Quantitative Economics 4 (2013), 157–196 1759-7331/20130157

An instrumental variable model of multiple discrete choice

Andrew Chesher
Department of Economics, University College London and CeMMAP

Adam M. Rosen
Department of Economics, University College London and CeMMAP

Konrad Smolinski
CeMMAP and Institute for Fiscal Studies

This paper studies identification 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 that delivers values of endogenous explanatory variables, and in this respect it is incomplete. Instead, the model employs instrumental variable restrictions that require the existence of instrumental variables that are excluded from latent utilities and distributed independently of the unobserved components of utilities.

We show that the model delivers set identification of latent utility functions and the distribution of unobserved heterogeneity, and we characterize sharp bounds on these objects. We develop easy-to-compute outer regions that, in parametric models, require little more calculation than what is involved in a conventional maximum likelihood analysis. The results are illustrated using a model that is essentially the conditional logit model of McFadden (1974), but with potentially endogenous explanatory variables and instrumental variable restrictions.

Andrew Chesher: andrew.chesher@ucl.ac.uk
Adam M. Rosen: adam.rosen@ucl.ac.uk
Konrad Smolinski: konrad_s@ifs.org.uk

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, the University of Manchester, and Vanderbilt, as well as audiences at the September 2010 CIREQ Conference on Revealed Preferences and Partial Identification, the 21st 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 thank three anonymous referees and the co-editor for comments that have led to substantial improvements. We are grateful to 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.

Copyright © 2013 Andrew Chesher, Adam M. Rosen, and Konrad Smolinski. Licensed under the Creative Commons Attribution-NonCommercial License 3.0. Available at http://www.qeconomics.org.
DOI: 10.3982/QE240

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.

JEL classification. C25, C26.

1. Introduction

This paper develops results on the identification of features of models of choice among 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 ground-breaking work of McFadden (1974). Individuals choose one of M alternatives $y \in \mathcal{Y} = \{1, \dots, M\}$, achieving utility $u_y(X, V_y)$ if choice y is made. Individuals observe the utility achieved from all choices and select the alternative that delivers 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 $u \equiv (u_1, \dots, u_M)$ and the distribution of $V \equiv (V_1, \dots, V_M)$, 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 $u_1, \dots, u_M$. 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 pairs of utility functions and distributions of unobservable utility shifters that are compatible with the distribution of the data.

In McFadden (1974), the utility functions are linear in covariates and additively separable in unobservables. The distribution of V is fully specified such that its components 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. Examples include 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 characterization of the identified set applies in all these

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

cases and, indeed, even in the absence of parametric restrictions. When parametric restrictions are imposed, this characterization consists of a collection of conditional moment inequalities to which recent methods for estimation and inference can be applied.

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 and Rosen (2013).

The IV model studied here is unrestrictive relative to many other models of multiple discrete-choice permitting endogeneity that have been used until now. In our IV model, there is no restriction placed on the process that generates the potentially endogenous explanatory variables. In this sense, the model is incomplete and 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*

Let $\mathcal{U}$ and $\mathcal{P}_V$ denote the collection of utility functions and distributions of unobserved heterogeneity admitted by the model. The set of utility functions and distributions of latent variables identified by our model is characterized by a system of inequalities that is convenient to express in terms of a conditional containment functional associated with a 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, $V = (V_1, \dots, V_M)$, that leads to a particular realization y of Y when the explanatory variables X take the value x and the utility functions u govern choices. The conditional containment functional $P_0[\mathcal{T}_v(Y, X; u) \subseteq \mathcal{S}|z]$ gives the probability conditional on instrumental variable $Z = z$ that $\mathcal{T}_v(Y, X; u)$ is a subset of the set $\mathcal{S}$, which can be any *test set* on the support of V. Because $V \in \mathcal{T}_v(Y, X; u)$ by construction, $\mathcal{T}_v(Y, X; u) \subseteq \mathcal{S}$ implies that $V \in S$, so that also making use of the independence of Z and V, we have

$$P_V(\mathcal{S}) = P_V(\mathcal{S}|z) \geq P_0\big[\mathcal{T}_v(Y, X; u) \subseteq \mathcal{S}|z\big] \quad \text{a.e. } z \in \mathcal{Z}, \tag{1.1}$$

where $\mathcal{Z}$ and $\mathcal{V}$ denote the support of Z and V, respectively. The notation $P_0[\cdot|z]$ denotes probabilities taken with respect to $F^0_{YX|Z}$—the distribution of (Y, X) given $Z = z$, which is identified from the data—and $P_V(\mathcal{S})$ is the probability mass that the distribution of unobservables P_V assigns to the set $\mathcal{S}$.

We show in Theorem 1 that application of (1.1) to *all closed test sets* $\mathcal{S} \subseteq \mathcal{V}$ characterizes the identified set of structures (u, P_V).[2] This in turn delivers sharp bounds on

[2]By the "identified set," we mean the set of admissible structures comprising pairs (u, P_V) that deliver the distributions $F^0_{YX|Z}$ for almost every z in the support of Z. Some authors term this the *sharp identified set*.

the counterfactual choice probability (CCP), or the selection probability as McFadden (1974) called it, that an individual drawn at random from the population would choose alternative y if his covariates were exogenously fixed at some x.

We show that in the context of any particular model, the only sets $\mathcal{S}$ that need to be considered when judging whether a particular pair (u, P_V) is 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 to enable 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.

To illustrate, consider a model for which u^* and P_V^* are the true (but unknown) utility functions and the joint distribution of V. Let

$$\wp(x, y) \equiv P_V^*\big[\big\{v : \forall y' \in \mathcal{Y}, u_y^*(x, v_y) \geq u_{y'}^*(x, v_{y'})\big\}\big] \tag{1.2}$$

denote the probability that $Y = y$ given $X = x$ when V and X are independently distributed, equivalently, the counterfactual choice probability that a randomly drawn individual from the population would choose alternative y if his covariates were exogenously set to x, holding P_V^* fixed. In the classical conditional logit model with utility functions $u_y^*(x, v_y) = x'\beta_y^* + v_y$, with β_M^* normalized to zero, the probabilities involved are the well known expressions

$$\wp(x, y) = \frac{\exp(x'\beta_y^*)}{1 + \sum_{y'=1}^{M-1} \exp(x'\beta_{y'}^*)}. \tag{1.3}$$

Our first outer region, applicable with discrete X, contains all utility functions u^* and distributions P_V^* such that

$$\wp(x, y) \geq \max_{z \in \mathcal{Z}} \big\{P_0[Y = y \wedge X = x | z]\big\} \tag{1.4}$$

holds for all y and x in the support of Y and X. In a parametric setting, any researcher able to calculate a likelihood function when explanatory variables X are assumed exogenous is able to calculate these outer regions directly. In the conditional logit case, this outer region is convex, which simplifies computation. In the absence of any parametric restrictions and, indeed, without even the utility-maximizing model of choice behavior but with only knowledge of distributions $F_{YX|Z}^0$, the outer region defined by (1.4) provides sharp nonparametric bounds on $\wp(x, y)$, which may be of interest in their own

right, for example, for counterfactual analysis.[3] Our second outer region provides a refinement of that given by (1.4), which can be informative with discrete and continuous X in models with parametrically specified or nonparametric shape-restricted utility functions. We provide illustrations of outer regions, identified sets, and nonparametric bounds on $\wp(x, y)$ in particular examples in Section 4.

1.2 *Related results*

The prior literature on multinomial choice is substantial. Only a small subset of this literature has allowed for endogeneity. An important early contribution is Matzkin (1993), where it is shown that if the unobservable components of utility from the different alternatives are identically distributed and conditionally independent, and if there are alternative-specific regressors with large support, then the latent utility functions can be nonparametrically identified. Lewbel (2000) showed 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 the regressors is independent of the special regressors conditional on the instrument. Some recent papers have provided sufficient conditions for point identification under alternative assumptions. Petrin and Train (2010) used a control function approach in a triangular model. Fox and Gandhi (2009) provided sufficient conditions for nonparametric identification in a recursive setting. Chiappori, Komunjer, and Kristensen (2011) provided an alternative route to nonparametric identification, relying on conditional independence and completeness conditions. In limited dependent variables models with simultaneity, Matzkin (2012) built on Matzkin (2008) to provide conditions for nonparametric identification 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 consist of a large number of markets. Berry and Haile (2009, 2010) established conditions for nonparametric identification, the latter when microlevel data are 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 requires neither variables that differ across alternatives nor covariates with large support.

[3] Let $Y = h(\cdot, V)$ denote the choice function of an individual with unobservable utility shifter V. By the law of total probability,

$$\wp(x, y) = P_V\big[h(x, V) = y \wedge X = x|z\big] + P_V\big[h(x, V) = y|X \neq x, Z = z\big]P_0[X \neq x|Z = z].$$

With $V \perp\!\!\!\perp Z$ but no restrictions on the choice function, $P_V[h(x, V) = y|X \neq x, Z = z]$ can take any value on $[0, 1]$, implying that for all z,

$$\wp(x, y) \geq P_V\big[h(x, V) = y \wedge X = x|z\big] = P_0[Y = y \wedge X = x|z]$$

subject to $\sum_{y\in\mathcal{Y}} \wp(y, x) = 1$. Without restrictions on the function $h(\cdot, V)$ across x, this provides sharp bounds on the collection of $\wp(x, y)$ across all (y, x).

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. 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 results on partially identifying models of multinomial response can be found in Manski (2007) and Beresteanu, Molchanov, and Molinari (2011), although the multinomial response models studied and the mechanisms by which partial identification is obtained in these papers are rather different and do not involve instrumental variables.[4] Papers with set-identifying results for parameters of binary choice models include Manski and Tamer (2002), Magnac and Maurin (2008), Komarova (2007), and, recently, Chesher and Rosen (2012a), where we show how the ideas used in both this paper and Chesher and Rosen (2013) can be used to establish identified sets in an instrumental variable random coefficient model.

To establish that our bounds are sharp, we make use of results from random set theory, in particular, Artstein's inequality (Artstein (1983)). Such results in random set theory have been previously used to establish set identification in other contexts by Beresteanu, Molchanov, and Molinari (2011, 2012). Beresteanu, Molchanov, and Molinari (2011) used the Aumann expectation of set-valued random variables to tractably characterize the identified set in models with convex moment predictions. Their characterization applies rather generally, covering as examples models of games with multiple equilibria, and best linear prediction and multinomial choice models with interval data on exogenous explanatory variables. In related work, Galichon and Henry (2011) used optimal transportation theory to 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 provided several approaches to facilitate the computational tractability of this approach, with further results pertaining to optimal transportation given in Ekeland, Galichon, and Henry (2010).

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 previously. 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 independent of unobserved heterogeneity and that may be excluded from agents' utility functions. Our use of random

[4]Specifically, Manski (2007) provided bounds on choice probabilities when agents face counterfactual choice sets, and Appendix F of Beresteanu, Molchanov, and Molinari (2011) established set identification results in a multinomial choice model with interval data on explanatory variables.

set theory provides a characterization of the identified set that applies in fully nonparametric, semiparametric, and parametric models, but that is more amenable to current estimation and inference approaches in parametric models. We use the notion of core-determining classes defined in Galichon and Henry (2011) to refine our characterization of the identified set. They showed how this can be done in econometric models of games under a monotonicity condition, which is not satisfied in our model. We provide a novel algorithm for the construction of core-determining classes in our setup.

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 and distributions of unobserved heterogeneity. 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 $P_0[Y = y|x]$ for all $(y, x) \in \mathrm{Supp}(Y, X)$, which are precisely likelihood contributions if the model is parametrically specified. In Section 3.3, we provide a theorem that defines a 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, 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 unobserved heterogeneity. We then specialize our results to the separable case, on which much of the previous literature on models of discrete choice has focused.

2.1 *Nonseparable utility*

An individual makes one choice from M alternatives, obtaining utility $u_y(X, V_y)$ from alternative $y \in \mathcal{Y} \equiv \{1, 2, \dots, M\}$, where for each $y \in \mathcal{Y}$, $U_y : \mathrm{Supp}(X, V_y) \to \mathbb{R}$, and where $\mathrm{Supp}(A, B)$ 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 on 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 A.5 below. In Section 2.2, we consider the common special case where the utility functions are additively separable in unobservables.

