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 CWP06/11

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† Adam M. Rosen‡ Konrad Smolinski§

UCL and CeMMAP UCL and CeMMAP UCL and CeMMAP

February 11, 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, not point, identification of the latent utility functions and we characterize sharp bounds on those functions. We develop easy-to-compute outer regions which 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.

*We thank seminar participants at the University of Manchester, the September 2010 CIREQ conference on revealed preferences and partial identification, and the 21^{st} EC^2 conference held in Toulouse for comments and discussion. 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.

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 maximising framework set down in the ground-breaking 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 $(u_1, \ldots, u_M) \equiv u$ and the distribution of $V \equiv (V_1, \ldots, V_M)$ and functionals of these features.

In the set up considered by McFadden the explanatory variables X and unobservable utility shifters V are independently distributed. Our model relaxes this restriction, permitting X to be endogenous. 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 and Z is excluded from the utility functions $u_1, \ldots, u_M$. 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 identifcation in models with multiple unob-

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

servables 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 almost no restriction placed on the process generating the potentially endogenous explanatory variables.[2] In this sense the model is incomplete. Because of this incompleteness the model is partially identifying but 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 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 random variable, or random set, $\mathcal{T}(Y, X; u)$. A realization of one of these random sets, $\mathcal{T}(y, x; u)$, is the set of values of differences in random utility shifters, $W_y \equiv V_y - V_M$, $y \in \{1, \ldots, M-1\}$ 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 $\Pr[\mathcal{T}(Y, X; u) \subseteq \mathcal{S}|z]$ gives the probability conditional on instrumental variable $Z = z$ that $\mathcal{T}(Y, X; u)$ is a subset of the set $\mathcal{S}$.

We show that a utility function u and a distribution P_W of utility shifter differences, $W \equiv (W_1, \ldots W_{M-1})$ lies in the identified set associated with conditional distributions of Y and X given Z, $F^0_{YX|Z}$, if and only if

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

for almost every z in the support of Z and all compact sets $\mathcal{S} \subseteq \mathbb{R}^{M-1}$. Here $\Pr_0$ indicates probabilities taken with respect to $F^0_{YX|Z}$ and $P_W(\mathcal{S})$ is the probability mass the distribution P_W assigns to the set $\mathcal{S}$. By the "identified set" we mean the set comprising all and only admissible duples (u, P_W) which deliver the distributions $F^0_{YX|Z}$ for all z in the support of Z.[3]

We show that the only sets $\mathcal{S}$ that need to be considered when judging whether a particular pair (u, P_W) are in the identified set are unions of the sets on the support of $\mathcal{T}(Y, X; u)$ which are connected. When X is

[2] All that is required is that the support of X and V are independent.
[3] Some authors term this the "sharp identified set".

continuous there remains an infinite number of inequalities but when X is discrete there is a finite number and it can be computationally feasible to obtain a complete characterization of the identified set.

We develop characterizations of two outer regions within which the identified set is guaranteed to lie. One of these is particularly simple to calculate when X is discrete. Consider a model which specifies P_W^* as the distribution of W and specifies $p(y, x; u^*, P_W^*)$ as the probability $Y = y$ when $X = x$ and the utility functions u take the value u^*. For example in the classical conditional logit model with utility functions

$$u_y^*(x) = x'\beta_y^*$$

the probabilities involved are the following well known expressions.

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

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

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

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 discrete explanatory variables X are exogenous is able to calculate our outer region directly. Moreover in the conditional logit case the outer region is convex which simplifies computation.

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 special 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 control function approaches as in Petrin and Train (2010) and Fox and Gandhi (2009), the latter focusing on identification in a fully nonparametric setting, and also making use of large support conditions. Chiappori and Komunjer (2009) provide an alternative route to nonparametric identification, relying on conditional independence and completeness conditions quite distinct from the marginal independence restrictions imposed here. In limited dependent variables models with simultaneity, Matzkin (2009) 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 nonparametric identification under completeness conditions in such settings with the presence of special regressors, 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 as previously stated requires neither variables that differ across alternatives nor covariates with large support.

There are a number of 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. As done there, we remain agnostic as to the joint determination of covariates and instruments, but here 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 (2009), although the mechanism by which partial identification is obtained in these papers is 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 (2009) provide sharp bounds on the parameters of multinomial response model with interval data on regressors, demonstrating general identification results derived from random set theory.

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 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 3.3 we provide two easy-to-compute outer regions for the case in which the explanatory variables are discrete.

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

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

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

This additively separable form is used throughout. Define $U \equiv (U_1, \ldots U_M)$ and $V \equiv (V_1, ..., V_M)$.

The elements of X are observed variables with support $\mathcal{X}$. The elements of V are unobservable variables that capture heterogeneity in tastes across individuals. The model restricts V to be continuously distributed with positive density with respect to Lebesgue measure on all of $\mathbb{R}^M$.

The elements of Z are observable variables which the IV model excludes from the utility functions and requires to be jointly independently distributed with V.

Individuals are utility maximisers, observing the value of U and choosing the alternative that gives the highest utility as follows. For each $y \in \mathcal{Y}$,

$$Y = y \Leftrightarrow \left\{ \begin{array}{l} \forall y' \in \mathcal{Y},\, y' < y\colon\ U_y \geq U_{y'}, \\ \forall y' \in \mathcal{Y},\, y' > y\colon\ U_y > U_{y'}. \end{array} \right\}. \tag{2.2}$$

This formulation breaks ties between maximal utilities by settling on the alternative with the highest index. Since V is absolutely continuously distributed this is not a substantive restriction because ties will occur with probability zero. Thus simpler representations that ignore ties and specify the chosen alternative as the strict utility maximizer will be used in what follows.

Since the optimal selection of alternatives is entirely determined by utility differences the normalization: $u_M(x) = 0$ for all $x \in \mathcal{X}$ can be employed. The model is comprised of the following restrictions.

Restriction A1: (Y, X, Z, V) are defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, where $\mathcal{F}$ contains the Borel Sets on Ω, the support of Y is a finite set $\mathcal{Y} \equiv \{1, 2, ..., M\}$, and the support of (V, X, Z) is $\mathbb{R}^M \times \mathcal{X} \times \mathcal{Z}$.

Restriction A2: For each value $z \in \mathcal{Z}$ there is a proper 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.

Restriction A3: Given (V, X, Z), Y is determined uniquely by (2.1) and (2.2).

Restriction A4: V is a continuously distributed random variable on $\mathbb{R}^M$ with everywhere positive density with respect to Lebesgue measure and with distribution belonging to a specified family of distributions $\mathcal{P}_V$.

Restriction A5: The utility functions $u = \{u_1, ..., u_{M-1}\}$ belong to a specified family of functions $\mathcal{U}$.

Restriction A6: V and Z are stochastically independent.

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 distributions of Restriction A2. Throughout the notation "$\Pr_0$" will indicate probabilities calculated using these distributions.

