Quantitative Economics 10 (2019), 105–144 1759-7331/20190105

Partial identification by extending subdistributions

Alexander Torgovitsky
Department of Economics, University of Chicago

I show that sharp identified sets in a large class of econometric models can be characterized by solving linear systems of equations. These linear systems determine whether, for a given value of a parameter of interest, there exists an admissible joint distribution of unobservables that can generate the distribution of the observed variables. The joint distribution of unobservables is not required to satisfy any parametric restrictions, but can (if desired) be assumed to satisfy a variety of location, shape, and/or conditional independence restrictions. To prove sharpness of the characterization, I generalize a classic result in copula theory concerning the extendibility of subcopulas to show that related objects—termed subdistributions—can be extended to proper distribution functions. I describe this characterization argument as partial identification by extending subdistributions, or PIES. One particularly attractive feature of PIES is that it focuses directly on the sharp identified set for a parameter of interest, such as an average treatment effect, without needing to construct the identified set for the entire model. I apply PIES to univariate and bivariate binary response models. A notable product of the analysis is a method for characterizing the sharp identified set for the average treatment effect in Manski's (1975, 1985, 1988) semiparametric binary response model.

Keywords. Partial identification, maximum score, bivariate probit, copulas, linear programming, discrete choice, semiparametric, endogeneity.

JEL classification. C14, C20, C51.

1. Introduction

Nonlinear models are common in empirical economics. Standard implementations of many classes of nonlinear models, such as discrete choice models, rely on parametric distributional assumptions to ensure point identification. Relaxing these assumptions raises the possibility of partial identification. A fundamental problem in studying the empirical content of a partially identified model is obtaining a tractable characterization of an identified set for the parameter of interest. To be fully informative, such a

Alexander Torgovitsky: torgovitsky@uchicago.edu

I am thankful to Marc Henry, Chuck Manski, Azeem Shaikh, Frank Wolak, and four anonymous referees for providing helpful comments. This paper was presented at Frontiers of Econometrics (a conference in honor of Donald W.K. Andrews), the 2015 CEME Interactions conference, the 2015 conference on Inverse Problems at Northwestern University, as well as seminars at Michigan St., Cornell, and Boston University. My thanks to those audiences. This research was supported in part by National Science Foundation Grant SES-1530538.

 https://doi.org/10.3982/QE634

characterization should be for the *sharp* identified set, that is, the set that exhausts all implications of the model and data.

Early work on partially identified models approached this characterization problem with a two-step argument. First, bounds are proposed for the parameter of interest. Second, the bounds are shown to be sharp by establishing the existence of model parameters that achieve these bounds, are consistent with the distribution of observables, and satisfy the researcher's assumptions. While this type of argument has provided many useful results for nonparametric models (e.g., Manski (1989, 1994), Hansen and Jagannathan (1991)), it can be analytically challenging to apply it to more complex semiparametric models. In particular, establishing sharpness is difficult when the relationship between the parameters of the model and the distribution of the observed data is complicated and/or when these parameters are assumed to satisfy many properties. This state of affairs characterizes many desirable semiparametric generalizations of widely used nonlinear econometric models.

A concrete example of such a model is a univariate or bivariate binary response model with a linear index function(s), but without any parametric distributional assumptions for the latent variables. Fully parametric versions of these models are widely used in the empirical literature. Examples include Neal (1997) and Grogger and Neal (2000), who used bivariate probit models to estimate the effect of Catholic school on high school graduation; Alesina and Fuchs-Schündeln (2007), who estimated bivariate probit models of the effect of receiving government transfers on views of the proper role of the state; Karlan and Zinman (2008), who used a probit model to estimate the demand for loan applications; Thornton (2008), who used a bivariate probit model to estimate the impact of learning HIV status on sexual behavior; and Ashraf, Field, and Lee (2014), who used a bivariate probit model to estimate the effect on fertility of providing access to contraception.

Point identification of the models used in these papers depends critically on the assumption that the latent terms follow a distribution in a parametric family (usually Gaussian) that has been correctly specified by the researcher. Manski (1975) provided an important early criticism of models with this property. This criticism is echoed in more recent discussions by Angrist (2001) and Angrist and Pischke (2009). Those authors go so far as to advocate using linear models even in cases with known nonlinearities, so as to avoid (explicitly) parameterizing the distributions of latent variables. Altonji, Elder, and Taber (2005) provided evidence that both fully parametric models and misspecified linear models can yield highly misleading empirical results. These concerns motivate studying identification in versions of these models that drop the parametric distributional assumptions on the latent terms.

This paper provides a new general method for constructing tractable characterizations of sharp identified sets in models without parametric distributional assumptions. The method is based on a fundamental result in copula theory due to Sklar (1959, 1996), which shows that any subcopula can be extended to a copula. I generalize this result to show that any L-dimensional *subdistribution*—that is, any function with the shape properties of an L-dimensional distribution function, but defined only on a subset of L-dimensional Euclidean space, $\mathbb{R}^L$—can be extended to a proper distribution function

that is defined everywhere on $\mathbb{R}^L$. The main contribution of this paper is to show how this subdistribution extension result can be usefully applied to provide tractable, sharp characterizations of identified sets in a wide variety of econometric models. I refer to this argument, and the computational method that it justifies, as partial identification by extending subdistributions, or PIES.

A distinguishing feature of PIES is the ability to focus directly on characterizing identified sets for scalar functionals of high-dimensional parameters. For example, the average treatment effect (ATE) in a binary response model is a function of both the index coefficients (β) and the distribution function for the unobservable (F). A standard way to construct an identified set for the ATE is to first construct an identified set for (β, F), then form the one-dimensional subset consisting of all values of the ATE that can be obtained as (β, F) varies over this identified set. This projection strategy is computationally intractable in a semiparametric model in which F (an infinite-dimensional object) is not finitely parameterized. PIES circumvents this problem by considering sufficient and necessary conditions for the *existence* of an F that yields a hypothesized value of the ATE.

