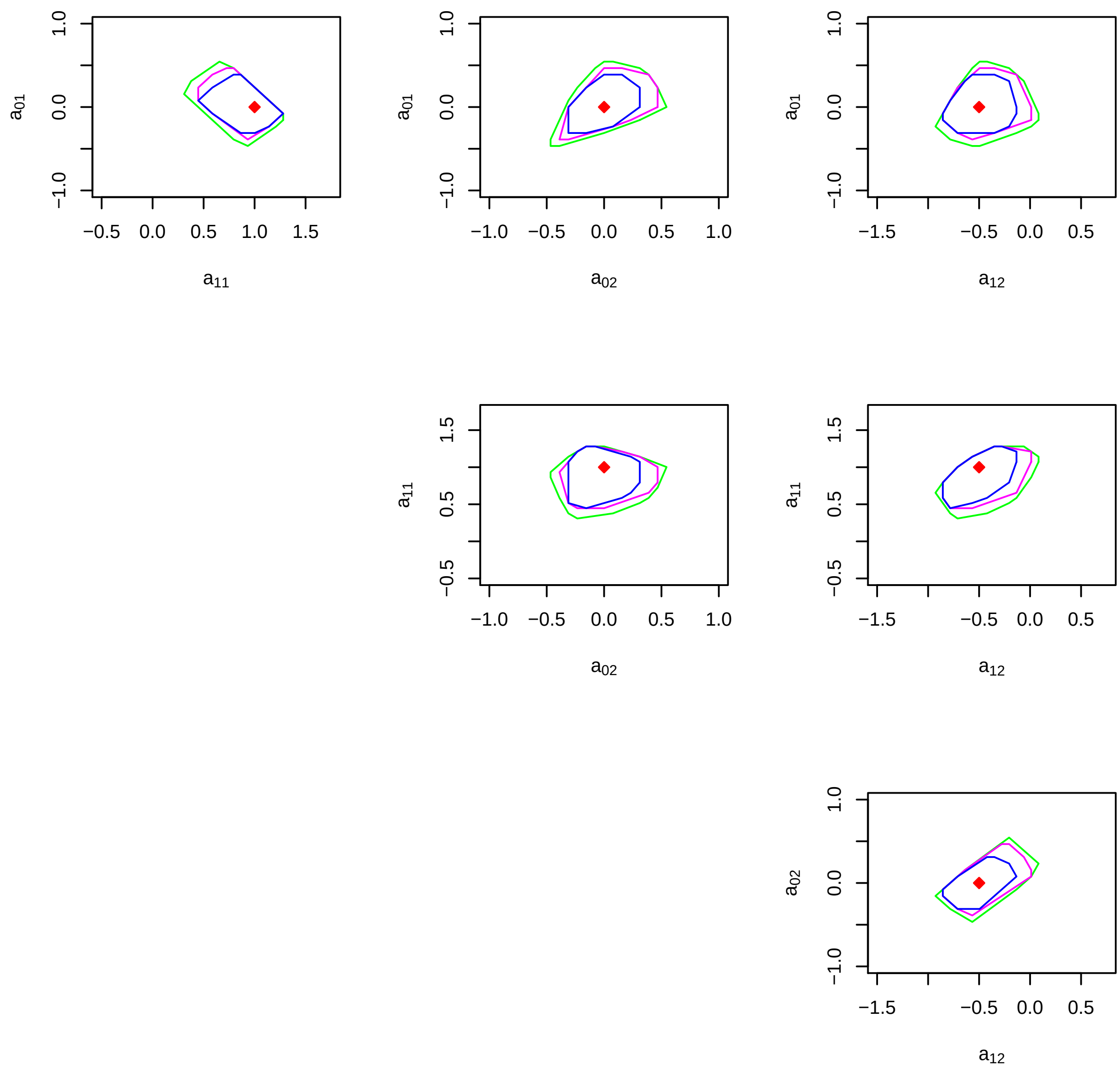


Figure 5: Case I.B. 2-D projections of the identified set and two outer regions, $M = 3$, $K = 2$, stronger instrument.