Restriction A6 requires V and the variables Z to be independently distributed and Restriction A3 excludes these variables from the structural function so the variables Z are instrumental variables in the classical sense. Of course these restrictions have 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.

Alternative models place more or less restrictions on the family of distributions $\mathcal{P}_V$. Here are two examples.

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.

In order to specify the selection of alternatives as a function of utility differences define for each $y \in \mathcal{Y}$:

$$W_y \equiv V_y - V_M$$

and, with $W \equiv (W_1, ..., W_{M-1}) \in \mathbb{R}^{M-1}$, define:

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

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

$$Y = h\left(X,W;u\right)$$

with h defined as follows.

$$h\left(x,w;u\right) \equiv \sum_{y=1}^{M} y \times 1\left[\min_{k\in\mathcal{Y},k\neq y}\left(\Delta U_y(x,w) - \Delta U_k(x,w)\right) > 0\right] \tag{2.4}$$

Because the dependence of the structural function $h(X,W;u)$ on the utility functions listed in u is crucial it is made explicit in the notation.

The model requires the random components of utility, V, to have a distribution in the family $\mathcal{P}_V$. Let $\mathcal{P}_W$ denote the corresponding family of probability distributions for the random utility *differences*, W.

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.

3 Identification

3.1 The identified set

We now develop results on the identifying power of the IV model of multiple discrete choice. Structures admitted by the model are characterized by a duple, $D \equiv (u, P_W)$, comprising a list of utility functions, u, and a distribution of differences in random utility shifters, P_W. It is shown that the IV model set identifies D. In general there are many admissible duples D that can generate a particular set of conditional distributions $F^0_{YX|Z}$ for $z \in \mathcal{Z}$. We develop a system of inequalities which characterize the identified set of duples.

Key in what follows are the sets of values of the unobservable variables W that, for a particular list of utility functions, u, deliver the value y of Y when $X = x$.

$$\mathcal{T}(y, x; u) \equiv \{w : h(x, w; u) = y\}$$

Note that for any admissible u and each value x, the sets $\mathcal{T}(y, x; u)$, $y \in \mathcal{Y}$, partition the support of W which is $\mathbb{R}^{M-1}$. These sets are illustrated for particular structural functions in Section 4.

Let $\mathsf{K}\left(\mathbb{R}^{M-1}\right)$ denote the collection of all compact subsets of $\mathbb{R}^{M-1}$.[4] Consider a probability distribution $P_W \in \mathcal{P}_W$ and for any $\mathcal{S} \in \mathsf{K}\left(\mathbb{R}^{M-1}\right)$ let $P_W(\mathcal{S})$ denote the probability under P_W of the event $\{W \in \mathcal{S}\}$.

Consider a family of conditional distributions $P_{W|XZ}$ for $z \in \mathcal{Z}$ and for any $\mathcal{S} \in \mathsf{K}\left(\mathbb{R}^{M-1}\right)$ let $P_{W|XZ}(\mathcal{S}|x, z)$ denote the associated conditional probability of the event $\{W \in \mathcal{S}\}$ 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, A6.

- **Independence**: The IV model requires V and Z to be independently distributed so W and Z must be independently distributed. It follows that for a choice $P_W \in \mathcal{P}_W$ all associated conditional distrib-

[4] 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.

utions $P_{W|XZ}$ that (i) are admitted by the IV model and (ii) can generate the particular probability distributions of Restriction 2 must satisfy the condition

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

for all values $z \in \mathcal{Z}$ and sets $\mathcal{S} \in \mathsf{K}\left(\mathbb{R}^{M-1}\right)$. The left hand side of (3.1) is the conditional probability $P_{W|Z}(S|z)$ 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 $P_W \in \mathcal{P}_W$ must support if they are to be capable of delivering the probability distributions of Restriction 2.

- **Observational equivalence**. Since for any value, x, of X the utility functions u deliver $Y = y$ if and only if $W \in \mathcal{T}(y, x; u)$, there is the requirement that, associated with P_W, there are conditional distributions $P_{W|XZ}$ such that for all $y \in \mathcal{Y}$, $x \in \mathcal{X}$ and $z \in \mathcal{Z}$:

$$P_{W|XZ}(\mathcal{T}(y, x; u)|x, z) = \Pr_0[Y = y|X = x, Z = z]. \tag{3.2}$$

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 2, that is all duples in the identified set associated with $F^0_{YX|Z}$ for $z \in \mathcal{Z}$. Let the identified set be denoted by $\mathcal{D}^0(\mathcal{Z})$. This system of inequalities is now derived.

Considering any compact set $\mathcal{S} \in \mathsf{K}\left(\mathbb{R}^{M-1}\right)$, equation (3.2) places restrictions on $P_{W|XZ}(\mathcal{S}|x, z)$ and the utility functions u associated with duples in $\mathcal{D}^0(\mathcal{Z})$.

First, if (3.2) is to be satisfied then for any set $\mathcal{S}$, the *smallest* value that $P_{W|XZ}(\mathcal{S}|x, z)$ can take is equal to the sum of the probabilities $\Pr_0[Y = y|X = x, Z = z]$ associated with all sets $\mathcal{T}(y, x; u)$ contained *entirely within* $\mathcal{S}$. This is expressed in the inequality

$$P_{W|XZ}(\mathcal{S}|x, z) \geq \sum_{y\in\mathcal{Y}} 1[\mathcal{T}(y, x; u) \subseteq \mathcal{S}] \Pr_0[Y = y|X = x, Z = z] \tag{3.3}$$

which holds for all $x \in \mathcal{X}$ and $z \in \mathcal{Z}$.

Second, for any compact set $\mathcal{S}$, the *largest* value that $P_{W|XZ}(\mathcal{S}|x, z)$ can take is equal to the sum of the probabilities $\Pr_0[Y = y|X = x, Z = z]$ associated with all sets $\mathcal{T}(y, x; u)$ that have a non-null intersection with $\mathcal{S}$. This is expressed in the following inequality which holds for all $x \in \mathcal{X}$ and $z \in \mathcal{Z}$. The symbol ϕ denotes the empty set.

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

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.

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

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

All duples (u, P_W) in the identified set $\mathcal{D}^0(\mathcal{Z})$ satisfy these inequalities for all $z \in \mathcal{Z}$ and all $\mathcal{S} \in \mathsf{K}\left(\mathbb{R}^{M-1}\right)$. So the inequalities (3.5) and (3.6) obtained as $\mathcal{S}$ passes across all sets in $\mathsf{K}\left(\mathbb{R}^{M-1}\right)$ 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 $\mathcal{S} \in \mathsf{K}\left(\mathbb{R}^{M-1}\right)$ and for any such S, given a choice $P_W \in \mathcal{P}_W$ the left hand sides of the inequalities can be calculated. We will shortly show that the system of inequalities define the sharp 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 (2009) and Galichon and Henry (2009).