This paper contributes to a literature on characterizing sharp identified sets in general classes of models. Other procedures that have been proposed include those using random sets (Beresteanu, Molchanov, and Molinari (2011), Chesher and Rosen (2012, 2013, 2014, 2017)), optimal transportation (Ekeland, Galichon, and Henry (2010), Galichon and Henry (2009, 2011, 2013), Henry, Méango, and Queyranne (2015)), and information theory (Schennach (2014)).[1] The primary attraction of PIES relative to these procedures is its approach to the projection problem discussed in the previous paragraph. This allows researchers to bound objects such as the ATE under a variety of different semiparametric restrictions. However, PIES does not apply to some models that feature prominently in this literature. In particular, it does not apply to models that are incomplete, such as entry games without an explicitly specified equilibrium selection function. PIES is most closely related to the discretization approach of Lafférs (2013), although it also applies more generally to settings in which his discretization argument cannot be implemented.[2]

In the next section, I illustrate the PIES argument by using it to characterize the sharp identified set for the ATE in a stylized version of Manski's (1975, 1985, 1988) semiparametric binary response model. In Section 3, I prove the subdistribution extension result, apply it to produce characterizations for sharp identified sets in a general econometric model, and discuss implementation of these characterizations. In Section 4, I apply the general arguments in Section 3 to provide characterizations for sharp identified sets in semiparametric bivariate binary response models. The results here contribute to the literature on identification in binary response instrumental variable models (e.g., Chesher

[1] Note that there are many other papers in the literature that employ computational approaches for *specific* partially identified models. Examples include Hansen, Heaton, and Luttmer (1995), Honoré and Lleras-Muney (2006), Honoré and Tamer (2006), Manski (2007, 2014), Molinari (2008), Kitamura and Stoye (2013), Freyberger and Horowitz (2015), Lafférs (2015), Demuynck (2015), Torgovitsky (2016), and Mogstad, Santos, and Torgovitsky (2017).

[2] For more detail, see the previous working paper version of this paper.

(2010, 2013); Shaikh and Vytlacil (2011)). Some brief concluding remarks are provided in Section 5.

2. PIES in a semiparametric binary response model

Consider the binary response model

$$Y = \mathbb{1}[U > \beta_0 + \beta_1 X], \tag{1}$$

where $Y \in \{0, 1\}$ is an observed binary outcome variable, X is an observed explanatory variable, $\beta \equiv (\beta_0, \beta_1)$ are unknown parameters, and $U \in \mathbb{R}$ is a scalar unobservable variable. For simplicity of exposition, I will assume that $X \in \{0, 1\}$ is binary, although this is not essential to the following discussion.

Let $\mathcal{F}$ denote the set of all conditional distribution functions $F : \overline{\mathbb{R}} \times \{0, 1\} \to [0, 1]$, and let $\mathcal{F}^\dagger$ denote all $F \in \mathcal{F}$ for which $F(0|x) = \frac{1}{2}$ for $x = 0, 1$.[3] Manski (1975, 1985, 1988) studied (1) under the assumption that the conditional distribution of U given X is an element of $\mathcal{F}^\dagger$. This assumption requires U to be exogenous in the sense that its median does not depend on X, however it still allows for heteroscedasticity in U. The choice of 0 as the median is a location normalization. To normalize the scale of β, I assume that $\beta \in \mathcal{B}^\dagger \equiv \{b \in \mathbb{R}^2 : \|b\| = 1\}$.

Manski (1975, 1985, 1988) provided conditions under which β is point identified. These conditions require at least one component of X to have a continuous distribution with sufficient variation. This can be restrictive in general, and in particular is not satisfied in the case considered here of binary X. Horowitz (2009, pp. 100–108) suggested a computational approach that can be used to compute the identified set for β when these conditions do not hold. Komarova (2013) developed Horowitz's approach into a more analytic argument, while Blevins (2015) considered estimation of this identified set.

This type of analysis is sometimes criticized on the grounds that the most natural parameter of interest is not the vector of index coefficients, β, but rather the average effect of X on Y.[4] This average effect (whether interpreted causally or not) is a function of both β and the conditional distribution of U given X, that is, F. The following argument directly addresses this criticism by providing a tractable method for constructing the sharp identified set for the average treatment effect (ATE).

The average treatment effect under (1) is defined as

$$\text{ATE}(\beta, F) \equiv \mathbb{P}_F[U > \beta_0 + \beta_1] - \mathbb{P}_F[U > \beta_0], \tag{2}$$

[3] The notation $\overline{\mathbb{R}} \equiv \mathbb{R} \cup \{\pm\infty\}$ denotes the extended real line.

[4] For example, Angrist and Pischke (2009, p. 201) wrote "…some researchers become distracted by an effort to estimate index coefficients instead of average causal effects. For example, a large literature in econometrics is concerned with the estimation of index coefficients without the need for distributional assumptions. Applied researchers interested in causal effects can safely ignore this work." This point was made early in the literature by Cosslett (1983, p. 767).

where $\mathbb{P}_F$ denotes probability when the conditional distribution of U given X is F. Using the law of iterated expectations, (2) can also be written as

$$\text{ATE}(\beta, F) = \sum_{x\in\{0,1\}} \big[F(\beta_0|x) - F(\beta_0+\beta_1|x)\big]\mathbb{P}[X=x]. \tag{3}$$

To define the identified set for the ATE, first define the identified set for $S=(\beta, F)$ as

$$\mathcal{S}^\star \equiv \big\{S=(\beta,F)\in\mathcal{B}^\dagger\times\mathcal{F}^\dagger : \mathbb{P}_S[Y=0|X=x] = \mathbb{P}[Y=0|X=x], x=0,1\big\}, \tag{4}$$

where $\mathbb{P}[Y=0|X=x]$ denotes the actual population distribution of (Y,X), and $\mathbb{P}_S[Y=0|X=x]$ denotes the same probability that would be obtained if the data were generated by the model with parameters S, that is,

$$\mathbb{P}_S[Y=0|X=x] \equiv \mathbb{P}_F[U\le\beta_0+\beta_1 x|X=x] = F\big(x'\beta|x\big). \tag{5}$$

The identified set for the ATE can then be defined as

$$\mathcal{P}^\star \equiv \text{ATE}\big(\mathcal{S}^\star\big) \equiv \big\{p : p=\text{ATE}(\beta,F) \text{ for some } (\beta,F)\in\mathcal{S}^\star\big\}. \tag{6}$$

Fix a potential value $p\in[-1,1]$ for the ATE. From (3)–(6) and the definition of $\mathcal{F}^\dagger$, it follows that $p\in\mathcal{P}^\star$ if and only if there exist a $\beta\in\mathcal{B}^\dagger$ and an $F\in\mathcal{F}$ such that

$$\begin{aligned}&\sum_{x\in\{0,1\}} \big[F(\beta_0|x) - F(\beta_0+\beta_1|x)\big]\mathbb{P}[X=x] = p\\ &\quad\text{with } F(\beta_0|0) = \mathbb{P}[Y=0|X=0], F(\beta_0+\beta_1|1) = \mathbb{P}[Y=0|X=1],\\ &\quad\text{and } F(0|0) = F(0|1) = \frac{1}{2}.\end{aligned} \tag{7}$$

Determining whether this is the case is an *infinite*-dimensional existence problem, because F, as a conditional distribution function, is an infinite-dimensional object. At the same time, it is clear from (7) that this problem only depends on the values of $F(\cdot|0)$ and $F(\cdot|1)$ on the three point set $\{0,\beta_0,\beta_0+\beta_1\}$. Thus, this infinite-dimensional existence problem can be solved by solving a related *finite*-dimensional existence problem.