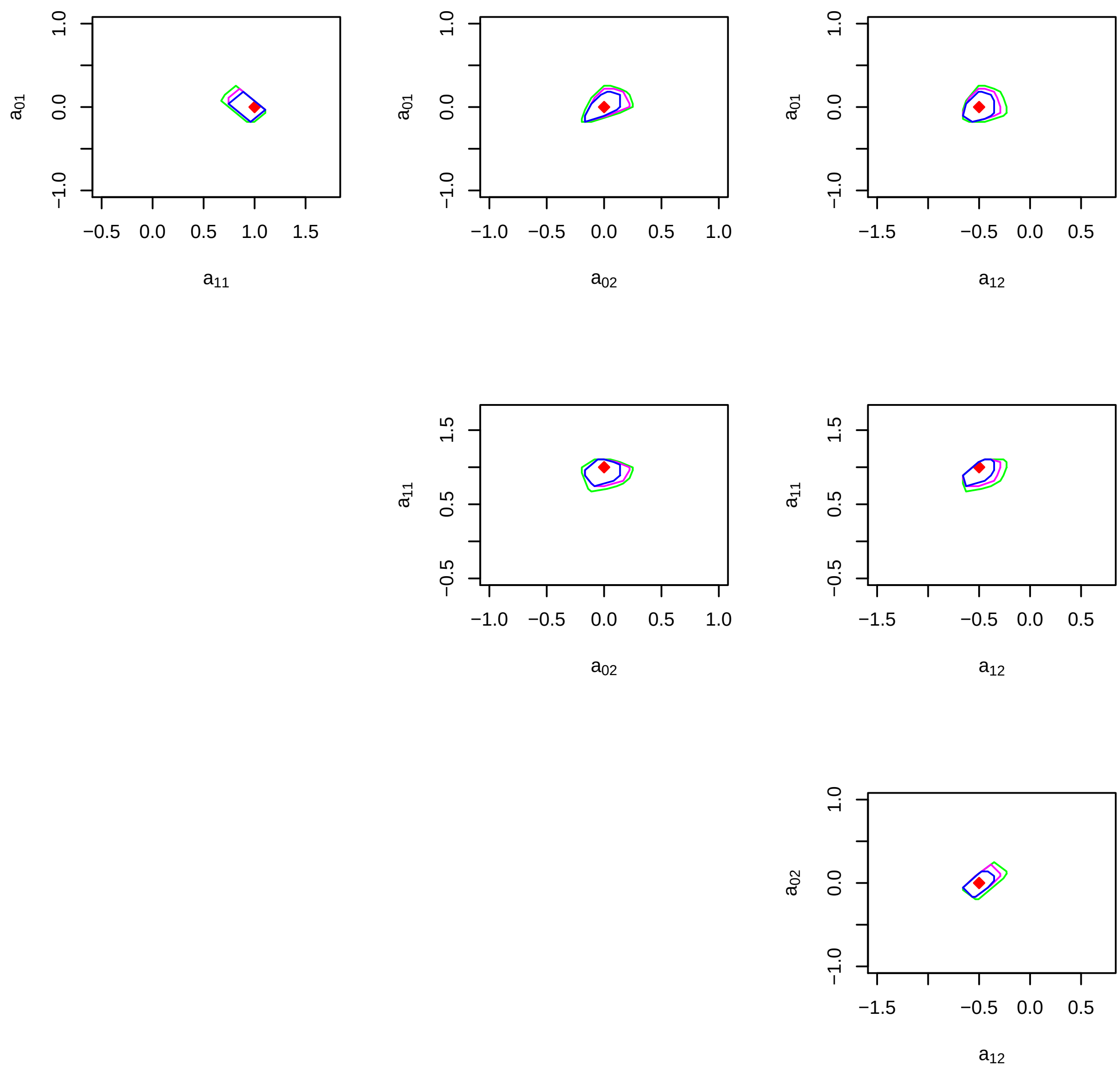


Figure 6: Case II.A. 2-D projections of the identified set and two outer regions, $M = 3$, $K = 4$, weaker instrument.