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

$$\mathcal{T}(Y, x; u) \equiv \{w : h(x, w; u) = Y\},$$

and

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

which are random closed sets on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ of Restriction A1.[5]

[5] 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 $\{\mathcal{T}(Y, x; u) \cap S \neq \phi\}$ and $\{\mathcal{T}(Y, X; u) \cap S \neq \phi\}$ are $\mathcal{F}$-measurable. For a formal definition of random closed sets see e.g. Molchanov (2005) or Beresteanu, Molchanov, and Molinari (2010) Appendix A.

Probability distributions of random sets are characterized either by containment functionals or by capacity functionals, see Molchanov (2005). The containment and capacity functionals of $\mathcal{T}(Y, x; u)$ conditional on $X = x$ and $Z = z$ under the particular probability distributions of Restriction 2 are respectively

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

and

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

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

Similarly the containment and capacity functionals of $\mathcal{T}(Y, X; u)$ conditional on $Z = z$ *alone*, under the particular probability distributions of Restriction 2 are respectively

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

and

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

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

It follows that all admissible duples (u, P_W) with probability distributions $P_W \in \mathcal{P}_W$ and utility functions $u \in \mathcal{U}$ that deliver the particular distributions in Restriction 2 satisfy the inequalities:

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

for all sets $\mathcal{S} \in \mathsf{K}\left(\mathbb{R}^{M-1}\right)$ and instrumental values $z \in \mathcal{Z}$.

Capacity and containment functionals are equivalent characterizations of the distribution of a random set because for all $\mathcal{S} \in \mathsf{K}\left(\mathbb{R}^{M-1}\right)$ and $z \in \mathcal{Z}$,

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

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* sets in $\mathsf{K}\left(\mathbb{R}^{M-1}\right)$ 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 containment functional of $\mathcal{T}(Y,X;u)$.

The following Theorem states that all and only duples (u,P_W) which satisfy the system of inequalities generated by the lower bound in (3.7) for all $z\in\mathcal{Z}$ and all $\mathcal{S}\in\mathsf{K}\left(\mathbb{R}^{M-1}\right)$ deliver the distributions of Restriction 2, that is that the system of inequalities defines the identified set of duples.

Theorem 1 *The identified set of admissible duples (u,P_W) associated with the conditional distributions $F^0_{YX|Z}$, $z\in\mathcal{Z}$, is*

$$\mathcal{D}^0(\mathcal{Z})\equiv\left\{(u,P_W):u\in\mathcal{U},\ P_W\in\mathcal{P},\ s.t.\ \ \Pr_0\left[\mathcal{T}(Y,X;u)\subseteq\mathcal{S}|Z=z\right]\leq P_W(\mathcal{S}),\ \forall\mathcal{S}\in\mathsf{K}\left(\mathbb{R}^{M-1}\right)\ \ a.e.\ z\in\mathcal{Z}\right\}$$

where $\mathsf{K}\left(\mathbb{R}^{M-1}\right)$ denotes the set of all compact subsets of $\mathbb{R}^{M-1}$.

Proof. *$\mathcal{D}^0(\mathcal{Z})$ contains all duples (u,P_W) with $u\in\mathcal{U}$ and $P_W\in\mathcal{P}$ such that for all $\mathcal{S}\in\mathsf{K}\left(\mathbb{R}^{M-1}\right)$, $\Pr_0\left[\mathcal{T}(Y,X;u)\subseteq\mathcal{S}|Z=z\right]\leq P_W(\mathcal{S})$ 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 provided by Artstein (1983) and Norberg (1992), Artstein's inequality, guarantees sharpness, see also Molchanov (2005) Section 1.4.8. To see why consider any $(u,P_W)\in\mathcal{D}^0(\mathcal{Z})$ and fix $z\in\mathcal{Z}$. Then with probability one we have that*

$$\Pr_0\left[\mathcal{T}(Y,X;u)\subseteq\mathcal{S}|Z=z\right]\leq P_W(\mathcal{S}),\ \forall\mathcal{S}\in\mathsf{K}\left(\mathbb{R}^{M-1}\right),$$

by definition of $\mathcal{D}^0(\mathcal{Z})$. This is equivalent to

$$\Pr_0\left[\mathcal{T}(Y,X;u)\cap\mathcal{S}\neq\phi|Z=z\right]\geq P_W(\mathcal{S}),\ \forall\mathcal{S}\in\mathsf{K}\left(\mathbb{R}^{M-1}\right),$$

see Beresteanu, Molchanov, and Molinari (2010, Theorem 2.1) for a formal proof of the equivalence. Then by Artstein (1983) and Norberg (1992) it follows that there exists a random variable $\tilde{W}$ and a random set $\tilde{\mathcal{T}}$ realized on the same probability space as $(W,\mathcal{T}(Y,X;u))$ such that conditional on $Z=z$, both $\tilde{W}\sim P_W$ and $\tilde{\mathcal{T}}\overset{d}{=}\mathcal{T}(Y,X;u)$ with $\tilde{W}\in\tilde{\mathcal{T}}$ 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 $\tilde{W}\in\mathcal{T}(\tilde{Y},\tilde{X};u)$ and $\left(\tilde{Y},\tilde{X}\right)\overset{d}{=}(Y,X)$ with probability one given $Z=z$. 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.* ■

Remarks

1. Key to the proof of sharpness is the result from random set theory that for any random set $\mathcal{T}$ and any random variable $W \in \mathbb{R}^{M-1}$ such that

$$\Pr\left[\mathcal{T} \cap \mathcal{S} \neq \phi\right] \geq P_W(\mathcal{S}), \forall \mathcal{S} \in \mathsf{K}\left(\mathbb{R}^{M-1}\right),$$

   we can couple with W and $\mathcal{T}$ a random variable $\tilde{W}$ and a random set $\tilde{\mathcal{T}}$, respectively, living on the same probability space and with the same distributions as the original random variable W and random set $\mathcal{T}$, such that $\tilde{W} \in \tilde{\mathcal{T}}$ with probability one. Our proof makes use of the existence of such a couple conditional on each instrumental value $z \in \mathcal{Z}$ to show that every duple (u, P_W) 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:

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

   can be replaced by the capacity functional inequality:

$$\Pr_0\left[\mathcal{T}(Y, X; u) \cap \mathcal{S} \neq \phi | Z = z\right] \geq P_W(\mathcal{S}).$$

3. The inequalities of Theorem 1 are required to hold for almost every $z \in \mathcal{Z}$ so for each $\mathcal{S} \in \mathsf{K}\left(\mathbb{R}^{M-1}\right)$ 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 $\mathcal{P}_W$. When there are parametric restrictions these classes of functions are indexed by a finite dimensional parameter.