Individuals know their values of X and V, and so choose Y to maximize $u_y(X, V_y)$. Due to monotonicity of the utility functions $u_y(\cdot, \cdot)$ in their second argument coupled with Restriction A.4 below, ties in the utility delivered by any two alternatives occur

with probability 0, and the utility-maximizing alternative is unique with probability 1 conditional on any realization of (X, Z).[5] We impose sufficient conditions for this both for convenience and because it is common in single agent discrete-choice models, and henceforth define the *choice function* $h_v(X, V; u)$ that delivers the utility maximizing choice with probability 1 so that

$$Y = h_v(X, V; u). \tag{2.1}$$

Thus the CCP of (1.2) for choice y at covariate value x is equivalently given by

$$\wp(x, y) \equiv P_V\big[h(x, V; u) = y\big]. \tag{2.2}$$

The model comprises the following restrictions.

Restriction A.1. The variables (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 $\mathcal{Y} \equiv \{1, 2, \ldots, M\}$, and the supports of X and Z are $\mathcal{X}$ and $\mathcal{Z}$, respectively. The joint support of (Y, X, Z) is a (possibly nonstrict) 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 $\mathrm{Supp}(V|X = x, Z = z)$, 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 A.2. 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 A.3. Given (V, X, Z), $Y \in \arg\max_{y \in \mathcal{Y}} u_y(X, V_y)$.

Restriction A.4. 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, $\mathrm{Supp}(V|X = x, Z = z) \subseteq \mathbb{R}^M$. The marginal distribution of V belongs to a specified family of distributions $\mathcal{P}_V$.

Restriction A.5. The utility functions $u = \{u_1, \ldots, u_M\}$ 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.

[5]Note, however, that the tools 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. 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 classes in Section 3.3 would require some modification. Chesher and Rosen (2012b) considered simultaneous equations models of discrete choice for which multiple or, indeed, no solutions are feasible. This raises further issues of coherence and completeness that are logically distinct from the study of multiple discrete choice.

Restriction A.6. The variables V and Z are stochastically independent.

Restriction A.1 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 on different realizations 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 identification analysis, we determine the set of observationally equivalent structures that are admitted by the model and deliver the probability distributions $F^0_{YX|Z}(y, x|z)$ of Restriction A.2. The notation P_0 indicates probabilities calculated using these distributions. Under Restriction A.2, 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 $z \in \mathcal{Z}$ are made with respect to this distribution.

Restriction A.6 requires V and the variables Z to be independently distributed. 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.

Our identification results permit the classes of distributions $\mathcal{P}_V$ and utility functions $\mathcal{U}$ of Restrictions A.4 and A.5 to each be either parametrically or nonparametrically specified. When they are parametrically specified, one can use inference methods for conditional moment inequalities, as we discuss after Theorem 1. We do not assume the existence of alternative-specific covariates in our analysis, but Restriction A.5 is 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. We impose strict monotonicity of all but one of the utility functions in its corresponding unobservable and impose weak monotonicity of the remaining utility function in its unobservable. Combined with Restriction A.4, this guarantees that conditional on any realization of (X, Z), there is a unique utility-maximizing choice of Y almost surely.

The classes of utility functions $\mathcal{U}$ and distributions $\mathcal{P}_V$ should incorporate the same sort of normalizations imposed to enable point identification in models in which X and V are independently distributed. Although not required for set identification analysis, this will ensure that the volume of resulting identified sets is not unduly inflated relative to the well studied exogenous X case. With nonseparable utility functions, one might impose the location and scale normalizations $U_M = 0$ and $\text{var}(U_1(x, V)) = 1$ for some known value of x, and if utilities are also nonparametrically specified, one may impose the normalization $V_y \sim U[0, 1]$, all $y \in \mathcal{Y}$. Other normalizations are possible, with the key consideration that utility-maximizing choices are insensitive to an increasing transformation of utility *differences*.

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 A.5*.

Restriction A.5* (Additive Separability). Restriction A.5 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, all $x \in \mathcal{X}$.

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.

Example 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

$$P_V\Big[\bigwedge_{y\in\mathcal{Y}}(V_y \le v_y)\Big] = \prod_{y\in\mathcal{Y}} \exp\big(-\exp(-v_y)\big). \tag{2.3}$$

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.

Example 2. An IV generalization of the conditional probit model studied in Hausman and Wise (1978) specifies $\mathcal{P}_V$ as a parametric family of multivariate normal, $N(0, \Sigma)$, distributions with a suitable normalization of Σ. Again one can impose $u_y(X) \equiv X'\beta_y$ for all $y \in \mathcal{Y}$.

Note that unlike the classical conditional logit and multinomial probit models, the specifications admitted by our model do not require X and V to be independent. The specification of $\mathcal{P}_V$ restricts the *unconditional* distribution of V, P_V, to be independent and identically distributed (i.i.d.) Type 1 extreme value or multivariate normal, respectively. Due to the independence Restriction A.6, 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 $V|(X = x, Z = z)$ 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 the realization of (X, Z). Thus the model need not adhere to independence of irrelevant alternatives once we condition on 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 $W \equiv (W_1, \dots, W_{M-1}) \in \mathbb{R}^{M-1}$, where for each $y \in \mathcal{Y}$,

$$W_y \equiv V_y - V_M.$$

To see why, define the utility differences

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

Then there is a convenient representation for the selection of Y equivalent to (2.1) given by

$$Y = h_w(X, W; u) \equiv \arg\max_{y\in\mathcal{Y}} \Delta U_y(X, W).$$

Because the dependence of $h_w(X, W; u)$ on the utilities u is crucial, it is made explicit in the notation. Again under Restriction A.4, the set $h_w(x, W; u)$ is singleton with probability 1 for all $x \in \mathcal{X}$.[6]

The model requires the random components of utility, V, to have a distribution in the family $\mathcal{P}_V$. From the above discussion, we see that when Restriction A.5* 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 A.2. 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 $\mathbb{R}^2$. Since the optimal selection of alternatives is entirely determined by utility differences, it is convenient here to impose the common location normalization that $u_M(x) = 0$ for all $x \in \mathcal{X}$. Further, $\mathcal{P}_W$ should incorporate a scale normalization, for instance, by setting the variance of W_1 to a known constant, as required for point identification in models with exogenous X. Such a normalization is already embedded in the restrictions of the conditional logit model of Example 1, but must be imposed in other models, such as the multinomial probit model of Example 2.[7]

3. Identification

3.1 *The identified set*

We now develop results on the identifying power of our model. The task is to infer what structures are admitted by the model given knowledge of $F^0_{YX|Z}$, $z \in \mathcal{Z}$. To characterize the identified set, 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 set of (u, P_v) pairs that satisfy these inequality restrictions taken over any collection of test sets $\mathcal{S}$ comprises valid, though possibly nonsharp, bounds. 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 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).[8]

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

[7]See, for example, Chapter 5 of Train (2009) for further discussion of such a scale normalization for the multinomial probit model and the implied restrictions on the variance of V.

[8]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.

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

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

For any admissible u and each x, the sets $\mathcal{T}_v(y, x; u)$, $y \in \mathcal{Y}$, form a partition of $\mathbb{R}^M$, ignoring shared boundaries that under Restriction A.4 have P_V measure zero.

In the additively separable case with Restriction A.5* imposed, we can likewise define

$$\mathcal{T}_w(y, x; u) \equiv \{w : \forall k \in \mathcal{Y}, u_y(x) + w_y \geq u_k(x) + w_k\}.$$

Using this set, we can then replace V with W, P_V with P_W, and $\mathcal{V}$ with $\mathcal{W} \equiv \text{Supp}(W)$, 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 A.5 is imposed.

For any test set $\mathcal{S} \subseteq \mathcal{V}$, let $P_{V|XZ}(\mathcal{S}|x, z)$ denote the conditional probability of the event $\{V \in \mathcal{S}\}$ given $X = x$ and $Z = z$. We first consider the import of the independence Restriction A.6.

- *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 probability distributions of Restriction A.2 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 $P_{V|Z}(\mathcal{S}|z)$, which by independence must be invariant with respect to z.

Now consider *observational equivalence* conditions that 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 A.2.

- *Observational equivalence*: For any $x \in \mathcal{X}$, 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 $(y, x, z) \in \text{Supp}(Y, X, Z)$,

$$P_{V|XZ}\big(\mathcal{T}_v(y, x; u)|x, z\big) = P_0[Y = y|x, z]. \tag{3.2}$$

These two implications of the IV model lead to a system of inequalities that must be satisfied by all admissible (u, P_V) that deliver the distributions of Restriction A.2, namely $F^0_{YX|Z}$ for $z \in \mathcal{Z}$. This system of inequalities is now derived.

First, if (3.2) is to be satisfied, then for any $\mathcal{S}$, the *smallest* value that $P_{V|XZ}(\mathcal{S}|x,z)$ can take is equal to the sum of the probabilities $P_0[Y=y|x,z]$ associated with all sets $\mathcal{T}_v(y,x;u)$ contained *entirely within* $\mathcal{S}$. This is expressed in the inequality

$$P_{V|XZ}(\mathcal{S}|x,z) \geq \sum_{y\in\mathcal{Y}} 1\big[\mathcal{T}_v(y,x;u)\subseteq\mathcal{S}\big]P_0[Y=y|x,z], \tag{3.3}$$

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

Marginalizing with respect to X given $Z=z$ on the left- and right-hand sides of (3.3) and simplifying using (3.1), there is

$$P_V(\mathcal{S}) \geq \int_{x\in\mathcal{X}} \left(\sum_{y\in\mathcal{Y}} 1\big[\mathcal{T}_v(y,x;u)\subseteq\mathcal{S}\big]P_0[Y=y|x,z]\right) dF^0_{X|Z}(x|z). \tag{3.4}$$

All (u,P_V) in the identified set $\mathcal{D}^0(\mathcal{Z})$ satisfy these inequalities for all $\mathcal{S}\subseteq\mathcal{V}$ and almost everywhere (a.e.) $z\in\mathcal{Z}$. Given a choice of $u\in\mathcal{U}$ with knowledge of the distributions $F^0_{YX|Z}$ of Restriction A.2, the right-hand side of (3.4) can be calculated for any test set $\mathcal{S}$. Likewise, for any such $\mathcal{S}$, given a choice $P_V\in\mathcal{P}_V$, the left-hand side of (3.4) can be calculated.

To show that this system of inequalities taken over all closed $\mathcal{S}\subseteq\mathcal{V}$ provides the identified set, we cast (3.4) in terms of random set theory. For nonstochastic (y,x), the set $\mathcal{T}_v(y,x;u)$ defined in (3.1) is a nonstochastic set of values in $\mathcal{V}$. Applying this set-valued mapping to random variables (Y,X) gives $\mathcal{T}_v(Y,X;u)$, a random closed set on $\mathcal{V}$.[9] Specifically, the right-hand side of (3.4) is its *conditional containment functional*, conditional on the instrument $Z=z$:

$$\begin{aligned} &P_0\big[\mathcal{T}_v(y,x;u)\subseteq\mathcal{S}|Z=z\big] \\ &\quad= \int_{x\in\mathcal{X}} \left(\sum_{y\in\mathcal{Y}} 1\big[\mathcal{T}_v(Y,X;u)\subseteq\mathcal{S}\big]P_0[Y=y|x,z]\right) dF^0_{X|Z}(x|z). \end{aligned}$$

Combining this with (3.4), it follows that the identified set must satisfy

$$P_0\big[\mathcal{T}_v(Y,X;u)\subseteq\mathcal{S}|Z=z\big] \leq P_V(\mathcal{S}) \tag{3.5}$$

for all sets $\mathcal{S}\subseteq\mathcal{V}$ and instrument values $z\in\mathcal{Z}$. The following theorem uses this information with Artstein's inequality (Artstein (1983)) to establish that all and only (u,P_V) that satisfy (3.5) for all closed $\mathcal{S}$ and a.e. $z\in\mathcal{Z}$ can deliver the distributions of Restriction A.2, equivalently, that these inequalities characterize the identified set.