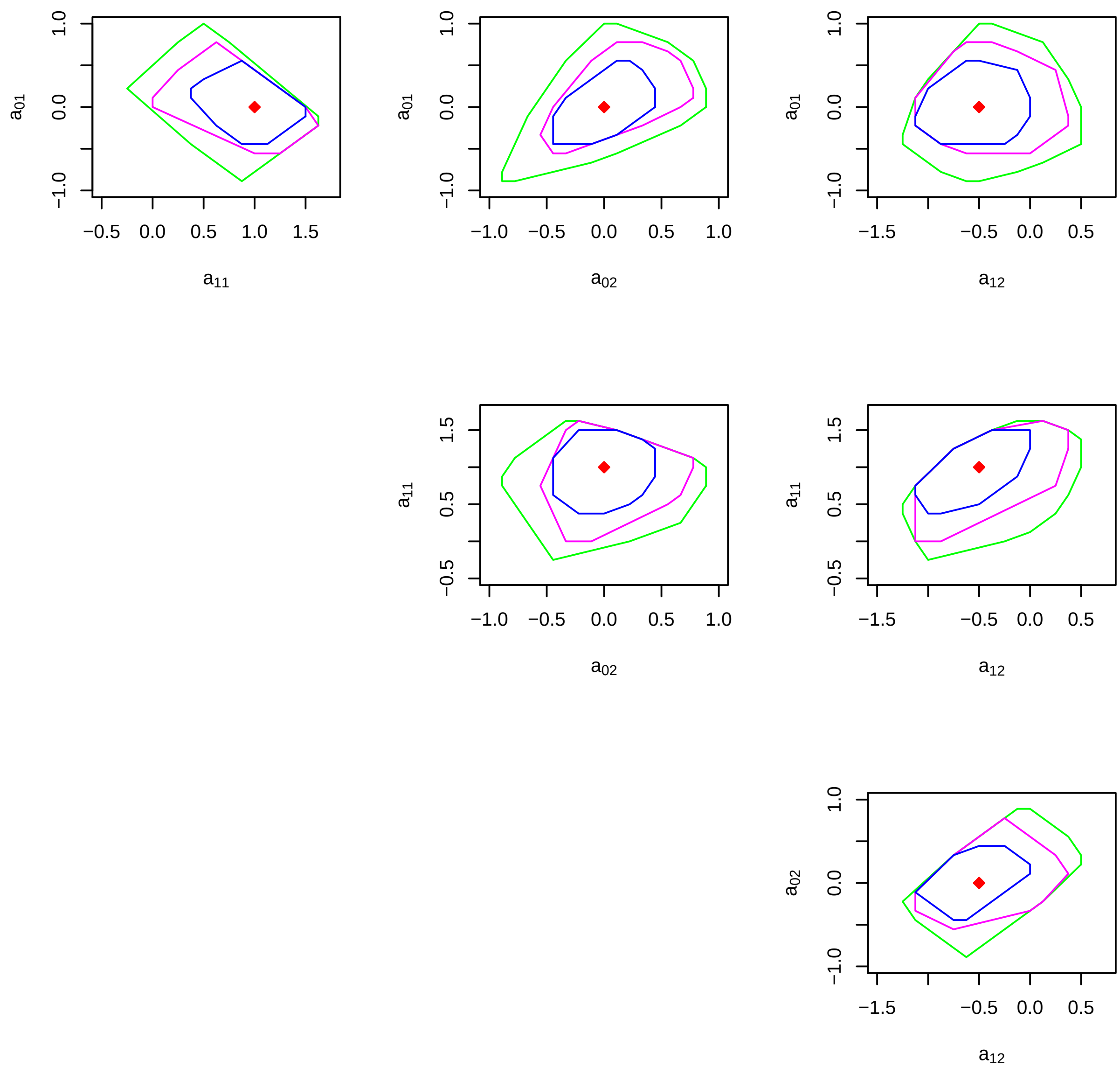