5. When X is exogenous (that is when X and V are stochastically independent) we can proceed as if $X = Z$, and (3.7) simplifies as follows.

$$\Pr_0\left[\mathcal{T}(Y, X; u) \subseteq \mathcal{S} | X = x\right] \leq P_W(S) \leq \Pr_0\left[\mathcal{T}(Y, X; u) \cap \mathcal{S} \neq \phi | X = x\right] \tag{3.9}$$

The upper and lower bounds are equal for sets S on the support of $\mathcal{T}(Y,X;u)$ because for all (y,x) on the support of (Y,X) and any $u \in \mathcal{U}$ there is the following.

$$\Pr_0[\mathcal{T}(Y,X;u) \subseteq \mathcal{T}(y,x;u)|X=x] = \Pr_0[Y=y|X=x] = \Pr_0[\mathcal{T}(Y,X;u) \cap \mathcal{T}(y,x;u) \neq \phi|X=x]$$

So for such sets the inequalities (3.7) become

$$\Pr_0[Y=y|X=x] = P_W[\mathcal{T}(y,x;u)]$$

which hold for $(y,x) \in \mathcal{Y} \times \mathcal{X}$ and with sufficient restrictions on $\mathcal{U}$ and $\mathcal{P}_W$ there may be point identification of u and P_W.

3.2 Core determining inequalities

It is often infeasible to consider the complete system of inequalities of Theorem 1 that are generated as $\mathcal{S}$ passes through *all* compact subsets of $\mathbb{R}^{M-1}$. 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 small collection of the sets in $\mathsf{K}\left(\mathbb{R}^{M-1}\right)$ whose inequalities define $\mathcal{D}^0(\mathcal{Z})$. This is a *core-determining class* of inequalities as studied by Galichon and Henry (2009) in obtaining identification regions in incomplete models.

The result of Theorem 2 below is useful in producing the sets that deliver core determining classes of inequalities for the models considered in this paper. We call these collections of sets *core determining sets* in what follows. The proof makes use of the following Lemma.

Lemma 1 *For the model defined by Restrictions A1-A6 the sets on the support of $\mathcal{T}(Y,X;u)$ are connected.*

Proof. *The sets are convex because each set $T(y,x;u)$ on the support of $\mathcal{T}(Y,X;u)$ is an intersection of linear half spaces as follows.*

$$T(y,x;u) = \{W : W_y - W_{y'} > u_{y'} - u_y, \quad \forall y' \neq y \in \mathcal{Y}\}$$

Since the sets are convex they are connected. ■

Theorem 2 *The identified set of Theorem 1 is defined by the inequalities generated by sets $\mathcal{S}$ which are (i) connected and (ii) unions of sets on the support of $\mathcal{T}(Y,X;u)$.*

Proof. *(i). Consider a set $\mathcal{S}$ which is the union of disjoint connected sets, S_j, $j \in \{1,\ldots,J\}$. Because the component sets are disjoint and, by Lemma 1 the sets on the support of $\mathcal{T}(Y,X;u)$ are connected, the lower bounding probability of the inequality of Theorem 1 is additive across the component connected sets.*

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

This is so because each of the connected sets $\mathcal{T}(y,x;u)$ is a subset of $\mathcal{S}$ if and only if it is a subset of one of the disjoint sets $\mathcal{S}_j$. Because the component sets that make up S are disjoint the probability assigned to $\mathcal{S}$ by any measure P_W is additive:

$$P_W(\mathcal{S}) = \sum_{j=1}^{J} P_W(\mathcal{S}_j).$$

Therefore:

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

implies

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

and so:

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

It follows that, if the inequalities of Theorem 1 hold for all connected sets $\mathcal{S} \in \mathsf{K}\left(\mathbb{R}^{M-1}\right)$ then they hold for all $\mathcal{S} \in \mathsf{K}\left(\mathbb{R}^{M-1}\right)$.[6]

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

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

[6] Here we have made use of our definition of the sets $\mathcal{T}(y,x;u)$ on the support of $\mathcal{T}(Y,X;u)$ as open sets, which follows from our use of strict inequalities in the definition of $h(y,x;u)$ in (2.4). This guarantees that two sets on the support of $\mathcal{T}(Y,X;u)$ that share only a common boundary are not connected, since neither contains their boundary. This can be amended to accommodate alternative definitions of $\mathcal{T}(y,x;u)$ that may contain boundary points, for example by explicitly working with their interior, which corresponds to our definition and only differs on a set of measure zero P_W.

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

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

It follows that if the inequalities of Theorem 1 hold for all unions of the sets on the support of $T(Y,X;u)$ *then they hold for all sets* $\mathcal{S} \in \mathsf{K}(\mathbb{R}^{M-1})$, *since then*

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

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

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_W. Let the support of discrete X be $\mathcal{X} \equiv \{x_1, \ldots, x_K\}$. X may be a finite dimensional vector.

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.[7] Let $\mathsf{C}_1 \| \mathsf{C}_2$ denote the collection of the sets that appear either in C_1 or in C_2.[8] Let $\mathsf{C}(u)$ denote the MK sets on the support of $T(Y,X;u)$. 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.

    (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 $\mathsf{G}(u) \| \mathsf{G}^*(u)$.

[7] 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

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

[8] 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.

Let Con(·) applied to a list of sets select the connected sets in the list. The algorithm recursively creates the following list of sets.

$$\mathsf{C}(u) \| \operatorname{Con}\left(\mathsf{C}(u) \otimes \mathsf{C}(u)\right) \| \operatorname{Con}\left(\operatorname{Con}\left(\mathsf{C}(u) \otimes \mathsf{C}(u)\right) \otimes \mathsf{C}(u)\right) \| \cdots$$

This is the same as the list

$$\operatorname{Con}\left(\mathsf{C}(u) \| \mathsf{C}(u) \otimes \mathsf{C}(u) \| \mathsf{C}(u) \otimes \mathsf{C}(u) \otimes \mathsf{C}(u) \| \cdots\right)$$

which is evidently the list of all connected unions of sets on the support of $\mathcal{T}(Y, X; u)$ as required by Theorem 2, but is more efficient computationally. The algorithm terminates in at most $MK - 1$ iterations.

The algorithm we use to produce tables later in the paper 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 $\operatorname{Con}\left(\mathsf{C}(u) \otimes \mathsf{C}(u)\right)$ is compared with every subsequent element in the list and elements in $\operatorname{Con}\left(\mathsf{C}(u) \otimes \mathsf{C}(u)\right)$ 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:

$$\sum_{\{(y,x):\mathcal{T}(y,x;u)\in\mathsf{C}_{\mathcal{G}}(u)\}} \Pr_0[Y = y \wedge X = x | Z = z].$$

The number of core determining sets is far smaller than the number of possible unions of sets on the support of $T(Y, X; u)$. For example in a 3 choice model with a binary explanatory variable, 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 $T(Y, X; u)$. 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.

Number of points of support of X	Number of core determining sets
2	12
3	33
4	82
5	188
6	406
7	842

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.

3.3 An easy-to-compute outer region

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

$$P_W[\mathcal{T}(y, x; u)] \geq \text{Pr}_0[Y = y \wedge X = x | Z = z]$$

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

$$P_W[\mathcal{T}(y, x; u)] \geq \max_{z \in \mathcal{Z}} \text{Pr}_0[Y = y \wedge X = x | Z = z] \tag{3.10}$$

must hold for all $(y, x) \in \mathcal{Y} \times \mathcal{X}$. The MK inequalities (3.10) define an *outer region* within which lies the identified set of duples (u, P_W).