The related finite-dimensional problem is whether there exist a $\beta\in\mathcal{B}^\dagger$ and functions $\overline{F}(\cdot|0)$ and $\overline{F}(\cdot|1)$ with domains $\{0,\beta_0,\beta_0+\beta_1\}$ that satisfy the conditions in (7), and which could be *extended* to proper distribution functions. Such an extension would be an element of $\mathcal{F}$ that also satisfies (7). The existence of a $\beta\in\mathcal{B}^\dagger$ for which some such extension exists is both sufficient and necessary for $p\in\mathcal{P}^\star$.

The subdistribution extension lemma proved in the next section provides a tractable method for solving this existence problem. Specifically, the lemma implies that a function $\overline{F}(\cdot|x)$ with domain $\{0,\beta_0,\beta_0+\beta_1\}$ can be extended to a proper distribution function if and only if $\overline{F}(\cdot|x)$ is a weakly increasing function. It follows that $p\in\mathcal{P}^\star$ if and only if there exist a $\beta\in\mathcal{B}^\dagger$ and functions $\overline{F}(\cdot|0)$, $\overline{F}(\cdot|1)$ with domains $\{0,\beta_0,\beta_0+\beta_1\}$ that are weakly increasing and satisfy (7). Given a fixed value of $\beta\in\mathcal{B}^\dagger$, the existence of weakly increasing functions that satisfy (7) is equivalent to the existence of a solution to a linear system of equations, which is a tractable and well-understood problem.

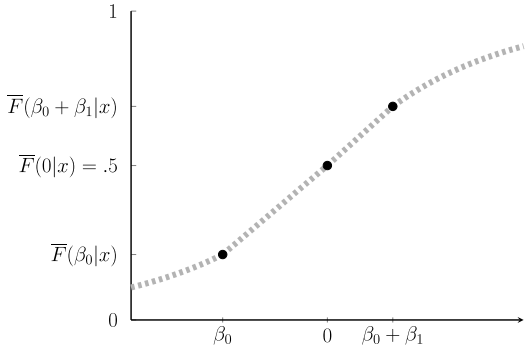

FIGURE 1. A function with domain $\{0, \beta_0, \beta_0+\beta_1\}$ (plotted in black dots) and one possible extension of that function to a distribution function (plotted in gray dashes).

This observation suggests a relatively efficient procedure for characterizing the identified set for the ATE, that is, $\mathcal{P}^\star$. Fix a given $\beta \in \mathcal{B}^\dagger$. Then solve the following linear program:

$$\begin{aligned}
\underline{p}^\star(\beta) \equiv \min_{\overline{F}} \ & \sum_{x\in\{0,1\}} \left[\overline{F}(\beta_0|x) - \overline{F}(\beta_0+\beta_1|x)\right]\mathbb{P}[X=x] \\
\text{subject to:} \quad & \overline{F}(\beta_0|0) = \mathbb{P}[Y=0|X=0], \\
& \overline{F}(\beta_0+\beta_1|1) = \mathbb{P}[Y=0|X=1], \\
& \overline{F}(0|0) = \overline{F}(0|1) = \frac{1}{2}, \\
& \overline{F}(u|x) \in [0,1] \text{ for all } u \in \{0, \beta_0, \beta_0+\beta_1\},\ x \in \{0,1\}, \\
& \overline{F}(\cdot|x) \text{ is increasing for each } x \in \{0,1\}. \qquad (8)
\end{aligned}$$

Let $\overline{p}^\star(\beta)$ denote the optimal value from the analogous maximization problem and construct the interval $[\underline{p}^\star(\beta), \overline{p}^\star(\beta)]$, letting this interval be the empty set if (8) is infeasible. Continue this procedure for all $\beta \in \mathcal{B}^\dagger$. Then the sharp identified set for the ATE can be characterized as $\mathcal{P}^\star = \bigcup_{\beta\in\mathcal{B}^\dagger}[\underline{p}^\star(\beta), \overline{p}^\star(\beta)]$.

The preceding argument is a simplified example of the general method of PIES that is introduced in the next section. The main simplification comes from the dimension of the latent variables. It is intuitive that a univariate function with values in $[0,1]$ can be extended to a proper distribution function if it is weakly increasing; see Figure 1. More surprising is that an analogous statement remains true in higher dimensions. This result, which will be proved in the next section, allows nearly the same argument to be applied to models with multiple latent variables, such as a bivariate binary response model. It is also shown in the next section that PIES can be modified to accommodate different and/or more assumptions on the distribution of unobservables.

It is worth emphasizing that a key advantage of this procedure is that it does not *ignore* the conditional distribution of U given X, that is, F. The ATE is determined in part by F, so treating it as a nuisance parameter is not appropriate. Other studies of

identification for this model, such as those referenced above, have focused exclusively on the identification of the index coefficients, β, without considering F. Perhaps due to this focus, it is not uncommon to see statements like "…we can only learn about relative sizes of the coefficients using the semiparametric approach…" Wooldridge (2010, p. 606). The preceding argument shows that this view is too pessimistic. Sharp bounds on parameters like the ATE can also be obtained using PIES.

3. The theory of PIES

3.1 *The subdistribution extension lemma*

In this section, I develop the subdistribution extension lemma alluded to in the previous section. This requires some definitions and intermediate results.[5] Throughout the discussion, I use $\overline{\mathbb{R}} \equiv \mathbb{R} \cup \{\pm\infty\}$ to denote the extended real line, and use $\overline{\mathbb{R}}^L$ to denote its L-fold Cartesian product for a positive integer L. The first key concept is that of an *L-increasing* function.

Definition 1. Let $\mathcal{U}$ be a subset of $\overline{\mathbb{R}}^L$ such that $\mathcal{U} = \mathcal{U}_1 \times \cdots \times \mathcal{U}_L$ for subsets $\mathcal{U}_l$ of $\overline{\mathbb{R}}$. A function F with domain $\mathcal{U}$ is called *L-increasing* if, for any $u', u'' \in \mathcal{U}$ with $u' \leq u''$ component-wise,

$$\mathrm{Vol}_F(u', u'') \equiv \sum_{u \in \mathrm{Vrt}(u', u'')} \mathrm{sgn}_{(u', u'')}(u) F(u) \geq 0, \tag{9}$$

where $\mathrm{Vrt}(u', u'')$ is the set of $u \in \mathcal{U}$ such that $u_l \in \{u'_l, u''_l\}$ for each l, and

$$\mathrm{sgn}_{(u', u'')}(u) \equiv \begin{cases} 1, & \text{if } u_l = u'_l \text{ for an even number of } l \in \{1, \ldots, L\}, \\ -1, & \text{if } u_l = u'_l \text{ for an odd number of } l \in \{1, \ldots, L\}. \end{cases}$$

The quantity $\mathrm{Vol}_F(u', u'')$ is the *F-volume* of the L-box $[u'_1, u''_1] \times \cdots \times [u'_L, u''_L]$ and the elements of the set $\mathrm{Vrt}(u', u'')$ are the *vertices* of the L-box.