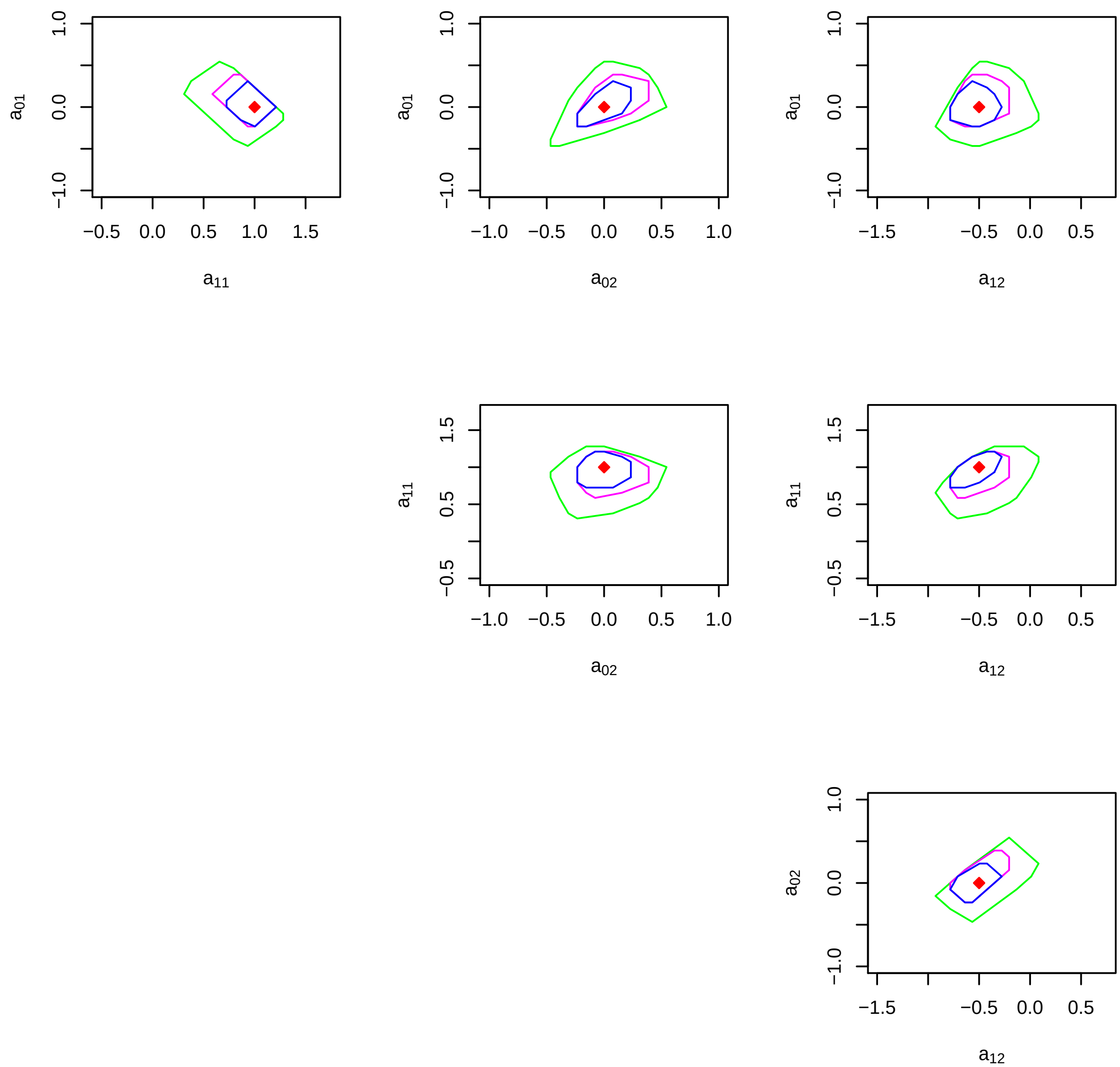