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

$$P_W[\mathcal{T}(y, x; u)] = \frac{\exp(u_y(x))}{1 + \sum_{y'=1}^{M-1} \exp(u_{y'}(x))}, \qquad y \in \{1, \ldots, M\}. \tag{3.11}$$

In general the probability $P_W[\mathcal{T}(y, x; u)]$ is the probability that would appear in a classical discrete choice likelihood function (for independent realisations) constructed using (u, P_W) and defined by conditioning on observed values of the explanatory variables X as if they were exogenous.

For all (u, P_W) in the identified set the inequalities (3.10) require that the probability $P_W[\mathcal{T}(y, x; u)]$ 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 a list L containing one or more index pairs (y', x') such that $\mathcal{T}(y', x'; u) \subseteq \mathcal{T}(y'', x''; u)$ because in such cases the containment functional inequality requires:

$$P_W[\mathcal{T}(y'', x''; u)] \geq \Pr_0[Y = y'' \wedge X = x'' | Z = z] + \sum_{(y',x') \in L} \Pr_0[Y = y' \wedge X = x' | Z = z].$$

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 $\mathcal{T}(Y, X; u)$. In general there are many cases in which such improvements can be obtained.

4 Illustration: Three choice models

4.1 Core determining sets

In this Section we provide examples of identified sets, focusing on models for choice among $M = 3$ alternatives and the case in which X is discrete with finite support $\mathcal{X} \equiv \{x_1, \ldots, x_K\}$. In this case we can give a graphical display of the support of the set valued random variable $\mathcal{T}(Y, X; u)$. 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:

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

for $x \in \mathcal{X}$. 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 situation with two values x_1 and x_2 of X is drawn in Figure 1. Values of W_1 are measured vertically and values of W_2 are measured horizontally. Sets $\mathcal{T}(1, x; u)$, $\mathcal{T}(2, x; u)$ and $\mathcal{T}(3, x; u)$ lie respectively northwest, southeast and southwest of the point $w(x)$.[9]

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 $K = 2$ there are 6 cases which can be grouped into 3 pairs distinguished by the *slope* of the line connecting $w(x_1)$ and $w(x_2)$: (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 $w(x_1)$ 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 6 orientations to be considered as in the case in which $K = 2$.

The three orientations with $w(x_2)$ above $w(x_1)$ are shown in Figures 1 - 3.[10] 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 3.3, in each case two of the six sets in $\mathsf{C}(u)$ are subsets of others. For example, in Case 1a $\mathcal{T}(1, x_2; u) \subseteq \mathcal{T}(1, x_1; u)$ and $\mathcal{T}(2, x_1; u) \subseteq \mathcal{T}(2, x_2; u)$ (see Figure 1) and, as explained earlier, our algorithm includes the "supersets"

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

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

[10] At the end of the paper.

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 T(Y,X;u) that appear in the unions generating the 12 core determining inequalities, M=3, K=2, Case 1a, 2a and 3a.

and

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

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

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 $\mathcal{X} = \{x_1, \ldots, x_K\}$ 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.

	Support	Unions											
Case	set	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 T(Y,X;u) that appear in the unions generating the 12 core determining inequalities, M=3, K=2, Case 1b, 2b and 3b.

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. A computational advantage of this choice of distribution is that probabilities can be calculated without using numerical integration methods.

In our initial 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[11]

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

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_{is} and w_{2t} by m_{st}. The probability mass placed by P_W on a set $S \subseteq [0, 1]^2$ is approximated by

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

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.

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

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

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

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

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

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 $\tilde{\mathcal{T}}(1, x; u)$ and $\tilde{\mathcal{T}}(2, x; u)$:

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

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)

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

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 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.[12] 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:

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

   implied by (3.10). 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

[12] We draw convex hulls of points calculated to lie in each 2-D set that is graphed.

are as in (4.2) but the right hand sides take account of the existence of any pairs (y', x') and (y'', 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 (y', x') and (y'', 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.

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. In a travel demand context these may be variables that influence choice of residential location but have no other role in determining propensities to travel by alternative transport modes.

We have shown that this instrumental variable multiple discrete choice model has set identifying power and we have characterized the sharp identified set. The characterization may involve an extremely large number of inequalities. We have characterized a smaller collection of core determining inequalities and we have provided an algorithm for calculating these in the case in which explanatory variables are discrete.

Easy-to-compute outer regions are available. 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.

A novel aspect of our results is that we have characterized the identifying power of an instrumental variable model which permits multiple unobservable variables in the structural function that delivers a discrete outcome. We develop a general approach to models of this sort in Chesher, Rosen, and Smolinski (2011).

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.

Topics being studied in ongoing research include the following.

1. The sensitivity of the identified set and outer regions to variations in the strength and support of instruments and the support of the endogenous variable.

2. Non-convexity and connectedness of identified sets when instruments are weak.

3. The geometry of identified sets and outer regions in the conditional probit model and in the nested

logit model.

4. Identification of random choice functions in nonparametric models.

5. The application of the methods employed here to other cases in which there are many unobservables in structural functions, for example discrete choice models with "random coefficients".

References

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 (2009): "Sharp Identification Regions in Models with Convex Predictions: Games, Individual Choice, and Incomplete Data," CEMMAP working paper CWP27/09.

——— (2010): "Partial Identification Using Random Set Theory," CEMMAP working paper CWP40/10.

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.

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

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

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., and I. Komunjer (2009): "On the Nonparametric Identification of Multiple Choice Models," working paper, Columbia University.

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

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

Galichon, A., and M. Henry (2009): "Set Identification in Models with Multiple Equilibria," working paper, Montreal and Ecoles Polytechnique.

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.

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.

Manski, C. F. (2007): *Identification for Prediction and Decision*. Harvard University Press, Cambridge, MA.

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.

——— (2009): "Identification in Nonparametric Limited Dependent Variable Models with Simultaneity and Unobserved Heterogeneity," working paper, UCLA.

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. North Holland, Amsterdam.

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.