L-increasingness reduces to the standard definition of weakly increasing for $L = 1$.[6] For $L > 1$, L-increasingness requires a function F to assign nonnegative volume to every L-box with vertices in $\mathcal{U}$. This is a key property of distribution functions, for which volumes are interpretable as probabilities of events. However, an L-increasing function must satisfy some additional conditions to be a distribution function. These are laid out in the next definition.

Definition 2. Let $\mathcal{U}$ be a subset of $\overline{\mathbb{R}}^L$ such that $\mathcal{U} = \mathcal{U}_1 \times \cdots \times \mathcal{U}_L$, where $\mathcal{U}_l \subseteq \overline{\mathbb{R}}$ is such that $\{\pm\infty\} \subset \mathcal{U}_l$ for each l. A function F with domain $\mathcal{U}$ is an L-dimensional *subdistribution function* if it satisfies the following three conditions:

2a. F is L-increasing.

[5] The discussion follows treatments by Schweizer and Sklar (1983) and Nelsen (2006).

[6] Note that the definition of $\mathrm{sgn}_{(u', u'')}$ uses the fact that 0 is an even integer.

2b. $F(u) = 0$ for any $u \in \mathcal{U}$ that has at least one component equal to $-\infty$.

2c. $F(+\infty, \ldots, +\infty) = 1$.

An L-dimensional *distribution function* (or, for emphasis, a *proper* L-dimensional distribution function) is an L-dimensional subdistribution function defined on $\mathcal{U} = \overline{\mathbb{R}}^L$.

This definition of a distribution function is standard.[7] The concept of a subdistribution function appears to be novel to this paper.[8] It is a natural counterpart to the concept of a subcopula, which is discussed in Appendix A. The only difference between an L-dimensional subdistribution function and a (proper) L-dimensional distribution function is that a subdistribution function could be defined on a strict subset of $\overline{\mathbb{R}}^L$, whereas a distribution function must be defined on the entirety of $\overline{\mathbb{R}}^L$. As a result, every distribution function is also a subdistribution function, but not conversely. In Figure 1, the gray dashes plot a proper one-dimensional distribution function, whereas a function defined only on the black dots (together with $\pm\infty$) is a one-dimensional subdistribution function.

Every L-dimensional distribution function generates a collection of L one-dimensional distribution functions called the marginal distribution functions. More generally, an L-dimensional subdistribution function generates a collection of L *margins*.

DEFINITION 3. Let F be an L-dimensional subdistribution function with domain $\mathcal{U} = \mathcal{U}_1 \times \cdots \times \mathcal{U}_L$. The lth *margin* of F is defined for each $l = 1, \ldots, L$ as

$$F_l : \mathcal{U}_l \to [0,1] : F_l(u_l) = F(+\infty, \ldots, u_l, \ldots, +\infty). \tag{10}$$

If F is a proper L-dimensional distribution function, then all of its margins are proper one-dimensional distribution functions. The next lemma records the more general fact that if F is an L-dimensional subdistribution function, then all of its margins are one-dimensional subdistribution functions. All proofs for this section are contained in Appendix B.

LEMMA 1. *Let F be an L-dimensional subdistribution function with domain $\mathcal{U} = \mathcal{U}_1 \times \cdots \times \mathcal{U}_L$. Then the lth margin of F is a one-dimensional subdistribution function with domain $\mathcal{U}_l$.*

The subdistribution extension lemma can now be stated as follows.

LEMMA 2 (Subdistribution extension). *Suppose that $\mathcal{U} \subseteq \overline{\mathbb{R}}^L$ can be written as $\mathcal{U} = \mathcal{U}_1 \times \cdots \times \mathcal{U}_L$ where each $\mathcal{U}_l$ is a closed subset of $\overline{\mathbb{R}}$ that contains $\{\pm\infty\}$. Let $\overline{F} : \mathcal{U} \to [0,1]$ be an*

[7] Although, note that the normalization of left- or right-continuity for each margin (defined ahead) is left unspecified here, in contrast to many treatments that define distribution functions as objects derived from random variables. This is innocuous, since the left- and right-continuous versions of a monotone real-valued function determine each other; see, for example, Section 2.2 of Schweizer and Sklar (1983).

[8] One occasionally encounters the phrase subdistribution as referring to a distribution-like function with largest value strictly smaller than 1. This is a distinct concept from the one introduced here, and no confusion between the two concepts will arise in this paper.

L-dimensional subdistribution function. Then there exists a proper L-dimensional joint distribution function F such that $F(u) = \overline{F}(u)$ for all $u \in \mathcal{U}$.

If $\overline{F}$ is continuous on $\mathcal{U}$, then there exists such an F that is continuous on $\overline{\mathbb{R}}^L$.[9] *If $\mathcal{U}$ is a finite set, then there exists such an F that has support contained in $\mathcal{U}_1^+ \times \cdots \times \mathcal{U}_L^+$, where $\mathcal{U}_l^+ \equiv (\mathcal{U}_l \setminus \{\pm\infty\}) \cup \{u_l^+\}$ for any point $u_l^+ \in \mathbb{R}$ such that $u_l^+ \geq \max(\mathcal{U}_l \setminus \{+\infty\})$.*[10,11]

The proof of Lemma 2 makes use of some fundamental results in copula theory, especially Sklar's theorem and a key intermediate result—referred to here as Sklar's lemma—that is used in the classical proof of Sklar's theorem.[12] Sklar's lemma establishes a result analogous to Lemma 2 for copulas and subcopulas, namely, that every subcopula can be extended to a copula. Since copulas and subcopulas are like distributions and subdistributions, but with fixed margins, Lemma 2 can be viewed as an extension of Sklar's lemma. The main contribution of this paper is to show how Lemma 2 (and its corollary in the next section) can be applied to provide tractable, sharp characterizations of identified sets in a wide variety of econometric models.[13]

3.2 *Extendibility and reducibility*

The following corollary to Lemma 2 provides some additional flexibility in choosing the extension F for a given subdistribution $\overline{F}$ for situations in which it is known that the margins of $\overline{F}$ satisfy certain properties. This is useful in partial identification analysis because it allows a researcher to maintain assumptions directly on the margins of a joint distribution. To state the corollary, I employ the following definition.

Definition 4. Suppose that $\mathcal{F}$ is a collection of proper one-dimensional distribution functions and $\overline{\mathcal{F}}$ is a collection of one-dimensional subdistribution functions with common domain $\mathcal{U}$. Then $\overline{\mathcal{F}}$ is *extendible* to $\mathcal{F}$ if, for every $\overline{F} \in \overline{\mathcal{F}}$, there exists an $F \in \mathcal{F}$ such that $F|_{\mathcal{U}} = \overline{F}$.[14]

The case of Lemma 2 with $L = 1$ established that the collection of all one-dimensional subdistribution functions on any common domain $\mathcal{U}$ is extendible to the entire collection of one-dimensional distribution functions. The definition of extendibility provides an additional layer of generality in allowing these collections to be restricted so

[9] Note that if u is an isolated point of $\mathcal{U}$, then $\overline{F}$ is regarded as being trivially continuous at u.