THEOREM 1. *Let Restrictions A.1–A.6 hold. Then the identified set for (u,P_V) is*

$$\begin{aligned} \mathcal{D}^0(\mathcal{Z}) = \big\{&(u,P_V)\in\mathcal{U}\times\mathcal{P}_V: \\ &\forall\mathcal{S}\in\mathsf{F}(\mathcal{V}),\ P_0\big[\mathcal{T}_v(Y,X;u)\subseteq\mathcal{S}|z\big]\leq P_V(\mathcal{S}),\ \textit{a.e. } z\in\mathcal{Z}\big\}, \end{aligned} \tag{3.6}$$

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

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

Corollary 1. *If Restriction A.5 is replaced with the additive separability Restriction A.5*, the identified set for* (u, P_W) *is*

$$\begin{aligned}\mathcal{D}_w^0(\mathcal{Z}) \equiv \big\{(u, P_W) \in \mathcal{U} \times \mathcal{P}_W :\\ \forall \mathcal{S} \in \mathsf{F}(\mathcal{W}), P_0\big[\mathcal{T}_w(Y, X; u) \subseteq \mathcal{S}|z\big] \le P_W(\mathcal{S}),\ a.e.\ z \in \mathcal{Z}\big\},\end{aligned} \tag{3.7}$$

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

Corollary 2. *Under the restrictions of Theorem 1, for any* $y \in \mathcal{Y}$ *and* $x \in \mathcal{X}$, *sharp bounds on the counterfactual choice probability* $\wp(x, y)$ *are given by*

$$\inf_{(u,P_V)\in\mathcal{D}^0(\mathcal{Z})} P_V\big(h(x, V) = y\big) \le \wp(x, y) \le \sup_{(u,P_V)\in\mathcal{D}^0(\mathcal{Z})} P_V\big(h(x, V) = y\big), \tag{3.8}$$

and the identified set for the collection of counterfactual choice probabilities

$$\big\{\wp(x, y) : (x, y) \in \mathrm{Supp}(X, Y)\big\},$$

is given by

$$\begin{aligned}\big\{\wp(x, y) : \exists (u, P_v) \in \mathcal{D}^0(\mathcal{Z})\ s.t.\\ \forall (x, y) \in \mathrm{Supp}(X, Y), \wp(x, y) = P_V\big(h(x, V) = y\big)\big\}.\end{aligned} \tag{3.9}$$

The proof of Theorem 1 employs Artstein's inequality to establish sharpness. Specifically, we use it to establish the existence of 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 $\mathcal{T}_v(Y, X; u)$, such that conditional on any realization of Z, $\tilde{V} \in \tilde{\mathcal{T}}$ with probability 1. This is then used to establish that every (u, P_V) in $\mathcal{D}^0(Z)$ can produce the distributions $F^0_{YX|Z}$ of Restriction A.2. Note that the same argument applies, so that sharpness is maintained, if the containment functional inequality in the definition of $\mathcal{D}^0(Z)$ is replaced by the *capacity* functional inequality $P_0[\mathcal{T}_v(Y, X; u) \cap \mathcal{S} \neq \emptyset | Z = z] \ge P_V(\mathcal{S})$ for all compact $\mathcal{S} \subseteq \mathcal{V}$. The first corollary of the theorem is an immediate consequence for the case of separable utility functions. Corollary 2, also an immediate consequence, provides the identified set for the resulting counterfactual choice probabilities.

Without appeal to Theorem 1, a definition of the identified set is given by the set of (u, P_V) such that for all $\mathcal{R}_x \subseteq \mathcal{X}$ and all $y \in \mathcal{Y}$,

$$\begin{aligned}&\int_{x\in\mathcal{R}_x} P_0(Y = y|x, z)\, dF^0_{X|Z}(x|z)\\ &\quad= \int_{x\in\mathcal{R}_x} P_{V|XZ}\big(\{v : h(X, v; u) = y\}|x, z\big)\, dF^0_{X|Z}(x|z), \quad \text{a.e. } z \in \mathcal{Z},\end{aligned}$$

with h the choice function defined in (2.1), for some collection of conditional distributions of V given X, Z, and $P_{V|XZ}$. Moreover, these distributions must conform with the independence restriction, Restriction A.6, so that for all $\mathcal{S} \subseteq \mathcal{V}$,

$$P_V(\mathcal{S}) = \int_{x \in \mathcal{X}} P_{V|XZ}(\mathcal{S}|x, z)\, dF^0_{X|Z}(x|z), \quad \text{a.e. } z \in \mathcal{Z}.$$

Directly establishing conditions for the existence of such a collection of conditional distributions, an infinite-dimensional object when either X or Z is continuous, is generally difficult. Theorem 1 shows that the identified set is, in fact, fully characterized by a collection of conditional moment inequalities $E[m(Y, X, \theta, \mathcal{S})|Z = z] \geq 0$ with moment function

$$m(Y, X, \theta, \mathcal{S}) = P_V(\mathcal{S}) - 1\big[\mathcal{T}_v(Y, X; u) \subseteq \mathcal{S}\big], \quad \theta \equiv (u, P_V).$$

The moment inequalities must hold for almost every $z \in \mathcal{Z}$ and for all closed $\mathcal{S} \in \mathcal{V}$. We shortly establish that in the context of any particular model, it suffices to consider a much smaller collection of test sets $\mathcal{S}$, and, in particular, that when X is discrete, the number of test sets required is finite. Given the representation of $\mathcal{D}^0(\mathcal{Z})$ via moment inequalities, at least when $\mathcal{P}_V$ and $\mathcal{U}$ are parametrically specified, recently developed methods for estimation and inference for conditional moment inequalities are applicable. When covariates and instruments are discrete, the identified set is characterized by a finite number of moment inequalities, and the inferential methods of Chernozhukov, Hong, and Tamer (2007), Beresteanu and Molinari (2008), Romano and Shaikh (2008), Rosen (2008), Galichon and Henry (2009), Bugni (2010), Canay (2010), or Henry, Meango, and Queyranne (2011) can be used, among others. When instruments are continuous, even with discrete X, the characterization embodies infinitely many unconditional moment inequalities. In this case, one could use the approach of either Andrews and Shi (2013) or Chernozhukov, Lee, and Rosen (2013) for inference on conditional moment inequalities. When X is continuous, there are infinitely many moment inequalities for each $z \in \mathcal{Z}$. This is related to, but somewhat different than, the problem of a conditional inequality restriction holding with a continuous conditioning variable. Though the structure of the problem with continuous X differs somewhat from that with continuous Z, we believe suitable modifications of Andrews and Shi (2013) or Chernozhukov, Lee, and Rosen (2013) should apply in this case. We focus in this paper on characterization of the identified set, leaving this topic to future research.[10]

Finally, we stress the necessity of exogenous instruments. Indeed, the following corollary illustrates that in our model, without restrictions on the distributions of unobservables beyond the support and absolute continuity conditions of Restriction A.4,

[10]With discrete X, the number of core-determining sets and hence conditional moment inequalities of our characterization will be finite, though possibly quite large. In such cases and under suitable conditions, the asymptotic approximations used by Andrews and Shi (2013) and Chernozhukov, Lee, and Rosen (2013) will be valid, yet the number of conditional moment inequalities may be large relative to sample size, a problem studied in a related context by Menzel (2009). The best way to select among these for accurate finite sample inference may thus pose an interesting question even with discrete X.

there is no identifying power for u without instruments Z. That is, the existence of instruments Z is required to achieve nontrivial set identification in the sense that without them, any identifying information comes solely from restrictions (e.g., parameterizations) placed on u and P_V.

Corollary 3. *Let the restrictions of Theorem 1 hold, let $\mathcal{P}_V$ admit all distributions that satisfy the support and absolute continuity conditions of Restriction A.4, and suppose there are no exogenous instruments Z. Then all $u \in \mathcal{U}$ are observationally equivalent, so that for all $u \in \mathcal{U}$, there exists $P_V \in \mathcal{P}_V$ such that $(u, P_V) \in \mathcal{D}^0(\mathcal{Z})$.*

3.2 *Relation to independent X and V*

When X and V are stochastically independent, the above characterization reduces to a collection of equalities involving maximum likelihood probabilities. To show this, set $X = Z$ and consider that for each $x \in \mathcal{X}$ and any $u \in \mathcal{U}$, we have from (3.5) applied to test sets $\mathcal{S} = \mathcal{T}_v(y, x; u)$ that

$$\forall y \in \mathcal{Y}, \quad P_0[Y = y|x] \leq P_V\big(\mathcal{T}_v(y, x; u)\big),$$

where

$$\sum_{y \in \mathcal{Y}} P_0[Y = y|x] = 1 \quad \text{and} \quad \sum_{y \in \mathcal{Y}} P_V\big(\mathcal{T}_v(y, x; u)\big) = 1.$$

It follows that

$$\forall y \in \mathcal{Y}, \quad P_0[Y = y|x] = P_V\big(\mathcal{T}_v(y, x; u)\big), \tag{3.10}$$

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 $y < M$, $u_M(x) = 0$, and $P_V[\mathcal{T}_v(y, x; u)]$ takes the familiar form

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

In this case, (3.10) 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).

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 to produce 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 on 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 $\mathrm{int}(\mathcal{S})$ and $\mathrm{cl}(\mathcal{S})$ to denote the interior and closure, respectively, of any set $\mathcal{S}$. The proof of Theorem 2 below makes use of the following lemma, which provides some properties of the sets $\mathcal{T}_v(y, x; u)$. In the following analysis, we make use of the support of the random set $\mathcal{T}_v(Y, X; u)$,

$$\mathsf{T}_v(Y, X; u) \equiv \big\{\mathcal{T}_v(y, x; u) : (x, y) \in \mathrm{Supp}(X, Y)\big\},$$

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

$$\mathsf{T}_w(Y, X; u) \equiv \big\{\mathcal{T}_w(y, x; u) : (x, y) \in \mathrm{Supp}(X, Y)\big\}.$$

Lemma 1. *Consider the model defined by Restrictions A.1–A.6. 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 A.5* holds, the sets $\mathcal{T}_v(y, x; u)$ and $\mathcal{T}_w(y, x; u)$ are convex.* (iii) *If Restriction A.5* holds and $\mathcal{V} = \mathbb{R}^M$, these sets are nonempty, with strictly positive Lebesgue measure whenever $u_{y'}(x) - u_y(x) < \infty$ for all $y' \in \mathcal{Y}$, $y' \neq y$.*

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

Theorem 2. *Let Restrictions A.1–A.6 hold. The identified set* (3.6) *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* (3.7) *in Corollary 1 if, additionally, Restriction A.5* holds, replacing the support of $\mathcal{T}_v(Y, X; u)$ with that of $\mathcal{T}_w(Y, X; u)$.*

Theorem 2 characterizes a collection of test sets $\mathcal{S}$ such that if the conditional containment inequality (3.5) holds for each of these, then it must hold for all closed $\mathcal{S}$. Thus the identified set is fully characterized by the set of pairs (u, P_V) that satisfy this inequality over this smaller collection of test sets. The reduction in the collection of required test sets stems from the structure of the utility functions u and the utility-maximizing framework, and will produce different collections for different u, as well as different z. It remains an open question whether the collection of sets in Theorem 2 is minimal, in consideration of either particular (u, P_v) or all $(u, P_v) \in \mathcal{U} \times \mathcal{V}$. Indeed, there will, in general, be particular (u, P_v) not in the identified set that violate inequalities for more than

one core-determining $\mathcal{S}$, and the possibility remains that there may be further, computationally useful methods to reduce the collection of sets to be checked for some of these, at least in some classes of models. Of course, once *any* of the requisite inequalities are violated, (u, P_v) can be excluded from the identified set and the other inequalities need not be checked.