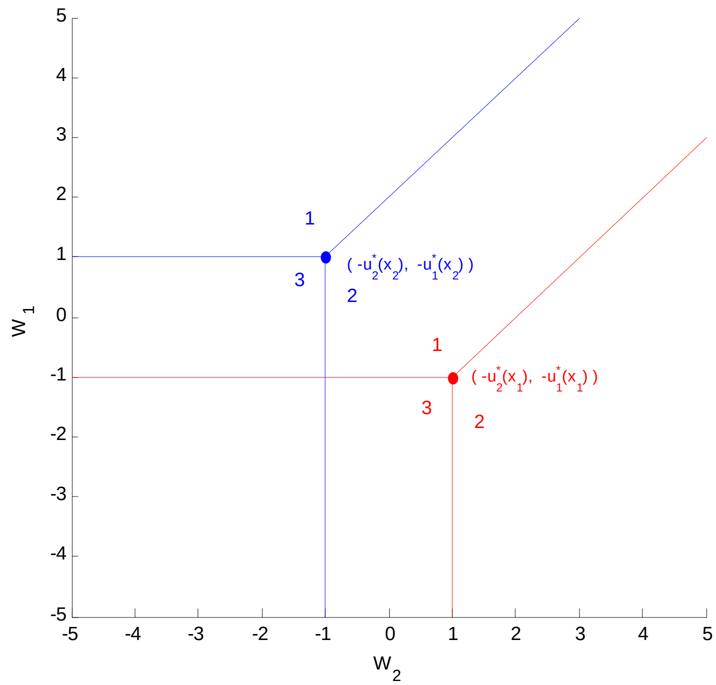

Figure 1: Support of $T(Y,X;u)$, $Y \in \{1,2,3\}$, $X \in \{x_1,x_2\}$. Case 1a.

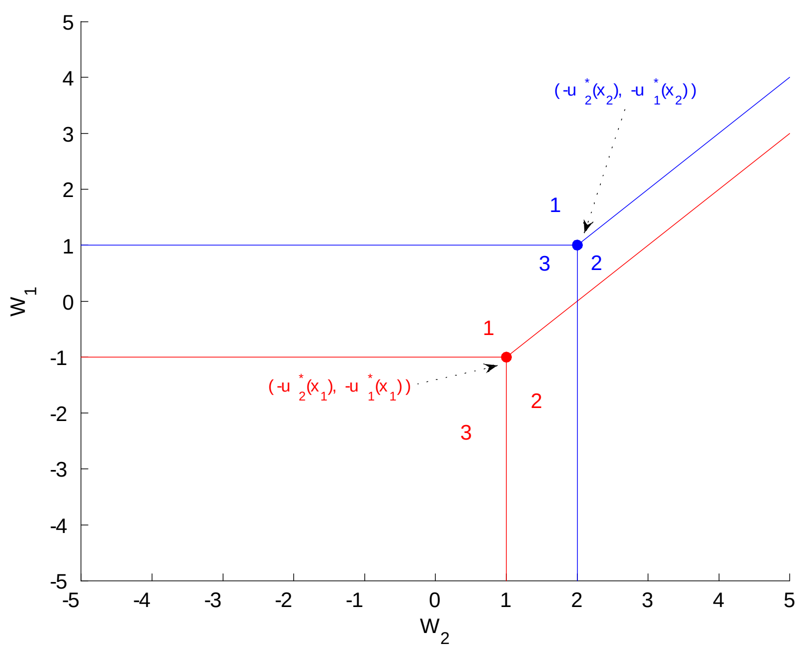

Figure 2: Support of $T(Y,X;u)$, $Y \in \{1,2,3\}$, $X \in \{x_1,x_2\}$. Case 2a.

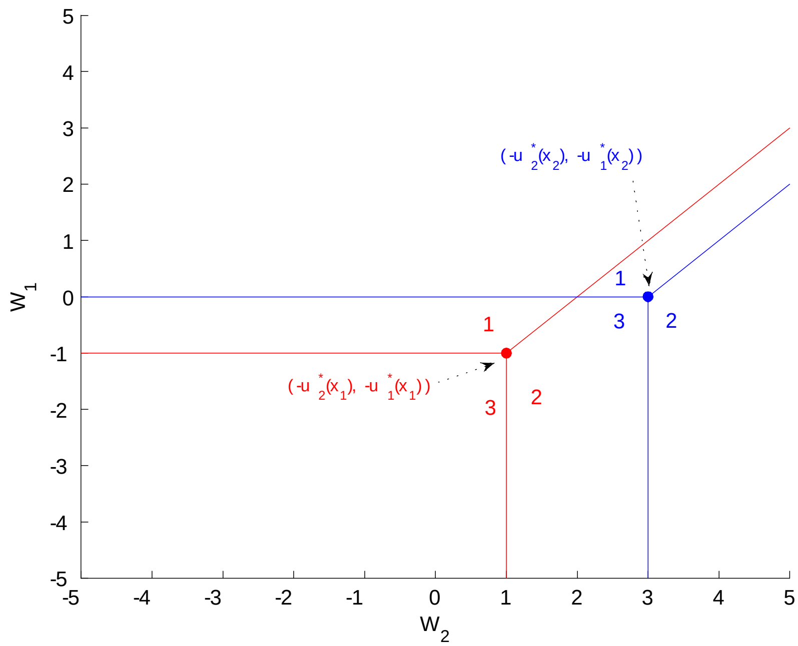

Figure 3: Support of $T(Y, X; u)$, $Y \in \{1, 2, 3\}$, $X \in \{x_1, x_2\}$. Case 3a.