Figure 7: Case II.B. 2-D projections of the identified set and two outer regions, $M = 3$, $K = 4$, stronger instrument.



5 Conclusion

We have considered multiple discrete choice models with potentially endogenous explanatory variables and an instrumental variable (IV) restriction. The IV restriction requires that there exist variables that are excluded from the random utilities and distributed independently of the latent variables that induce stochastic variation in utilities. Our model does not rely on special regressor, large support, triangularity or control function restrictions. Nor does it require the existence of aggregate, e.g. market level, data. Indeed the model imposes quite minimal restrictions, being incomplete in the sense that the model is silent about the genesis of the potentially endogenous explanatory variables.

We have shown that this instrumental variable multiple discrete choice model has set identifying power and we have characterized the (sharp) identified set. The general characterization may involve a large number of inequalities. We have characterized a smaller collection of core-determining inequalities which in the context of any particular model serve to define the identified set, and we have provided an algorithm for calculating these in the case in which explanatory variables are discrete.

We also provide easy-to-compute outer regions that can further facilitate computation of the identified set. These may be of interest in their own right, potentially sufficient to address the qualitative economic questions pursued in some applications. In parametric models with discrete explanatory variables these only require calculation of probability expressions which appear in a conventional likelihood function and calculation of probabilities of the joint occurrence of values of the outcome and the explanatory variables conditional on the instrumental variables. This was demonstrated in the conditional logit model in Section 4, and in continuing work we are investigating the geometry of identified sets and outer regions in IV conditional probit and nested logit models.

A novel aspect of our results is that we have characterized the identifying power of an IV model which permits multiple unobservable variables in a structural function that delivers a discrete outcome. We develop a general approach to models of this sort in Chesher, Rosen, and Smolinski (2011), in which we extend the methods employed here to other IV models in which there are many unobservables in structural functions. Examples include random coefficient models that allow for general stochastic dependence between random coefficients and covariates with either continuous or discrete outcomes, and discrete choice models in which individuals' choices among alternatives need not be mutually exclusive.

References

Andrews, D. W. K., and X. Shi (2009): "Inference for Parameters Defined by Conditional Moment Inequalities," working paper, Cowles Foundation.

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

Ben-Akiva, M. (1973): "Structure of Passenger Travel Demand Models," MIT PhD Dissertation.

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.

Beresteanu, A., and F. Molinari (2008): "Asymptotic Properties for a Class of Partially Identified Models," *Econometrica*, 76(4), 763–814.

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.

——— (2004): "Differentiated Products Demand Systems from a Combination of Micro and Macro Data: The New Car Market," *Journal of Political Economy*, 112(1), 68–105.

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

Bugni, F. (2010): "Bootstrap Inference for Partially Identified Models Defined by Moment Inequalities: Coverage of the Identified Set," *Econometrica*, 78(2), 735–753.

Canay, I. (2010): "EL Inference for Partially Identified Models: Large Deviations Optimality and Bootstrap Validity," *Journal of Econometrics*, 156(2), 408–425.

Chernozhukov, V., and C. Hansen (2005): "An IV Model of Quantile Treatment Effects," *Econometrica*, 73, 245–261.

Chernozhukov, V., H. Hong, and E. Tamer (2007): "Estimation and Confidence Regions for Parameter Sets in Econometric Models," *Econometrica*, 75(5), 1243–1284.

Chernozhukov, V., S. Lee, and A. Rosen (2009): "Intersection Bounds, Estimation and Inference," CeMMAP working paper CWP19/09.

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

Chesher, A., and A. Rosen (2011): "Simultaneous Equations Models for Discrete Outcomes: Coherence, Completeness, and Identification," in preparation.

Chesher, A., A. Rosen, and K. Smolinski (2011): "Generalized Instrumental Variable Models," in preparation.