The characterization of core-determining sets in Theorem 2 applies whether X is discrete or continuous. When X is discrete, the collection of core-determining sets is finite and can be computed via the following algorithm. This collection *varies with the specific utility functions u under consideration*, but is invariant with respect to changes in P_V. Let the support of discrete X be $\mathcal{X} \equiv \{x_1, \dots, x_K\}$; X may be a finite-dimensional vector. The algorithm can 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 $\mathcal{T}(y, x; u)$ to denote either $\mathcal{T}_v(Y, X; u)$ 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.[11] Let $\mathsf{C}_1 \parallel \mathsf{C}_2$ denote the collection of the sets that appear either in C_1 or in C_2.[12] Let $\mathsf{C}(u)$ denote the collection of the interiors of the sets on the support of $\mathcal{T}(Y, X; u)$:

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

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

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

Let $\operatorname{Con}(\cdot)$ applied to a list of sets select the connected sets in the list. Step 2 of the algorithm recursively creates the list of sets

$$\mathsf{C}(u) \parallel \operatorname{Con}\big(\mathsf{C}(u) \otimes \mathsf{C}(u)\big) \parallel \operatorname{Con}\big(\operatorname{Con}\big(\mathsf{C}(u) \otimes \mathsf{C}(u)\big) \otimes \mathsf{C}(u)\big) \parallel \cdots.$$

[11]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

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

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

This is the same as the list

$$\mathrm{Con}\big(\mathsf{C}(u) \parallel \mathsf{C}(u) \otimes \mathsf{C}(u) \parallel \mathsf{C}(u) \otimes \mathsf{C}(u) \otimes \mathsf{C}(u) \parallel \cdots\big),$$

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. In the examples in Section 4, the algorithm ran very quickly relative to other computations, though this could change if X had richer support.

The number of core-determining sets that result is far smaller than the number of possible unions of sets on the support of $\mathcal{T}(Y, X; u)$. For example, in a three-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.[13]

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 $\mathcal{T}_v(Y, X; u)$, equivalently, with each set in the collection $\mathsf{C}(u)$. Application of the containment functional inequality to such sets produces the inequalities

$$P_V\big[\mathcal{T}_v(y, x; u)\big] \geq P_0[Y = y \wedge X = x|z],$$

which must hold for all $(y, x, z) \in \mathrm{Supp}(Y, X, Z)$, for all (u, P_V) in the identified set. Noting that $P_V[\mathcal{T}_v(y, x; u)] = \wp(x, y)$, it follows that

$$\wp(x, y) \geq \max_{z \in \mathcal{Z}} P_0[Y = y \wedge X = x|z] \tag{3.11}$$

must hold for all $(y, x) \in \mathrm{Supp}(Y, X)$. These inequalities define an *outer region* within which lies the identified set for (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 (3.11) are 0. With discrete X, but in the absence of any parametric restrictions, these inequalities provide nonparametric bounds on the counterfactual choice probability $\wp(x, y)$, with an upper bound implied by the requirement that for all x, $\wp(x, y)$ must sum to 1 across values of y. In a parametric model, $\wp(x, y)$ is simply 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, for example, the familiar logit

[13]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.

probabilities of (1.3) in the conditional logit model. When X is endogenous, $\wp(x, y)$ is the counterfactual choice probability for alternative y were a randomly drawn member of the population to have its covariates set to x, keeping P_V fixed. Whenever a model is considered for which, under an exogeneity restriction, there is a well defined parametric likelihood function, this can be substituted for $\wp(x, y)$ in (3.11) above, so that the outer region defined by these inequalities is very easy and quick to compute.[14]

Our second outer region can be useful with either continuous or discrete X. This outer region provides a refinement of the previous one 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

$$P_V\big[\mathcal{T}_v(y, x; u)\big] \geq \int_{(x':\mathcal{T}_v(y,x';u)\subseteq\mathcal{T}_v(y,x;u))} P_0\big[Y = y \wedge X = x'|z\big]\, dF^0_{X|Z}\big(x'|z\big).$$

In the three-choice models with binary X considered in Section 4, this improvement is obtained for two of the six 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 both discrete and continuous X.

4. Illustration: Three-choice models

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 to be 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 cases where $K = 2$ and $K = 4$.

4.1 *Core-determining sets*

In the three-choice model, utilities are determined as

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

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

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

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

[14]Another interpretation of the inequality (3.4) is that the event $V \in \mathcal{T}_v(y, x; u)$ is a *necessary* but not sufficient condition for the event $\{Y = y \wedge X = x\}$. In this sense, the collection of inequalities (3.4) across all $z \in \mathcal{Z}$ parallels the use of necessary but not sufficient conditions for equilibrium proposed by Andrews, Berry, and Jia (2004) to bound parameters in econometric models of discrete games.

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 $\mathbb{R}^2$ "centered" on a point denoted $w(x)$ with coordinates $W_1 = -u_1(x)$ and $W_2 = -u_2(x)$. The collection of sets $\mathsf{G}(u)$ 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.[15] 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 $w(x)$ for each of the two possible values of x.[16] 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 $w(x_1)$ lies higher (in the W_1 direction) than $w(x_2)$ and another in which these positions are reversed. When K is much greater 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 $K = 2$.

The 12 core-determining sets for the case where $K = 2$ when $w(x_2)$ lies northwest of $w(x_1)$ 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 nonsingleton unions of sets on the support of $\mathcal{T}(Y, X; u)$ obtained by following the algorithm provided above. Other orientations likewise deliver 12 core-determining sets.[17]

4.2 *Illustrations of identified sets*

In this section, we give examples of identified sets for particular distributions $F^0_{YX|Z}$. To keep the dimension of the identified set small enough to allow a graphical display, we impose a linear index restriction, with utility functions determined by a parameter $\alpha = (a_{01}, a_{02}, a_{11}, a_{12})$ as

$$U_1 = a_{01} + a_{11}x + V_1, \qquad U_2 = a_{02} + a_{12}x + V_2, \qquad U_3 = V_3.$$

Probability distributions are generated with $a_{01} = 0$, $a_{11} = 1$, $a_{02} = 0$, and $a_{12} = -0.5$. In each example the components of V are distributed i.i.d. Type 1 extreme value.

The endogenous variable X is generated according to

$$X = x_k \quad \text{if and only if} \quad c_{k-1} < X^* \le c_k,$$
$$X^* = d_0 + d_1 Z + d_2^{1/2} W + \lambda g(V), \quad W \sim N(0, 1),$$

[15]The pdf version of this article, available on-line, has color graphics which are easier to read and interpret.

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

[17]Tables 2 and 3 of an on-line working paper version of this paper, Chesher, Rosen, and Smolinski (2011), enumerate core-determining sets for all possible orientations in three-choice models with $K = 2$.

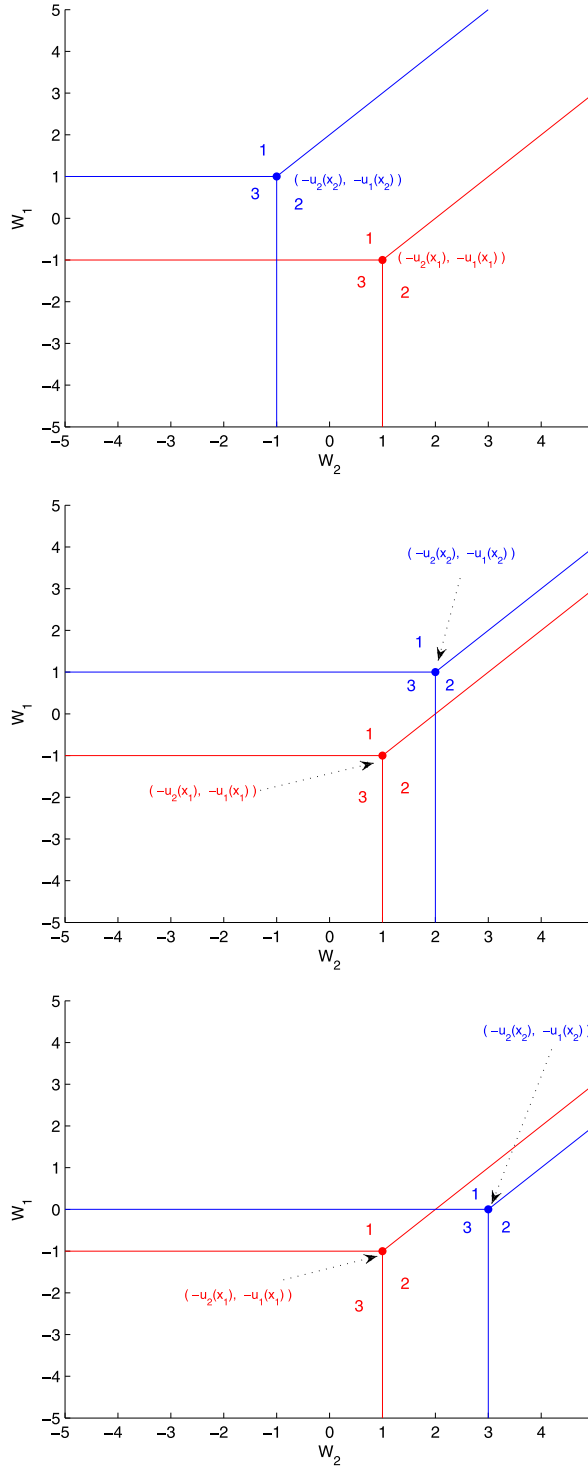

FIGURE 1. Orientations of $w(x) = (-u_1(x), -u_2(x))$ when $M = 3$ and $K = 2$.

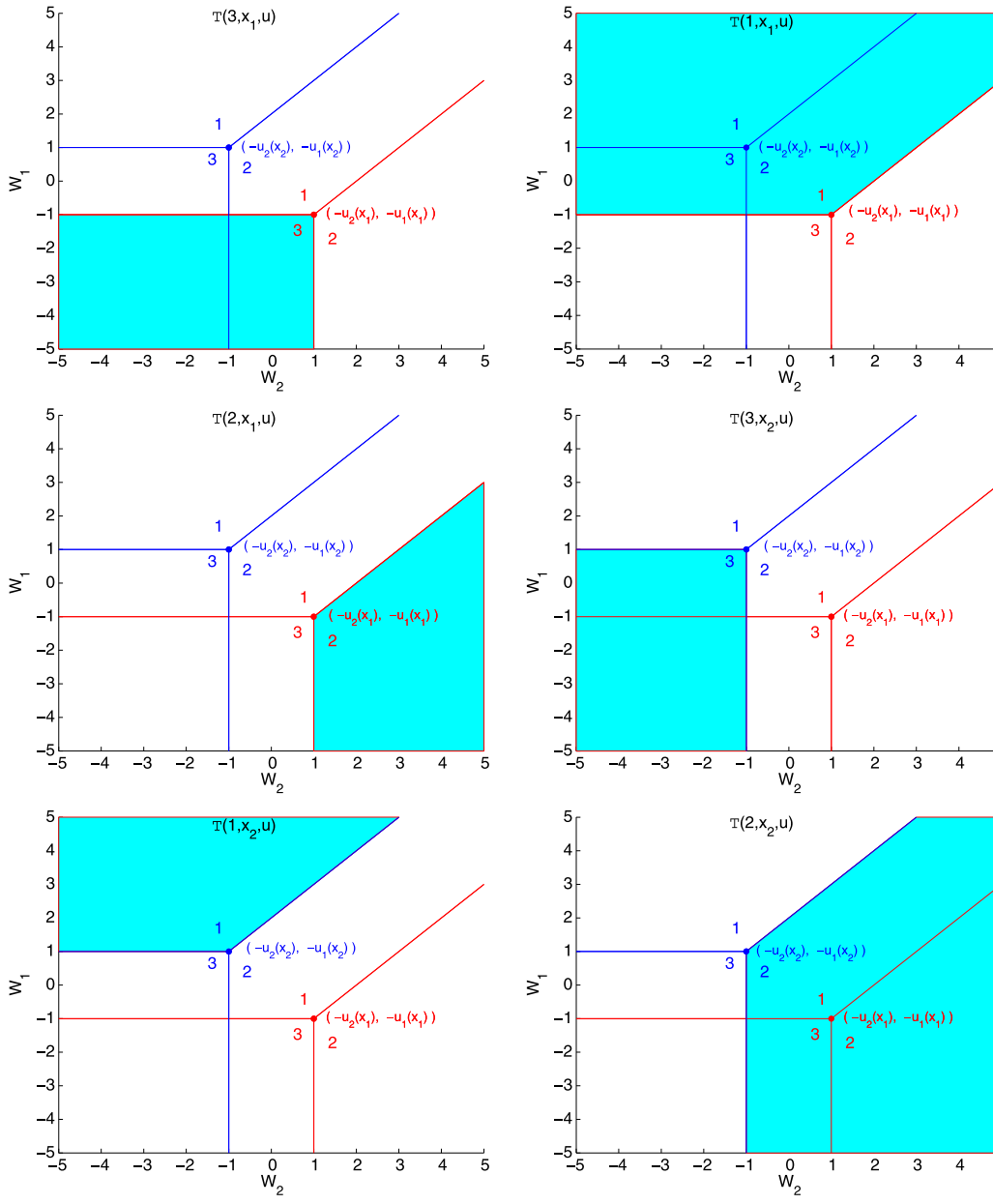

FIGURE 2. Core-determining sets for binary X: sets on the support of $\mathcal{T}(y, x; u)$.