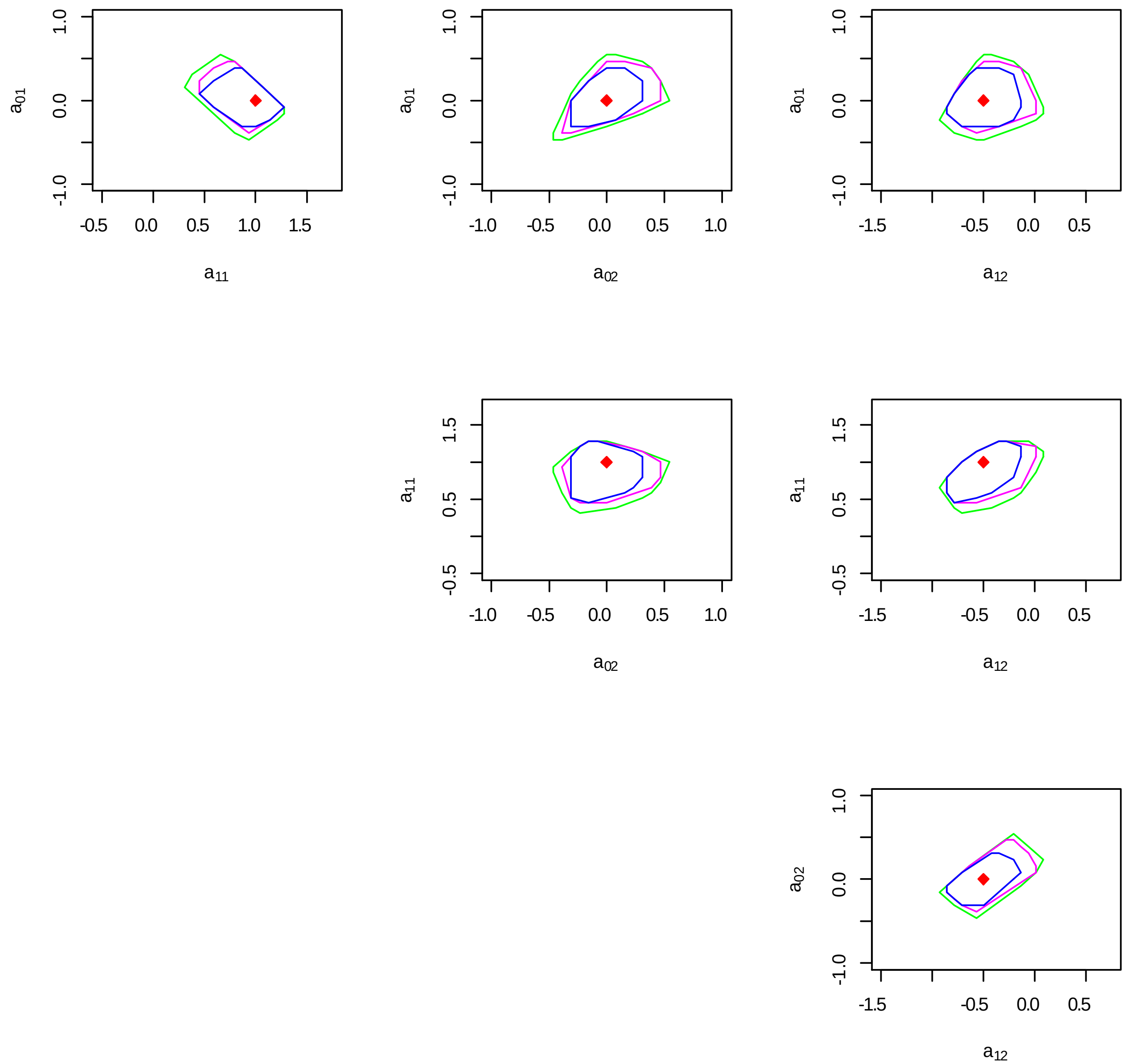

Figure 4: Case I.A. 2-D projections of the identified set and two outer regions, $M = 3$, $K = 2$, weaker instrument.

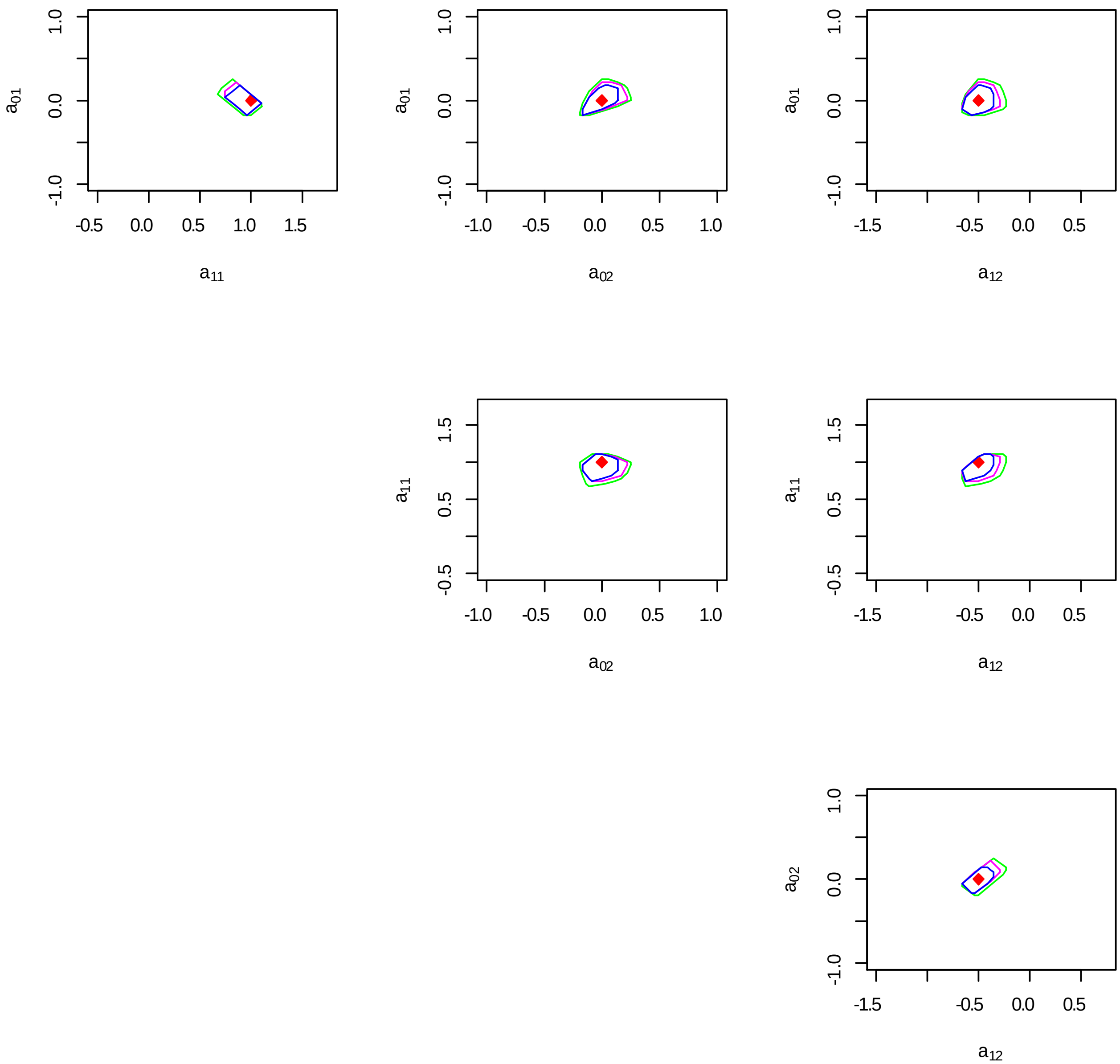

Figure 5: Case I.B. 2-D projections of the identified set and two outer regions, $M = 3$, $K = 2$, stronger instrument.