[10] The support of a distribution function F is defined as the minimal closed set $\mathcal{V}$ for which $\mathbb{P}[U \in \mathcal{V}] = 1$ if U is a random variable distributed according to F; see, for example, Shorack (2000, p. 110).

[11] I thank an anonymous referee for suggesting this additional result.

[12] See Appendix A for a brief review of the relevant concepts and results in copula theory.

[13] Chiburis (2010, p. 271) also noted briefly that Sklar's lemma could be used to simplify characterizations of identified sets in a specific type of nonparametric binary response model with two unobservable terms that have marginal distributions normalized to be uniform over [0, 1]. However, Chiburis (2010) did not develop this insight formally, and does not appear to have recognized that the argument could be generalized to apply to the much broader class of semiparametric models discussed in this paper. See also Mourifié (2015), who utilized Chiburis's (2010) insight to develop an analytic characterization of the sharp identified set for the model of Shaikh and Vytlacil (2011).

[14] For any function f with domain $\mathcal{A}$ and $\mathcal{B} \subseteq \mathcal{A}$, the notation $f|_{\mathcal{B}}$ denotes the restriction of f to $\mathcal{B}$.

as to exclude subdistribution and distribution functions that do not satisfy prespecified properties. Examples of such restrictions are given below.

The corollary to Lemma 2 can now be stated as follows.

COROLLARY 1. *Suppose that $\mathcal{U}$ and $\overline{F}$ are as in Lemma 2 and that, for each $l = 1, \ldots, L$, $\overline{F}_l \in \overline{\mathcal{F}}_l$, where $\overline{\mathcal{F}}_l$ is a set of one-dimensional subdistribution functions with common domain $\mathcal{U}_l$ that is extendible to a collection $\mathcal{F}_l$ of proper one-dimensional distribution functions. Then there exists a proper L-dimensional distribution function F defined on $\overline{\mathbb{R}}^L$ such that $F(u) = \overline{F}(u)$ for all $u \in \mathcal{U}$ and such that $F_l \in \mathcal{F}_l$ for each $l = 1, \ldots, L$.*

Before providing some concrete examples of extendibility, consider also the following companion definition of reducibility, which is in some sense the inverse of extendibility. This definition is also useful for the partial identification analysis in the next section. As the examples below show, it is typically easier to establish than extendibility.

DEFINITION 5. Suppose that $\mathcal{F}$ is a collection of proper one-dimensional distribution functions and $\overline{\mathcal{F}}$ is a collection of one-dimensional subdistribution functions with common domain $\mathcal{U}$. Then $\mathcal{F}$ is *reducible* to $\overline{\mathcal{F}}$ if $F|_{\mathcal{U}} \in \overline{\mathcal{F}}$ for every $F \in \mathcal{F}$.

From Definition 2, it follows immediately that the collection of all one-dimensional distribution functions is reducible to any collection of one-dimensional subdistribution functions defined on a common domain, $\mathcal{U}$. Like the definition of extendibility, the definition of reducibility allows one to also consider an analogous relationship for restricted collections of distribution and subdistribution functions.

The following are some concrete examples of extendibility and reducibility. The first example is trivial, but useful in contexts where a researcher wishes to maintain a known marginal distribution for an unobservable. This is frequently done for identification in classical applications of parametric models, and is also often imposed as a normalization in nonparametric models; see, for example, Matzkin (2003, 2007), Chernozhukov and Hansen (2005), Chesher (2010), or Torgovitsky (2015, 2017).

EXAMPLE 1. Suppose that $\mathcal{F} = \{\delta\}$ is the set consisting of a single proper, one-dimensional distribution function δ, and suppose that $\overline{\mathcal{F}} = \{\delta|_{\mathcal{U}}\}$ for any $\mathcal{U} \subseteq \overline{\mathbb{R}}$. Then $\overline{\mathcal{F}}$ is extendible to $\mathcal{F}$, and $\mathcal{F}$ is reducible to $\overline{\mathcal{F}}$.

The next example says that a collection of subdistributions with a common compact domain (leaving aside $\{\pm\infty\}$) is extendible to the collection of distributions with mean 0, and that the converse reducibility property also holds.

EXAMPLE 2. Suppose that $\mathcal{F}$ is the set of all proper one-dimensional distribution functions such that $\int u\, dF(u) = 0$. Suppose that $\overline{\mathcal{U}} \subseteq \mathbb{R}$ is compact and that $\overline{\mathcal{F}}$ is any collection of subdistributions $\overline{F}$ with domains $\mathcal{U} \equiv \overline{\mathcal{U}} \cup \{\pm\infty\}$ for which $\overline{F}(u) \in (0, 1)$ for every $u \in \overline{\mathcal{U}}$. Then $\overline{\mathcal{F}}$ is extendible to $\mathcal{F}$, and $\mathcal{F}$ is reducible to $\overline{\mathcal{F}}$.

To see this, observe that if $F \in \mathcal{F}$, then its restriction to $\mathcal{U}$ is a subdistribution regardless of whether F has mean zero. Conversely, if $\overline{F} \in \overline{\mathcal{F}}$, then because $\overline{\mathcal{U}}$ is compact

and $1 > \overline{F}(u) > 0$ for every $u \in \overline{\mathcal{U}}$, one can construct an F that agrees with $\overline{F}$ on $\mathcal{U}$ and places sufficient mass sufficiently far out in either its left or right tail to ensure that $\int u\, dF(u) = 0$.

Manski (1988) observed that the type of conditional mean zero conditions used in linear models have no identifying content for the index coefficients in the semiparametric binary response model of Section 2. Manski's intuition is essentially what is contained in Example 2. In the absence of any additional identifying assumptions, the subdistribution of interest for a binary response model always has a bounded domain (leaving aside $\{\pm\infty\}$). As a result, restricting the set of distributions for the unobservables to have mean 0 will not place any restrictions on the set of underlying subdistributions that determines the identified set.

In contrast, if $\mathcal{F}$ satisfies a median (or other quantile) restriction, then it is only reducible to collections of subdistributions that satisfy the analogous condition.