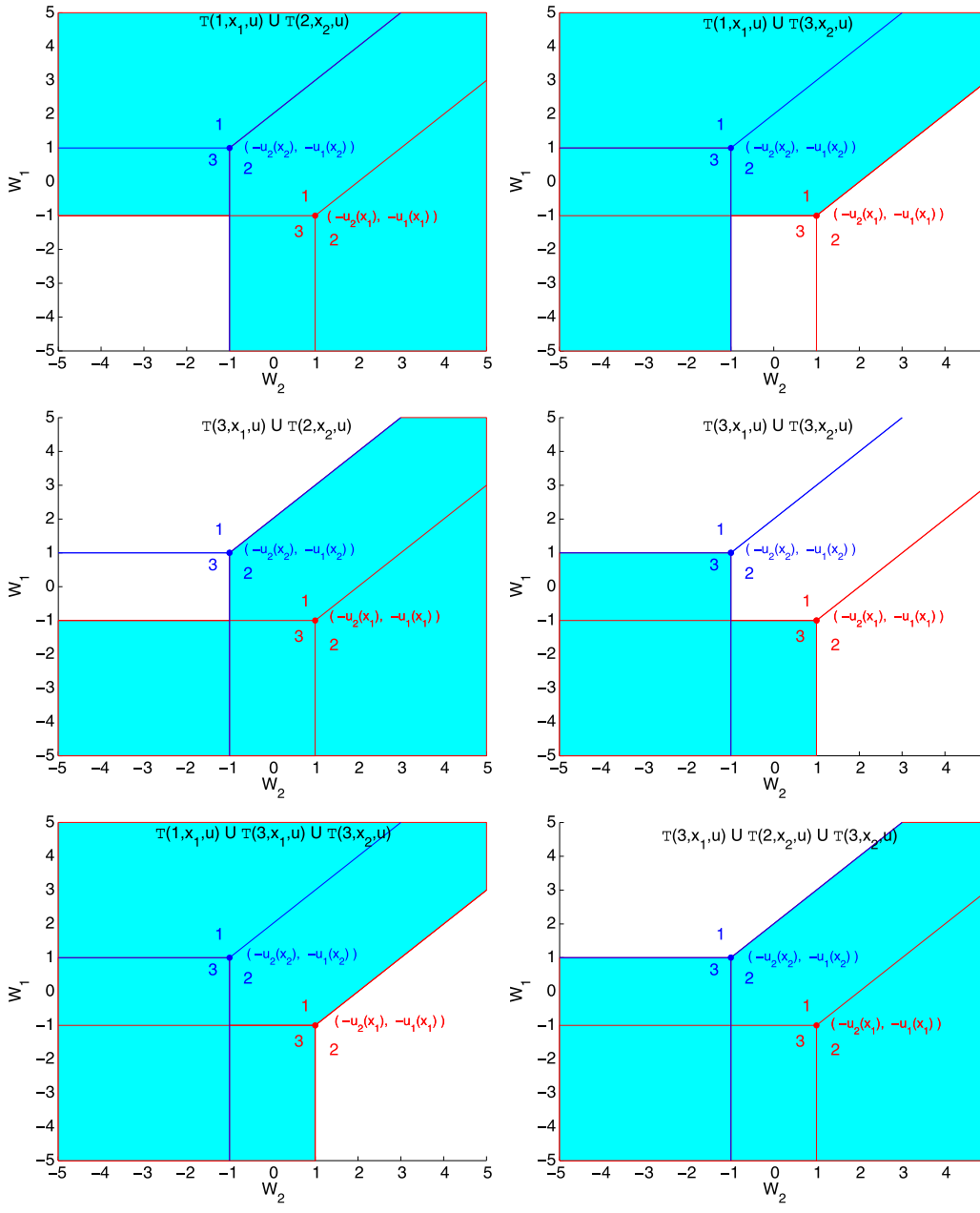

FIGURE 3. Core-determining sets for binary X: nonsingleton unions of sets on the support of $\mathcal{T}(y, x; u)$.

with $(V, W) \perp\!\!\!\perp Z$, $V \perp\!\!\!\perp W$, $g(V) = V_1 + V_2 + V_3$, and $\lambda = \pi^{-1}$.[18] This specification is intended to reflect a situation where an individual can choose X with knowledge of his unobservable V, rendering X dependent on V and therefore endogenous. The probabilities $P_0[Y = y \wedge X = x_k|z]$ in the examples below were computed via numerical integration in Mathematica using the NIntegrate function.[19]

The scalar instrumental variable takes two values, -1 and $+1$. The parameters in the model for X are set to $d_2 = 1/\sqrt{2}$ with $d_1 = 1$ 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 (I), $\mathcal{X} = \{-1, 1\}$, 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 $c_3 = 1/2$.

We begin by reporting both fully nonparametric and parametric bounds on the CCPs $\wp(x, y)$ in Figure 4, which give a feel for the relative identifying power of the parametric restrictions. The nonparametric bounds in each panel of the figure correspond to the CCP bounds of (3.11) and are illustrated with solid boundaries. The parametric bounds, shown with dashed boundaries, correspond to those of Corollary 2. The panel in the top row of the figure depicts these bounds at both support points for X when $K = 2$, for both the weaker and the stronger instrument cases. The panels in the bottom row provide bounds for the case where $K = 4$, where we illustrate the weak and the strong instrument cases separately so as to keep the panels from becoming overcrowded. Each panel illustrates joint regions for $\wp(x, 1)$ and $\wp(x, 2)$; the corresponding value for $\wp(x, 3)$ can be computed for any such point in these regions as $\wp(x, 3) = 1 - \wp(x, 1) - \wp(x, 2)$.

In the top-left panel of Figure 4 with binary X, the nonparametric bounds for $\wp(x, 1)$ and $\wp(x, 2)$ have a triangular shape.[20] Imposing the parametric structure cuts away part of these triangles, dispensing with higher values for $\wp(-1, 1)$ and lower values for $\wp(-1, 2)$, while discarding lower values for $\wp(1, 1)$ and higher values for $\wp(1, 2)$. The effect of the stronger instrument is more pronounced for the probabilities with $x = -1$. The bounds on the CCPs for $x = -1$ $(+1)$ are drawn in the upper (lower) part of the top panel in Figure 4. It is worth noting that in this binary X model, the parametric structure is only restrictive through the specification for the distribution of unobserved heterogeneity.

The bottom panels of Figure 4 depict designs where X has four points of support. In these figures the nonparametric bounds on choice probabilities are quite wide for intermediate values of x but significantly more informative at the extreme values of x.

[18]The values of λ and d_2 are chosen to ensure that the variance of X^* given Z is 1, the same as it was in a related design that appeared in the on-line working paper Chesher, Rosen, and Smolinski (2011), so as to enable transparent comparison to identified sets from that design. The illustrations in that version pertain to the same setup, but where $\lambda = 0$, in which case X is, in fact, exogenous, though this is not assumed by the econometrician. The resulting bounds are qualitatively similar to those reported here and have thus been omitted from this version.

[19]Computational details for probabilities $P_V[\mathcal{S}]$ can be found in Chesher, Rosen, and Smolinski (2011).

[20]The color figure in the pdf version of this article, available on-line, is easier to read and interpret.

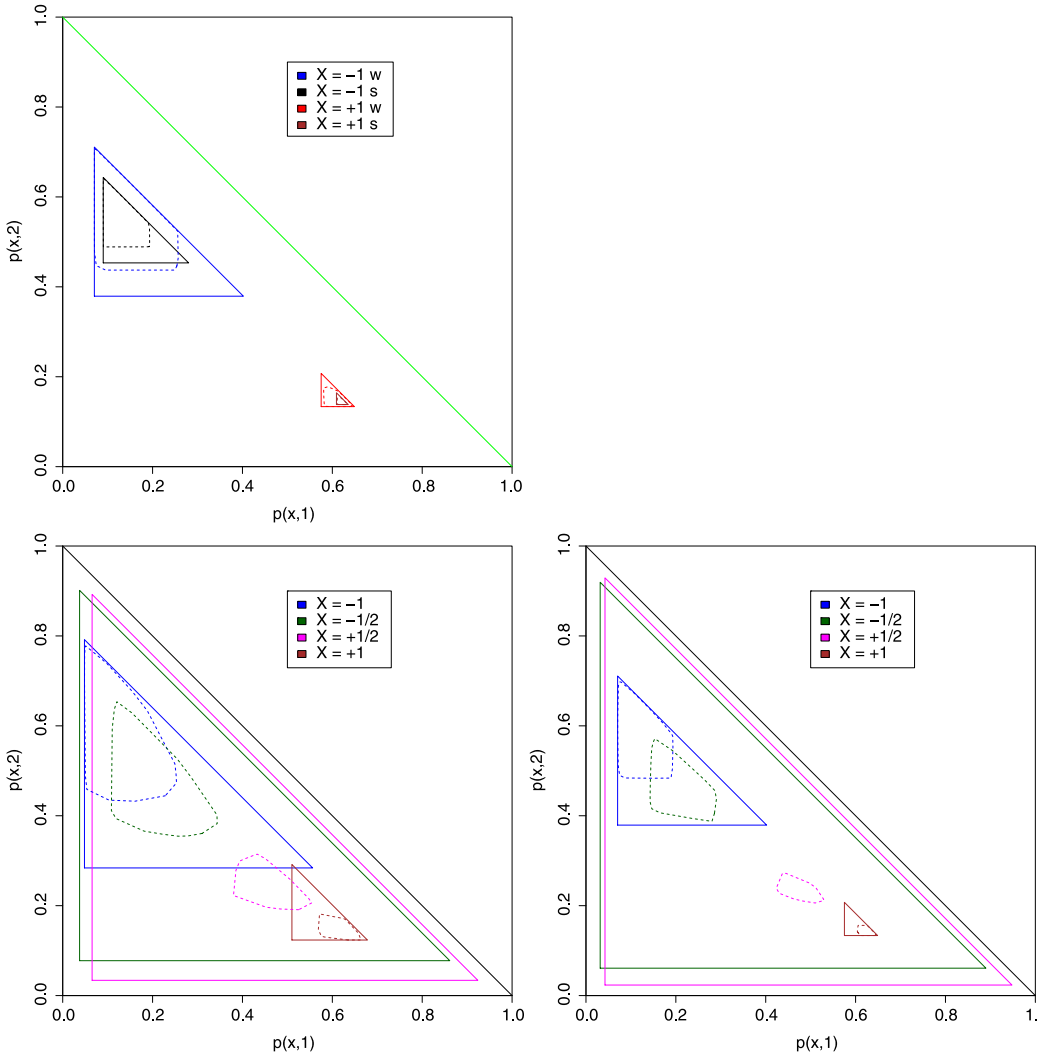

FIGURE 4. Nonparametric (solid) and parametric (dashed) bounds for counterfactual choice probabilities. The top-left panel provides illustrations for $\mathcal{X} = \{-1, 1\}$ for both a weaker (w) and a stronger (s) instrument, and the bottom two panels provide illustrations for $\mathcal{X} = \{-1, -1/2, 1/2, 1\}$ for a weaker and a stronger instrument, respectively.

The extreme value -1 $(+1)$ of the instrumental variable delivers the sets located in the upper (lower) parts of the lower two panels in Figure 4. In the nonparametric cases the intermediate values for the endogenous variable $x = \pm 1/2$ deliver the large triangular sets occupying most of the lower triangle of the lower panels of Figure 4. In the parametric case the identified set for the CCPs for these values are smaller and lie between the sets for the values $x = \pm 1$. This is due to the ordered response nature of the specification for X, where, in particular, $Z = 1$ has relatively high predictive power for $X = 1$, and $Z = -1$ has relatively high predictive power for $X = -1$. With more support points for Z and richer variation in the conditional probability of intermediate realizations of X, the

bounds for $x = -1/2$ and $x = 1/2$ could shrink substantially. Unsurprisingly, the stronger instrument design leads to tighter identified sets for the probabilities when $x = \pm 1$, due to increases in $P_0[X = 1|Z = 1]$ and $P_0[X = -1|Z = -1]$.