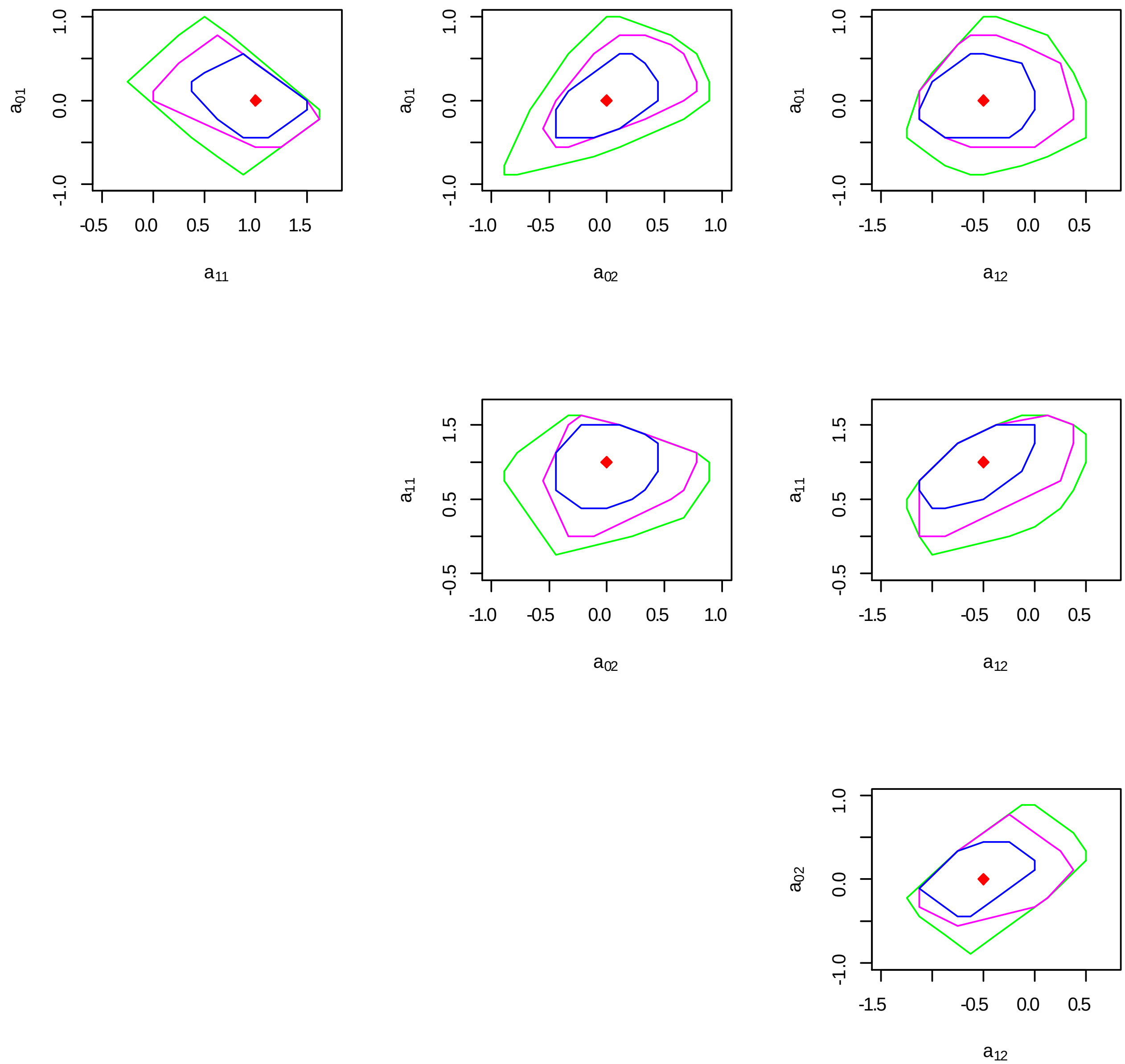

Figure 6: Case II.A. 2-D projections of the identified set and two outer regions, $M = 3$, $K = 4$, weaker instrument.

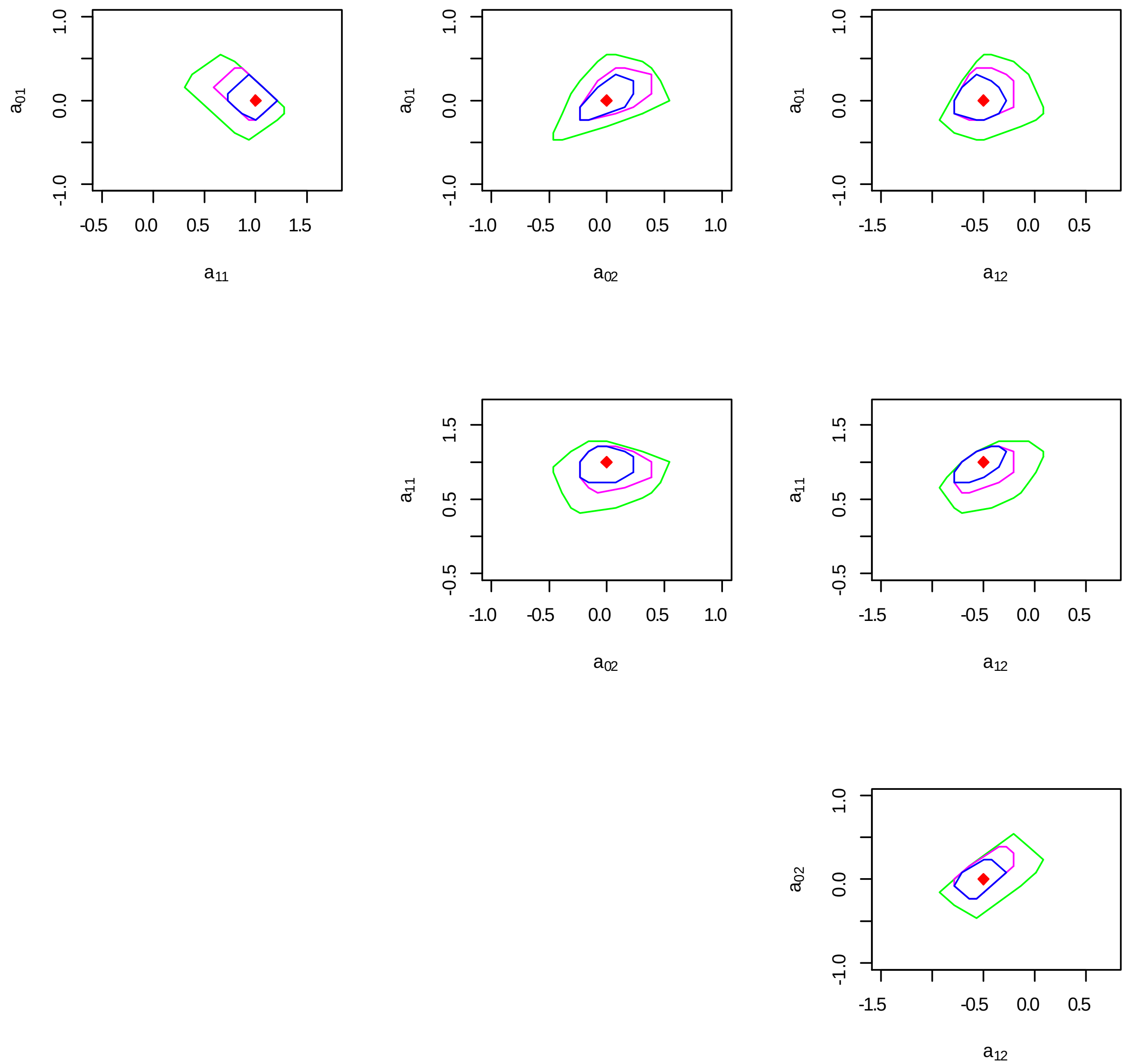

Figure 7: Case II.B. 2-D projections of the identified set and two outer regions, $M = 3$, $K = 4$, stronger instrument.