Chesher, A., and K. Smolinski (2010): "Sharp Identified Sets for Discrete Variable IV Models," CeMMAP working paper CWP11/10.

Chiappori, P.-A., I. Komunjer, and D. Kristensen (2011): "On the Nonparametric Identification and Estimation of Multiple Choice Models," working paper, Columbia University.

Choquet, G. (1954): "Theory of Capacities," *Annales de l'Institut Fourier*, 5, 135–295.

Domencich, T., and D. McFadden (1975): *Urban Travel Demand: A Behavioural Analysis*. North-Holland, Amsterdam.

Ekeland, I., A. Galichon, and M. Henry (2010): "Optimal Transportation and the Falsifiability of Incompletely Specified Economic Models," *Economic Theory*, 42, 355–374.

Fox, J. T., and A. Gandhi (2009): "Identifying Heterogeneity in Economic Choice Models," NBER working paper 15147.

Galichon, A., and M. Henry (2009): "A Test of Non-identifying Restrictions and Confidence Regions for Partially Identified Parameters," *Journal of Econometrics*, 152(2), 186–196.

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

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.

Kim, K. i. (2009): "Set Estimation and Inference with Models Characterized by Conditional Moment Inequalities," working paper, University of Minnesota.

Komarova, T. (2007): "Binary Choice Models with Discrete Regressors: Identification and Misspecification," working paper, LSE.

Koning, R., and G. Ridder (2003): "Discrete Choice and Stochastic Utility Maximization," *Econometrics Journal*, 6(1), 1–27.

Lewbel, A. (2000): "Semiparametric Qualitative Choice Models with Instrumental Variables and Unknown Heteroscedasticity," *Journal of Econometrics*, 97, 145–177.

Magnac, T., and E. Maurin (2008): "Partial Identification in Binary Models: Discrete Regressors and Interval Data," *Review of Economic Studies*, 75(3), 835–864.

Manski, C. F. (2007): "Partial Identification of Counterfactual Choice Probabilities," *International Economic Review*, 48(4), 1393–1410.

Manski, C. F., and E. Tamer (2002): "Inference on Regressions with Interval Data on a Regressor or Outcome," *Econometrica*, 70(2), 519–546.

Matzkin, R. (1993): "Nonparametric Identification and Estimation of Polychotomous Choice Models," *Journal of Econometrics*, 58, 137–168.

——— (2008): "Identification in Nonparametric Simultaneous Equations Models," *Econometrica*, 76, 945–978.

——— (2012): "Identification in Nonparametric Limited Dependent Variable Models with Simultaneity and Unobserved Heterogeneity," *Journal of Econometrics*, 166(1), 106–115.

McFadden, D. (1974): "Conditional Logit Analysis of Qualitative Choice Behavior," in *Frontiers in Econometrics*, ed. by P. Zarembka. New York: Academic Press.

——— (1978): "Modelling the Choice of Residential Location," in *Spatial Interaction Theory and Residential Location*, ed. by A. Karlvist, L. Ludvist, F. Snickars, and J. Weibull, pp. 75–96. North Holland, Amsterdam.

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.

Newey, W. K., and J. L. Powell (2003): "Instrumental Variable Estimation of Nonparametric Models," *Econometrica*, 71, 1565–1578.

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

Petrin, A., and K. Train (2010): "A Control Function Approach to Endogeneity in Consumer Choice Models," *Journal of Marketing Research*, 47, 1–11.

Romano, J. P., and A. M. Shaikh (2008): "Inference for Identifiable Parameters in Partially Identified Econometric Models," *Journal of Planning and Statistical Inference*, 138, 2786–2807.

Rosen, A. M. (2008): "Confidence Sets for Partially Identified Parameters that Satisfy a Finite Number of Moment Inequalities," *Journal of Econometrics*, 146, 107–117.

Sutherland, W. A. (2009): *Introduction to Metric and Topological Spaces*. Oxford Univerity Press, New York.