EXAMPLE 3. Suppose that $\mathcal{F}$ is the set of all proper one-dimensional distribution functions F such that $F(0) = \frac{1}{2}$. Suppose that $\overline{\mathcal{F}}$ is a collection of subdistributions $\overline{F}$ with domain $\mathcal{U}$ such that $0 \in \mathcal{U}$ and such that $\overline{F}(0) = \frac{1}{2}$. Then $\overline{\mathcal{F}}$ is extendible to $\mathcal{F}$, and $\mathcal{F}$ is reducible to $\overline{\mathcal{F}}$.[15]

The binary response model in Section 2 featured a median zero restriction as in Example 3. The analysis there employed Corollary 1 in the special case in which $L = 1$. Intuitively, the idea is simply that any one-dimensional subdistribution function with 0 in its domain that evaluates to $\frac{1}{2}$ at 0 can be extended to a one-dimensional distribution function that does the same. Conversely, a one-dimensional distribution function that evaluates to $\frac{1}{2}$ at 0 will, when restricted to a subset that contains 0, generate a one-dimensional subdistribution function that does the same.

The next example concerns symmetry restrictions.

EXAMPLE 4. Suppose that $\mathcal{F}$ is the set of all proper one-dimensional distribution functions that are symmetric around 0, that is, such that $F(u) = 1 - F(-u)$ for all $u \in \mathbb{R}$. Suppose that $\overline{\mathcal{F}}$ is a collection of subdistributions with domain $\mathcal{U}$ such that $0 \in \mathcal{U}$, $u \in \mathcal{U}$ if and only if $-u \in \mathcal{U}$, and such that $\overline{F}(u) = 1 - \overline{F}(-u)$ for all $\overline{F} \in \overline{\mathcal{F}}$ and $u \in \mathcal{U}$. Then $\overline{\mathcal{F}}$ is extendible to $\mathcal{F}$, and $\mathcal{F}$ is reducible to $\overline{\mathcal{F}}$.[16]

There are two requirements for the subdistribution class in Example 4. First, the domain on which its members are defined, $\mathcal{U}$, must be symmetric around 0 in the sense that $u \in \mathcal{U}$ if and only if $-u \in \mathcal{U}$. Second, functions in the subdistribution class must themselves be symmetric on this domain. The conclusion of the example is that any such one-dimensional subdistribution function can be extended to the entire real line

[15] More generally, a similar statement could be made for the concept of "$\mathcal{T}$ independence" introduced by Masten and Poirier (2016).

[16] A proof of this statement can be found in Appendix B.

in such a way that preserves monotonicity and symmetry around 0. The converse reducibility statement, that is, that any symmetric distribution function is still symmetric when restricted to a symmetric domain $\mathcal{U}$, is immediate.

The last two examples concern support restrictions.[17] The first example is about classes of distribution and subdistribution functions that have support contained within some prespecified interval.

EXAMPLE 5. Suppose that $\mathcal{F}$ is the set of all proper one-dimensional distribution functions that have support contained within $[\underline{u}, \overline{u}]$ for some $\underline{u}, \overline{u} \in \overline{\mathbb{R}}$. Suppose that $\overline{\mathcal{F}}$ is any collection of subdistributions with domain $\mathcal{U} \subseteq [\underline{u}, \overline{u}]$ such that $\underline{u} \in \mathcal{U}$, $\overline{u} \in \mathcal{U}$, and such that $\overline{F}(\underline{u}) = 0$ and $\overline{F}(\overline{u}) = 1$. Then $\overline{\mathcal{F}}$ is extendible to $\mathcal{F}$, and $\mathcal{F}$ is reducible to $\overline{\mathcal{F}}$.

Example 5 simply says that if a class of distribution functions is known to concentrate its mass on some interval, then a class of subdistribution functions that concentrates its mass strictly within the same interval is extendible to this class of distribution functions. The converse reducibility statement follows because the restriction of any such distribution function to a set contained within this interval must concentrate its mass on this interval as well.

The final example concerns distribution functions with support $\mathbb{R}$, that is, "full support." As in Example 2, these are reducible to any collection of subdistributions with compact domains (leaving aside $\{\pm\infty\}$), and any such collection of subdistributions is extendible to the set of distribution functions with full support.

EXAMPLE 6. Suppose that $\mathcal{F}$ is the set of all proper one-dimensional distribution functions that have support equal to $\mathbb{R}$. Suppose that $\overline{\mathcal{U}} \subseteq \mathbb{R}$ is compact and that $\overline{\mathcal{F}}$ is any collection of subdistributions $\overline{F}$ with domains $\mathcal{U} \equiv \overline{\mathcal{U}} \cup \{\pm\infty\}$ for which $\overline{F}(u) \in (0, 1)$ for every $u \in \overline{\mathcal{U}}$. Then $\overline{\mathcal{F}}$ is extendible to $\mathcal{F}$, and $\mathcal{F}$ is reducible to $\overline{\mathcal{F}}$. The reducibility follows immediately, while the extendibility can be established by using the construction in the proof of Lemma 2.

Similarly to Example 2, an implication of Example 6 in the context of the binary response model of Section 2 is that requiring the distribution of the unobservable U to have full support is not restrictive. The subdistribution of interest in these models already has bounded domain (apart from $\{\pm\infty\}$), so the content of Example 6 is that requiring these subdistribution functions to be extendible to a class of distribution functions with full support would not entail any additional restrictions.

3.3 *PIES in a general econometric model*

In this section, I use the subdistribution extension lemma and its corollary to develop sharp characterizations of identified sets for a general model of a random vector Y with support $\mathcal{Y}$, conditional on a random vector X with support $\mathcal{X}$. The distinction between Y and X is that the determination of Y is modeled, while the determination of X is

[17] These examples were suggested by an anonymous referee.

not. Both Y and X should be viewed as observable random variables. The following discussion of identification is premised on knowledge of the distribution of (Y, X).

The parameters of the model are $S = (\theta, F)$, where θ is a finite-dimensional parameter, and $F : \overline{\mathbb{R}}^L \times \mathcal{X} \to [0, 1]$ is an unknown conditional distribution function for an L-dimensional latent vector, U.[18] I assume that every $S = (\theta, F)$ generates a single distribution of Y given $X = x$ for every supported x, denoted as

$$\mathbb{P}_S[Y \leq y | X = x] \equiv \omega_{y|x}(\theta, F) \tag{11}$$

for some known function $\omega_{y|x}$ of θ and F. This assumption requires the model to be complete, so rules out, for example, models that have multiple equilibria but an unspecified equilibrium selection rule, such as the entry games studied by Tamer (2003) and Ciliberto and Tamer (2009). For the binary response model in Section 2, $L = 1$, $\theta = \beta$, and $\omega_{y|x}$ corresponds to (5) when $y = 0$, and is identically 1 when $y = 1$.