The parametric bounds on CCPs in the bottom panels are much narrower than the nonparametric ones, especially for the intermediate x values $x = \pm 1/2$. With four support points for X, linear utility is now a substantive restriction, leading to a decrease in the volume of the identified sets for $\{\wp(x, y), y \in \mathcal{Y}\}$ relative to the nonparametric bounds. The functional form for the utility function now carries implications for choice probabilities *across* different values of x. Consider, for example, the first orientation in the binary X example illustrated in Figure 1. Under that parameter configuration, whenever an individual with $X = x_1$ is observed to have chosen $Y = 2$, we can conclude that he also would have chosen $Y = 2$ if his X had instead been equal to x_2, holding V fixed. In the example illustrated in the bottom panels of Figure 4 with four support points, there are a number of such relationships implied by any value of the utility parameters. Thus, bounds on the CCPs at $x = \pm 1$ also serve to narrow the bounds on CCPs at $x = \pm 1/2$, whereas in the fully nonparametric model, they did not.

We now turn to identified sets for the parameters of the utility functions when the restrictions of the parametric conditional logit model are imposed. Figure 5 shows two-dimensional projections of the four-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 5. Cases I.B, II.A, and II.B are illustrated in Figures 6, 7, and 8.

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

In each case, three sets are drawn. The inner set is the identified set obtained using all the core-determining inequalities of Theorem 2. The outer set is the outer region obtained using the $3K$ inequalities

$$\frac{\exp(a_{0y} + a_{1y}x)}{1 + \sum_{y'=1}^{2} \exp(a_{0y'} + a_{1y'}x)} \geq \max_{z \in \mathcal{Z}} P_0[Y = y \wedge X = x|z], \quad y \in \{1, 2, 3\}, x \in \mathcal{X}, \tag{4.1}$$

implied by (3.11). Since, as shown in McFadden (1974), the logarithms of the choice probabilities on the left-hand side of (4.1) are concave functions of the parameters $a \equiv (a_{01}, a_{11}, a_{02}, a_{12})$, these inequalities define a convex set. The intermediate set is the set obtained using $3K$ inequalities in which the left-hand sides are as in (4.1) but the right-hand sides take into account 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 (4.1) with no change in the values on the left-hand sides. This set cannot be guaranteed to be convex because the identity of the values x' that

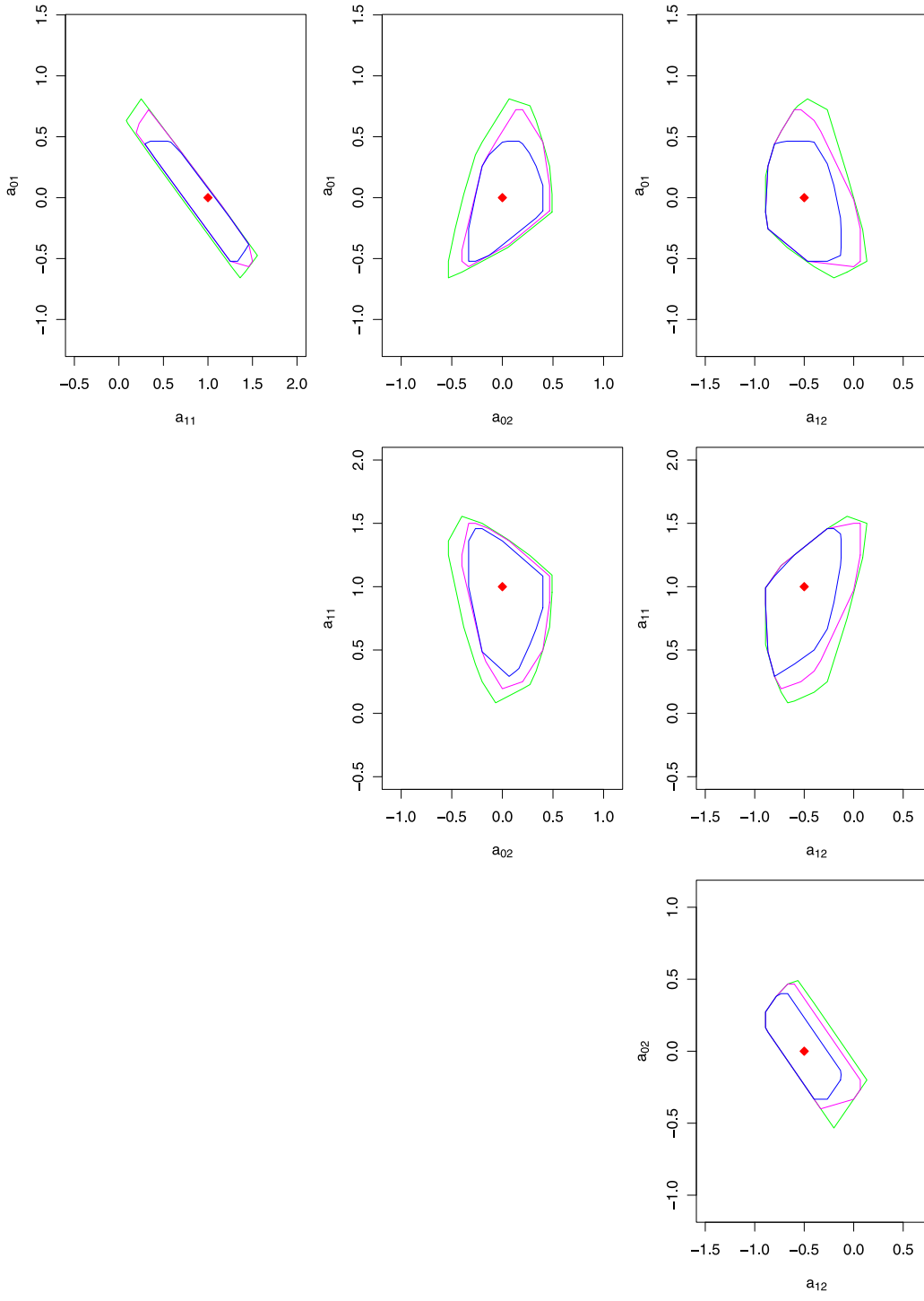

FIGURE 5. Case I.A. Two-dimensional projections of the identified set and two outer regions, $M = 3$, $K = 2$, weaker instrument.

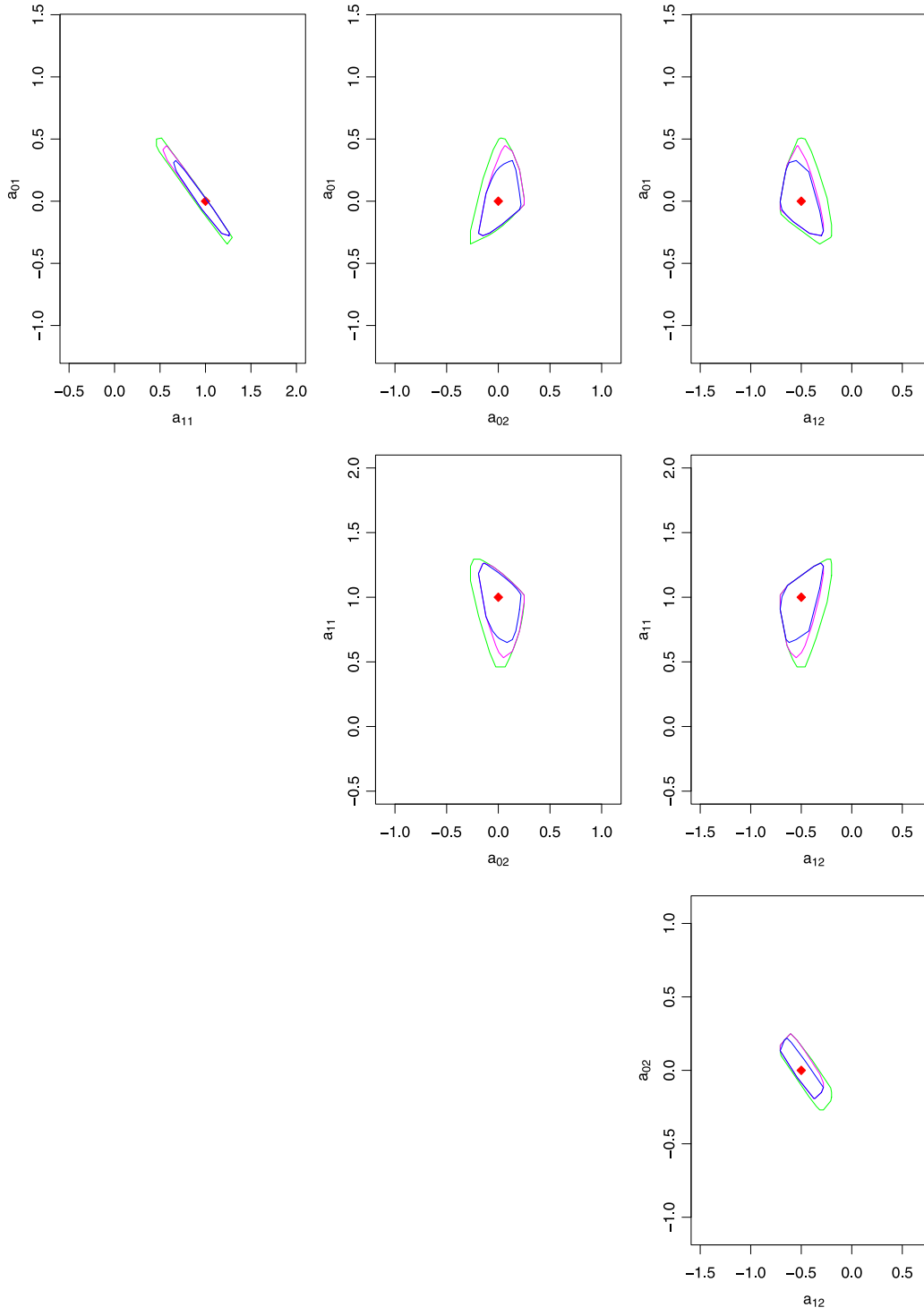

FIGURE 6. Case I.B. Two-dimensional projections of the identified set and two outer regions, $M = 3$, $K = 2$, stronger instrument.

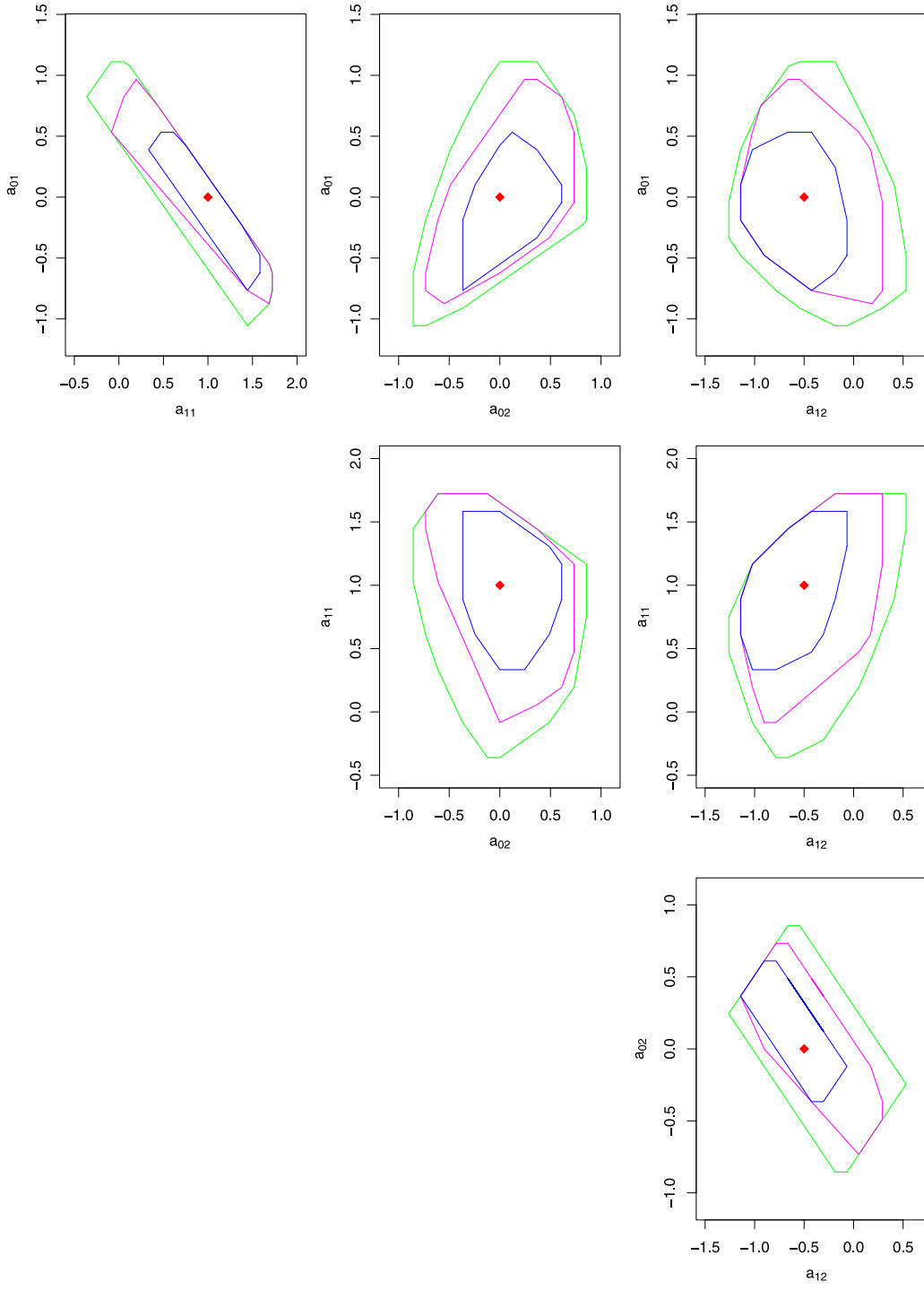

FIGURE 7. Case II.A. Two-dimensional projections of the identified set and two outer regions, $M = 3$, $K = 4$, weaker instrument.

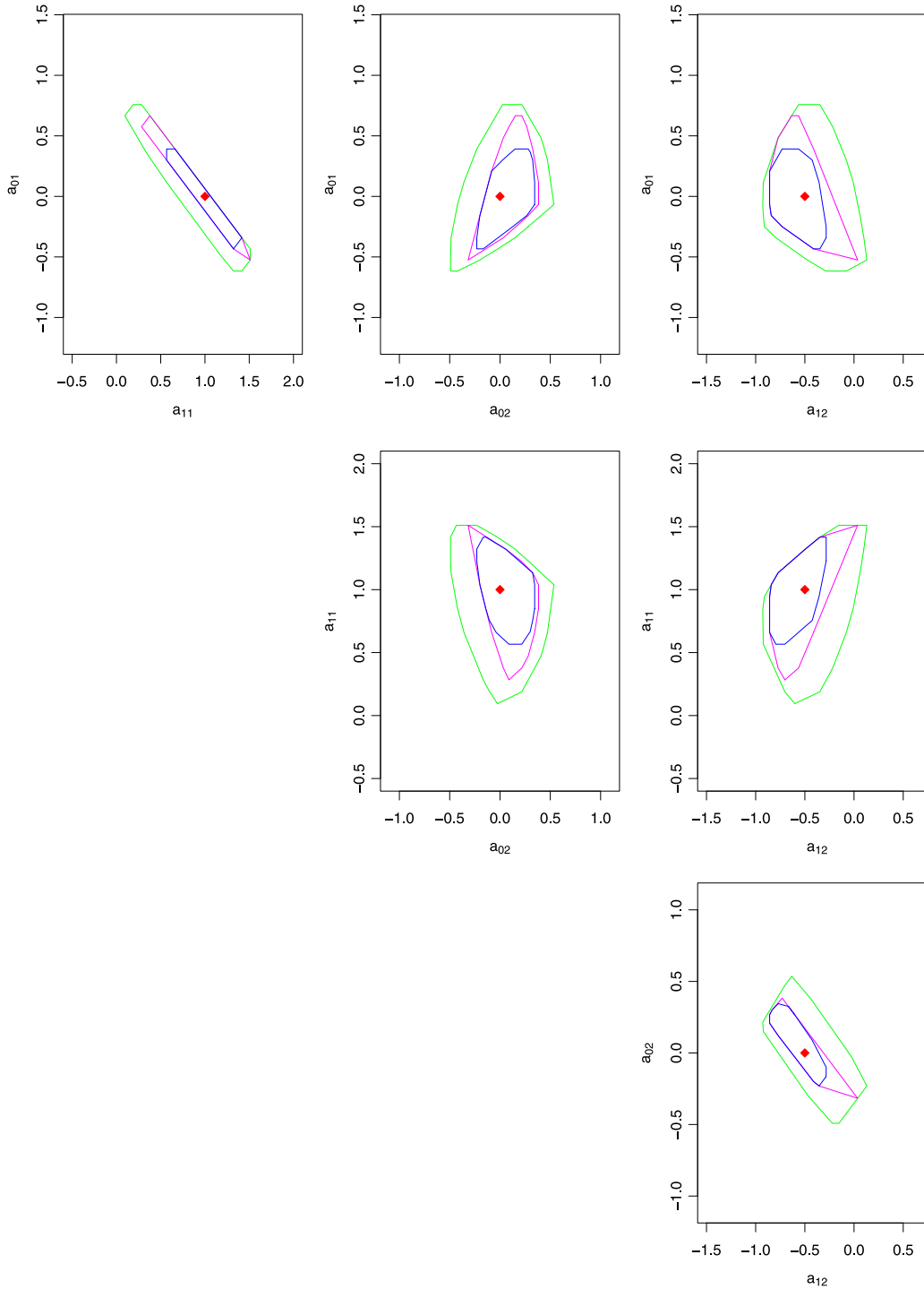

FIGURE 8. Case II.B. Two-dimensional projections of the identified set and two outer regions, $M = 3$, $K = 4$, stronger instrument.

are involved in subset relationships depends on the relative signs and magnitudes of the parameters a_{11} and a_{12}.

In all four cases examined, the calculations suggest that all the two-dimensional projections are convex. Accordingly, boundaries drawn are the convex hulls of the points on the grids that are calculated to lie in the each of the projected two-dimensional sets. In each panel of the figure, the solid diamond locates the parameter value that generates the probability distributions used in this analysis.

The IV model is quite informative for utility parameters. For example, the slope coefficients can be signed in the sense that all values of a_{11} and a_{12} in the identified set have $a_{11} > 0$ and $a_{12} < 0$. This also holds true for some, but not all, of the outer regions. Comparing Figure 5 with Figure 6 ($K = 2$) and comparing Figure 7 with Figure 8 ($K = 4$), it is clear that the identified set and the outer regions are much smaller in the stronger instrument case, as expected and as previously noted in examination of the bounds on choice probabilities reported in Figure 4.

The sets in Figure 5 ($K = 2$) are substantially smaller than those in Figure 7 ($K = 4$), as are those in Figure 6 relative to those in Figure 8. Again as discussed with regard to the CCP bounds, this seems to occur 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. Even so, Figure 4 shows that the implied bounds on CCPs can remain informative. The bounds on both the utility parameters and the CCPs are sensitive to changes in the support of the instrumental variable and to changes in the underlying specification of the relationship between potentially endogenous X and the instrumental variable Z.

The outer regions 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 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 that define the extreme outer region.

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. 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 that, 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 being 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 that 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. In addition, 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 that 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 and Rosen (2013), in which we extend the methods employed here to other IV models in which there are many unobservables in structural functions and possibly continuous outcomes.

Appendix A: Proofs

Proof of Theorem 1. The set $\mathcal{D}^0(\mathcal{Z})$ contains all $(u, P_V) \in \mathcal{U} \times \mathcal{P}_V$ that satisfy for all $\mathcal{S} \in \mathsf{F}(\mathcal{V})$,

$$P_0\big[\mathcal{T}_v(Y, X; u) \subseteq \mathcal{S} | Z = z\big] \leq P_V(\mathcal{S})$$

for almost every $z \in \mathcal{Z}$. The development in the text shows that all admissible (u, P_V) 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 $(u, P_V) \in \mathcal{D}^0(\mathcal{Z})$ 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 the proof of Theorem 2.1 in Beresteanu, Molchanov, and Molinari (2012) to show that the containment functional inequalities of (3.6) 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 1, we have that

$$\forall \mathcal{S} \in \mathsf{F}(\mathcal{V}), \quad P_0\big[\mathcal{T}_v(Y, X; u) \subseteq \mathcal{S} | Z = z\big] \leq P_V(\mathcal{S}) \tag{A.1}$$

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

$$P_0\big[\mathcal{T}_v(Y, X; u) \subseteq \mathcal{S} | Z = z\big] = 1 - P_0\big[\mathcal{T}_v(Y, X; u) \cap \mathcal{S}^c \neq \emptyset | Z = z\big],$$

it follows that (A.1) holds if and only if

$$\forall \mathcal{S} \in \mathsf{F}(\mathcal{V}), \quad P_0\big[\mathcal{T}_v(Y, X; u) \cap \mathcal{S}^c \neq \emptyset | Z = z\big] \geq P_V\big(\mathcal{S}^c\big)$$

or, equivalently,

$$\forall \mathcal{S} \in \mathsf{G}(\mathcal{V}), \quad P_0\big[\mathcal{T}_v(Y, X; u) \cap \mathcal{S} \neq \emptyset | Z = z\big] \geq P_V(\mathcal{S}),$$

where $\mathsf{G}(\mathcal{V})$ here denotes 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

$$\forall \mathcal{S} \in \mathsf{K}(\mathcal{V}), \quad P_0\big[\mathcal{T}_v(Y, X; u) \cap \mathcal{S} \neq \emptyset | Z = z\big] \geq P_V(\mathcal{S}),$$

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 1. This implies that conditional on $Z = z$, there exist random variables $(\tilde{Y}, \tilde{X})$ defined on the same probability space with $\tilde{V} \in \mathcal{T}_v(\tilde{Y}, \tilde{X}; u)$ and $(\tilde{Y}, \tilde{X})$ 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 $(\tilde{Y}, \tilde{X})$ conditionally distributed $F^0_{YX|Z}$ a.e. $z \in \mathcal{Z}$ so that Restriction A.2 is satisfied. □

The proof of Corollary 1 is identical to the proof of Theorem 1 on replacing V with W and P_V with P_W.

The proof of Corollary 2 is a direct consequence of Theorem 1.

Proof of Corollary 3. Consider any $u \in \mathcal{U}$. For each $x \in \mathcal{X}$, let the conditional distribution $P_{V|X}(\cdot|x)$ of V given $X = x$ be absolutely continuous with respect to Lebesgue measure, with everywhere positive density on $\mathrm{Supp}(V|X = x)$ such that for all $y \in \mathcal{Y}$, $P_{V|X}(\mathcal{T}_v(Y, x; u)|x) = P_0[Y = y|x]$. Because V and X are allowed to be arbitrarily correlated, any such conditional distributions for $V|X$ coupled with u produce the observed distributions of $Y|X$. The inequalities that define the identified set hold, since for any test set $\mathcal{S}$,

$$\begin{aligned}
&P_0\big[\mathcal{T}_v(Y, X; u) \subseteq \mathcal{S}|z\big] \\
&\quad = P_0\big[\mathcal{T}_v(Y, X; u) \subseteq \mathcal{S}\big] \\
&\quad = \int_{x\in\mathcal{X}} \left(\sum_{y\in\mathcal{Y}} 1\big[\mathcal{T}_v(y, x; u) \subseteq \mathcal{S}\big] \times P_0[Y = y|x] \right) dF^0_X(x) \\
&\quad = \int_{x\in\mathcal{X}} \left(\sum_{y\in\mathcal{Y}} 1\big[\mathcal{T}_v(y, x; u) \subseteq \mathcal{S}\big] \times P_{V|X}\big(\mathcal{T}_v(y, x; u)|x\big) \right) dF^0_X(x) \\
&\quad \leq \int_{x\in\mathcal{X}} P_{V|X}(\mathcal{S}|x)\, dF^0_X(x) \\
&\quad = P_V(\mathcal{S}),
\end{aligned}$$

where the inequality holds because the sets $\mathcal{T}_v(y, x; u)$ have zero measure intersection with respect to $P_{V|X}(\cdot|x)$. □

Proof of Lemma 1. (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 A.5, 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,

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

Likewise utility from any alternative $k \neq y$ is weakly lower at $V = v^*$ than at either of v or v'. Restriction A.5 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 points. 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.[21]

(ii) If Restriction A.5* 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.[22]

(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 nonempty. Since $\mathcal{T}_w(y, x; u)$ contains its interior and any nonempty 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 $\mathcal{T}_v(y, x; u)$. □

Proof of Theorem 2. We provide the proof for the more general case where Restrictions A.1–A.5 hold with regard to the characterization (3.6). We separate the proof into two cases, depending on whether or not the set

$$\mathcal{Z}^{\emptyset} \equiv \{z \in \mathcal{Z} : P_0[\mathcal{T}_v(Y, X; u) = \emptyset | Z = z] > 0\}$$

has positive measure Z and, equivalently, on whether $\mathcal{T}_v(Y, X; u)$ is empty with positive probability. The proof for the characterization (3.7), where, in addition, Restriction A.5* holds, follows identical steps, replacing V with W.