The researcher only considers θ that lie in some admissible set, $\Theta^\dagger$. For the model in Section 2, $\Theta^\dagger$ was given by $\mathcal{B}^\dagger \equiv \{b \in \mathbb{R}^2 : \|b\| = 1\}$. The dimension of the latent variables, L, is assumed to be known by the researcher, but is otherwise unrestricted in relationship to the dimensions of Y and X. The space of all proper L-dimensional conditional distribution functions is denoted by $\mathcal{F}$. The researcher only considers F lying in some *admissible* set $\mathcal{F}^\dagger \subseteq \mathcal{F}$, which is typically a strict subset of $\mathcal{F}$. The admissible set contains only those F which satisfy the a priori (or "identifying") assumptions maintained by the researcher.

Assumption A describes the types of identifying assumptions that the researcher can maintain on F. This puts some abstract structure on $\mathcal{F}^\dagger$ and so delimits the generality of the models to which the subsequent analysis applies.

ASSUMPTION A. *The admissible set $\mathcal{F}^\dagger$ is the set of all $F \in \mathcal{F}$ that satisfy all of the following properties. In the following, for any $F \in \mathcal{F}$, the function $F_l(\cdot|x)$ denotes the lth margin of $F(\cdot|x)$ (Definition 3).*

A1. *$F(u|x) = F(u|x')$ for all $u \in \overline{\mathbb{R}}^L$, all $x, x' \in \mathcal{X}^\dagger_{0,m_0}$, and all $m_0 = 1, \ldots, M_0$, where $\mathcal{X}^\dagger_{0,m_0}$ are known (possibly empty) disjoint subsets of $\mathcal{X}$.*

A2. *For each $l = 1, \ldots, L$, $F_l(u_l|x) = F_l(u_l|x')$ for all $u_l \in \overline{\mathbb{R}}$, all $x, x' \in \mathcal{X}^\dagger_{l,m_l}$, and all $m_l = 1, \ldots, M_l$, where $\mathcal{X}^\dagger_{l,m_l}$ are known (possibly empty) disjoint subsets of $\mathcal{X}$.*

A3. *For each $l = 1, \ldots, L$ and $x \in \mathcal{X}$, $F_l(\cdot|x) \in \mathcal{F}^\dagger_{l,x}$, where $\mathcal{F}^\dagger_{l,x}$ is a known collection of proper one-dimensional distribution functions.*

A4. *$\rho(\theta, F) \geq 0$ for some known vector-valued function ρ, where the inequality is interpreted component-wise.*

Assumption A1 is an independence restriction with respect to some known (possible empty) subsets $\mathcal{X}^\dagger_{0,m_0}$ of the support of X. The leading case is $M_0 = 1$ with $\mathcal{X}^\dagger_{0,1} = \mathcal{X}$, under which Assumption A1 requires independence between U and X. Specifying multiple

[18] All of the results still hold if θ is an infinite-dimensional parameter. However, as will become clear, implementing Theorem 1 would be computationally prohibitive in this case.

such sets ($M_0 > 1$) can be used to require independence between U and some component of X, conditional on some other components of X. Assumption A2 is the same as Assumption A1, but for a single component of U rather than the entire vector. Restrictions of type A3 require the marginal distributions of U to lie in known classes that are extendible and reducible in a sense described ahead in Theorem 1. Assumption A4 allows for miscellaneous restrictions, represented here by a function ρ chosen by the researcher. Like Assumption A3, Assumption A4 will interact with additional conditions stated ahead in Theorem 1.

In the formal analysis ahead, all of the restrictions on $\mathcal{F}^\dagger$ in Assumption A are maintained simultaneously. However, notice that any of these restrictions can also be made trivial (nonrestrictive) by using specific choices of $\mathcal{X}^\dagger_{0,m_0}$, $\mathcal{X}^\dagger_{l,m_l}$, $\mathcal{F}^\dagger_{l,x}$, and/or ρ.[19] Also, for some assumptions, there may be more than one category A1–A4 under which it could be classified. For example, the conditional median 0 restriction for the binary response model in Section 2 would be most naturally written as Assumption A3 using Example 3, but it could also be written as Assumption A4 by taking $\rho(\theta, F) = [F(0|0) - \frac{1}{2}, \frac{1}{2} - F(0|0), F(0|1) - \frac{1}{2}, \frac{1}{2} - F(0|1)]'$.

An example of an assumption that one might wish to consider, but which *cannot* be characterized as one of A1–A4, is positive quadrant dependence (Lehmann (1966)) between the components of U in a case where $L > 1$, say $L = 2$. Positive quadrant dependence (conditional on $X = x$) is satisfied if and only if $F(u_1, u_2|x) \geq F_1(u_1|x)F_2(u_2|x)$ for all $(u_1, u_2) \in \overline{\mathbb{R}}^2$. This condition cannot be expressed as a finitely-valued function ρ, nor can it characterized as placing a restriction solely on the margins of F. Hence, while Assumptions A1–A4 are general, they are still restrictive in the sense of constraining the types of assumptions that a researcher can consider maintaining.

The identified set for S, denoted as $\mathcal{S}^\star$, is the collection of all admissible θ and F that generate the observed distributions of Y, conditional on X. Formally,

$$\mathcal{S}^\star \equiv \big\{S = (\theta, F) : \theta \in \Theta^\dagger, F \in \mathcal{F}^\dagger \text{ and } \omega_{y|x}(\theta, F) = \mathbb{P}[Y \leq y|X = x] \text{ for all } y \in \mathcal{Y}, x \in \mathcal{X}\big\}. \tag{12}$$

The researcher's object of interest is a function π of S. The identified set associated with a given parameter π is denoted by $\mathcal{P}^\star = \{\pi(\theta, F) : (\theta, F) \in \mathcal{S}^\star\}$. In Section 2, the parameter of interest was the ATE, defined in (3), with $\theta = \beta$.

Theorem 1 ahead uses the subdistribution extension lemma to characterize $\mathcal{P}^\star$ in this general model. The proof of Theorem 1, which is given in Appendix C, is an extensive generalization of the argument in Section 2. The statement of the theorem is premised on the existence of a collection of subsets $\{\mathcal{U}_x(\theta) : x \in \overline{\mathcal{X}}\}$—which could depend on θ—that satisfies the following conditions.

CONDITION U. *Suppose that $\mathcal{F}^\dagger$ can be represented as in Assumption A. A collection of subsets $\{\mathcal{U}_x(\theta) : x \in \overline{\mathcal{X}}\}$ satisfies Condition U if the following five statements are true.*

[19] In particular, Assumptions A1 or A2 become vacuous if $M_0 = 1$ with $\mathcal{X}^\dagger_{0,1} = \emptyset$ or $M_l = 1$ with $\mathcal{X}^\dagger_{l,m_l} = \emptyset$. Assumption A3 is trivially satisfied by taking $\mathcal{F}^\dagger_{l,x}$ to be the set of all proper, one-dimensional distribution functions, for every l and x. Assumption A4 can be made tautological by taking $\rho(\theta, F) \equiv 0$.