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

$$0 < P_0[\mathcal{T}_v(Y, X; u) \subseteq \emptyset | Z = z] \leq P_V(\emptyset) = 0,$$

[21] See, for example, 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.

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

which is a contradiction.

Case 2: Again fix $(u, P_V) \in \mathcal{U} \times \mathcal{P}_V$ and now suppose that $\mathcal{Z}^{\emptyset}$ 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 A.1, which requires that the support of $V|(X = x, Z = z)$ is open, in conjunction with Restriction A.5, which requires for all $(y, x) \in \mathrm{Supp}(Y, X)$ 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 $\mathcal{S}$, let $\mathsf{C}_{\mathcal{S}}(u)$ denote the collection of sets on the support of $\mathcal{T}_v(Y, X; u)$ that are subsets of $\mathcal{S}$. Let

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

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

$$P_0\big[\mathcal{T}_v(Y, X; u) \subseteq \mathcal{S} | Z = z\big] = P_0\big[\mathcal{T}_v(Y, X; u) \subseteq \mathcal{G}_{\mathcal{S}}(u) | Z = z\big].$$

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 $\mathcal{S}$,

$$P_0\big[\mathcal{T}_v(Y, X; u) \subseteq \mathcal{G}_{\mathcal{S}}(u) | Z = z\big] \leq P_V\big(\mathcal{G}_{\mathcal{S}}(u)\big) \leq P_V(\mathcal{S}),$$

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

(ii) We now show that the inequalities associated with those sets $\mathcal{G}_{\mathcal{S}}(u)$ such that Condition (ii) does not hold are redundant. Define

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

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

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

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

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

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

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

so that $\mathcal{G}_{\mathcal{S}}(u) = \bigcup_{j=1}^{J} \mathcal{G}_{\mathcal{S},j}(u)$. 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 $\mathcal{Z}^{\emptyset}$ has zero

measure, by the above reasoning. Therefore, it cannot be contained in both $\mathcal{G}_{\mathcal{S},j}(u)$ and $\mathcal{G}_{\mathcal{S},k}(u)$ for any $j \neq k$ since $\mathcal{G}^0_{\mathcal{S},j}(u) \cap \mathcal{G}^0_{\mathcal{S},k}(u) = \emptyset$. Thus

$$P_0\big[\mathcal{T}_v(Y, X; u) \subseteq \mathcal{G}_{\mathcal{S}}(u) | Z = z\big] = \sum_{j=1}^{J} P_0\big[\mathcal{T}_v(Y, X; u) \subseteq \mathcal{G}_{\mathcal{S},j}(u) | Z = z\big] \tag{A.2}$$

and

$$P_V\big(\mathcal{G}_{\mathcal{S}}(u)\big) = \sum_{j=1}^{J} P_V\big(\mathcal{G}_{\mathcal{S},j}(u)\big). \tag{A.3}$$

Therefore,

$$\forall j \in \{1, \dots, J\}, \quad P_0\big[\mathcal{T}_v(Y, X; u) \subseteq \mathcal{G}_{\mathcal{S},j}(u) | Z = z\big] \leq P_V\big(\mathcal{G}_{\mathcal{S},j}(u)\big)$$

implies

$$\sum_{j=1}^{J} P_0\big[\mathcal{T}_v(Y, X; u) \subseteq \mathcal{G}_{\mathcal{S},j}(u) | Z = z\big] \leq \sum_{j=1}^{J} P_V\big(\mathcal{G}_{\mathcal{S},j}(u)\big),$$

and so by (A.2) and (A.3),

$$P_0\big[\mathcal{T}_v(Y, X; u) \subseteq \mathcal{G}_{\mathcal{S}}(u) | Z = z\big] \leq P_V\big(\mathcal{G}_{\mathcal{S}}(u)\big). \qquad \square$$

References

Andrews, D. W. K., S. T. Berry, and P. Jia (2004), "Confidence regions for parameters in discrete games with multiple equilibria, with an application to discount chain store location." Working paper, Yale University. [176]

Andrews, D. W. K. and X. Shi (2013), "Inference based on conditional moment inequalities." *Econometrica*, 81 (2), 609–666. [171]

Artstein, Z. (1983), "Distributions of random sets and random selections." *Israel Journal of Mathematics*, 46 (4), 313–324. [162, 169, 189, 190]

Ben-Akiva, M. (1973), *Structure of Passenger Travel Demand Models*. Ph.D. dissertation, MIT. [158]

Beresteanu, A., I. Molchanov, and F. Molinari (2011), "Sharp identification regions in models with convex moment predictions." *Econometrica*, 79 (6), 1785–1821. [162, 164]

Beresteanu, A., I. Molchanov, and F. Molinari (2012), "Partial identification using random set theory." *Journal of Econometrics*, 166 (1), 17–32. [162, 169, 189]

Beresteanu, A. and F. Molinari (2008), "Asymptotic properties for a class of partially identified models." *Econometrica*, 76 (4), 763–814. [171]

Berry, S. and P. Haile (2009), "Nonparametric identification of multiple choice demand models with heterogeneous consumers." Working Paper w15276, NBER. [161]

Berry, S. and P. Haile (2010), "Identification in differentiated markets using market level data." Working Paper w15641, NBER. [161]

Berry, S., J. Levinsohn, and A. Pakes (1995), "Automobile prices in market equilibrium." *Econometrica*, 63 (4), 841–890. [161]

Berry, S., J. Levinsohn, and A. Pakes (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. [161]

Berry, S. T. (1994), "Estimating discrete choice models of product differentiation." *RAND Journal of Economics*, 25 (2), 242–262. [161]

Bugni, F. (2010), "Bootstrap inference for partially identified models defined by moment inequalities: Coverage of the identified set." *Econometrica*, 78 (2), 735–753. [171]

Canay, I. (2010), "EL inference for partially identified models: Large deviations optimality and bootstrap validity." *Journal of Econometrics*, 156 (2), 408–425. [171]

Chernozhukov, V. and C. Hansen (2005), "An IV model of quantile treatment effects." *Econometrica*, 73, 245–261. [159]

Chernozhukov, V., H. Hong, and E. Tamer (2007), "Estimation and confidence regions for parameter sets in econometric models." *Econometrica*, 75 (5), 1243–1284. [171]

Chernozhukov, V., S. Lee, and A. Rosen (2013), "Intersection bounds: Estimation and inference." *Econometrica*, 81 (2), 667–737. [171]

Chesher, A. (2010), "Instrumental variable models for discrete outcomes." *Econometrica*, 78 (2), 575–601. [159, 162]

Chesher, A. and A. Rosen (2012a), "An instrumental variable random coefficients model for binary outcomes." Working Paper CWP34/12, CeMMAP. [162]

Chesher, A. and A. Rosen (2012b), "Simultaneous equations models for discrete outcomes: Coherence, completeness, and identification." Working Paper CWP21/12, CeMMAP. [164]

Chesher, A. and A. Rosen (2013), "Generalized instrumental variable models." In preparation. [159, 162, 189]

Chesher, A., A. Rosen, and K. Smolinski (2011), "An instrumental variable model of multiple discrete choice." Working Paper CWP39/11, CeMMAP. [177, 181]

Chesher, A. and K. Smolinski (2010), "Sharp identified sets for discrete variable IV models." Working Paper CWP11/10, CeMMAP. [162]

Chiappori, P.-A., I. Komunjer, and D. Kristensen (2011), "On the nonparametric identification and estimation of multiple choice models." Working paper, Columbia University. [161]

Domencich, T. and D. McFadden (1975), *Urban Travel Demand: A Behavioural Analysis*. North-Holland, Amsterdam. [158]

Ekeland, I., A. Galichon, and M. Henry (2010), "Optimal transportation and the falsifiability of incompletely specified economic models." *Economic Theory*, 42, 355–374. [162]

Fox, J. T. and A. Gandhi (2009), "Identifying heterogeneity in economic choice models." Working Paper w15147, NBER. [161]

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. [171]

Galichon, A. and M. Henry (2011), "Set identification in models with multiple equilibria." *Review of Economic Studies*, 78 (4), 1264–1298. [162, 163, 164, 167, 173]

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. [158, 166]

Henry, M., R. Meango, and M. Queyranne (2011), "Combinatorial bootstrap inference in partially identified incomplete structural models." Working paper, University of Montreal. [171]

Komarova, T. (2007), "Binary choice models with discrete regressors: Identification and misspecification." Working paper, LSE. [162]

Koning, R. and G. Ridder (2003), "Discrete choice and stochastic utility maximization." *Econometrics Journal*, 6 (1), 1–27. [177]

Lewbel, A. (2000), "Semiparametric qualitative response model estimation with unknown heteroscedasticity or instrumental variables." *Journal of Econometrics*, 97, 145–177. [161]

Magnac, T. and E. Maurin (2008), "Partial identification in binary models: Discrete regressors and interval data." *Review of Economic Studies*, 75 (3), 835–864. [162]

Manski, C. F. (2007), "Partial identification of counterfactual choice probabilities." *International Economic Review*, 48 (4), 1393–1410. [162]

Manski, C. F. and E. Tamer (2002), "Inference on regressions with interval data on a regressor or outcome." *Econometrica*, 70 (2), 519–546. [162]

Matzkin, R. (1993), "Nonparametric identification and estimation of polychotomous choice models." *Journal of Econometrics*, 58, 137–168. [161]

Matzkin, R. (2008), "Identification in nonparametric simultaneous equations models." *Econometrica*, 76, 945–978. [161]

Matzkin, R. (2012), "Identification in nonparametric limited dependent variable models with simultaneity and unobserved heterogeneity." *Journal of Econometrics*, 166 (1), 106–115. [161]

McFadden, D. (1974), "Conditional logit analysis of qualitative choice behavior." In *Frontiers in Econometrics* (P. Zarembka, ed.), Academic Press, New York. [157, 158, 160, 163, 166, 172, 183]

McFadden, D. (1978), "Modelling the choice of residential location." In *Spatial Interaction Theory and Residential Location* (A. Karlvist, L. Ludvist, F. Snickars, and J. Weibull, eds.), 75–96, North-Holland, Amsterdam. [158]

Menzel, K. (2009), "Estimation and inference with many weak moment inequalities." Working paper, MIT. [171]

Molchanov, I. S. (2005), *Theory of Random Sets*. Springer, London. [169, 189, 190]

Newey, W. K. and J. L. Powell (2003), "Instrumental variable estimation of nonparametric models." *Econometrica*, 71, 1565–1578. [159]

Norberg, T. (1992), "On the existence of ordered couplings of random sets—With applications." *Israel Journal of Mathematics*, 77, 241–264. [189, 190]

Petrin, A. and K. Train (2010), "A control function approach to endogeneity in consumer choice models." *Journal of Marketing Research*, 47, 1–11. [161]

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. [171]

Rosen, A. M. (2008), "Confidence sets for partially identified parameters that satisfy a finite number of moment inequalities." *Journal of Econometrics*, 146, 107–117. [171]

Sutherland, W. A. (2009), *Introduction to Metric and Topological Spaces*. Oxford University Press, New York. [191]

Train, K. (2009), *Discrete Choice Methods With Simulation*, second edition. Cambridge University Press, Cambridge. [167]

Submitted December, 2011. Final version accepted September, 2012.