U1. $\mathcal{U}_x(\theta) = \mathcal{U}_{1,x}(\theta) \times \cdots \times \mathcal{U}_{L,x}(\theta)$, *where* $\mathcal{U}_{l,x}(\theta) \subseteq \overline{\mathbb{R}}$ *is closed and such that* $\{\pm\infty\} \subseteq \mathcal{U}_{l,x}(\theta)$ *for each* $l = 1, \ldots, L$ *and every* $x \in \overline{\mathcal{X}}$.

U2. *There exist functions* $\{\overline{\omega}_{y|x} : y \in \overline{\mathcal{Y}}, x \in \overline{\mathcal{X}}\}$, $\overline{\pi}$, *and* $\overline{\rho}$ *such that for every* $F \in \mathcal{F}$:

$$\omega_{y|x}(\theta, F) = \overline{\omega}_{y|x}\big(\theta, F(\cdot|x)|_{\mathcal{U}_x(\theta)}\big), \tag{U2.ω}$$

$$\pi(\theta, F) = \overline{\pi}\big(\theta, \big\{F(\cdot|x)|_{\mathcal{U}_x(\theta)} : x \in \overline{\mathcal{X}}\big\}\big), \tag{U2.π}$$

$$\rho(\theta, F) = \overline{\rho}\big(\theta, \big\{F(\cdot|x)|_{\mathcal{U}_x(\theta)} : x \in \overline{\mathcal{X}}\big\}\big). \tag{U2.ρ}$$

U3. *For each* $l = 1, \ldots, L$ *and every* $x \in \overline{\mathcal{X}}$, *there exists a collection of subdistributions* $\overline{\mathcal{F}}^{\dagger}_{l,x}$ *with common domain* $\mathcal{U}_{l,x}(\theta)$ *such that* $\mathcal{F}^{\dagger}_{l,x}$ *is reducible to* $\overline{\mathcal{F}}^{\dagger}_{l,x}$ *and* $\overline{\mathcal{F}}^{\dagger}_{l,x}$ *is extendible to* $\mathcal{F}^{\dagger}_{l,x}$.

U4. $\mathcal{U}_x(\theta) = \mathcal{U}_{x'}(\theta)$ *for all* $x, x' \in (\overline{\mathcal{X}} \cap \mathcal{X}^{\dagger}_{0,m_0})$ *and all* $m_0 = 1, \ldots, M_0$.

U5. $\mathcal{U}_{l,x}(\theta) = \mathcal{U}_{l,x'}(\theta)$ *for all* $x, x' \in (\overline{\mathcal{X}} \cap \mathcal{X}^{\dagger}_{l,m_l})$ *and each* $l = 1, \ldots, L$ *and* $m_l = 1, \ldots, M_l$.

Although Theorem 1 will be valid for any collection of subsets that satisfies Condition U, in practice one wants to find the *smallest* such collection. I elaborate on this point after the statement of the result.

THEOREM 1 (PIES). *Let* $\overline{\mathcal{Y}} \subseteq \mathcal{Y}$ *and* $\overline{\mathcal{X}} \subseteq \mathcal{X}$. *Suppose that* $\mathcal{F}^{\dagger}$ *can be represented as in Assumption A. For any* $\theta \in \Theta^{\dagger}$, *let* $\{\mathcal{U}_x(\theta) : x \in \overline{\mathcal{X}}\}$ *be any collection of subsets of* $\overline{\mathbb{R}}^L$ *that satisfy Condition U. If* $p \in \mathcal{P}^{\star}$, *then there exist a* $\theta \in \Theta^{\dagger}$ *and function* $\overline{F}(\cdot|x) : \mathcal{U}_x(\theta) \to [0, 1]$ *for each* $x \in \overline{\mathcal{X}}$ *such that*:

$$\overline{\omega}_{y|x}\big(\theta, \overline{F}(\cdot|x)\big) = \mathbb{P}[Y \leq y | X = x] \quad \text{for all } y \in \overline{\mathcal{Y}} \text{ and } x \in \overline{\mathcal{X}}, \tag{T1.1}$$

$$\overline{F}(\cdot|x) \text{ is an } L\text{-dimensional subdistribution for each } x \in \overline{\mathcal{X}}, \tag{T1.2}$$

$$\overline{F}(u|x) = \overline{F}\big(u|x'\big) \quad \text{for all } x, x' \in \big(\overline{\mathcal{X}} \cap \mathcal{X}^{\dagger}_{0,m_0}\big), \text{ all } u \in \mathcal{U}_x(\theta), m_0 = 1, \ldots, M_0, \tag{T1.3}$$

$$\overline{F}_l(u_l|x) = \overline{F}_l\big(u_l|x'\big) \quad \text{for all } x, x' \in \big(\overline{\mathcal{X}} \cap \mathcal{X}^{\dagger}_{l,m_l}\big), \text{ all } u_l \in \mathcal{U}_{l,x}(\theta), m_l = 1, \ldots, M_l, \text{ and all } l = 1, \ldots, L, \text{ where } \overline{F}_l(\cdot|x) \text{ is the } l\text{th margin of } \overline{F}(\cdot|x), \tag{T1.4}$$

$$\overline{F}_l(\cdot|x) \in \overline{\mathcal{F}}^{\dagger}_{l,x} \quad \text{for each } l = 1, \ldots, L \text{ and } x \in \overline{\mathcal{X}}, \tag{T1.5}$$

$$\overline{\rho}\big(\theta, \big\{\overline{F}(\cdot|x) : x \in \overline{\mathcal{X}}\big\}\big) \geq 0, \tag{T1.6}$$

$$\overline{\pi}\big(\theta, \big\{\overline{F}(\cdot|x) : x \in \overline{\mathcal{X}}\big\}\big) = p. \tag{T1.7}$$

If $\overline{\mathcal{X}} = \mathcal{X}$ *and* $\overline{\mathcal{Y}} = \mathcal{Y}$, *then the existence of a* $\theta \in \Theta^{\dagger}$ *and functions* $\{\overline{F}(\cdot|x) : \mathcal{U}_x(\theta) \to [0, 1]\}_{x \in \mathcal{X}}$ *satisfying* (T1.1)–(T1.7) *also implies that* $p \in \mathcal{P}^{\star}$.

Theorem 1 provides a system of equations (T1.1)–(T1.7) that determines whether a given value p is in the sharp identified set $\mathcal{P}^\star$. For a fixed $\theta \in \Theta^\dagger$, the system of equations is in terms of functions $\{\overline{F}(\cdot|x) : x \in \overline{\mathcal{X}}\}$ with domains $\{\mathcal{U}_x(\theta) : x \in \overline{\mathcal{X}}\}$. These domain sets are constructed by the researcher, subject to U1–U5 in Condition U. This condition describes the set of points on which a collection of functions $\{\overline{F}(\cdot|x) : x \in \overline{\mathcal{X}}\}$ would need to be evaluated to determine (via Corollary 1 and a generalization of the argument in Section 2) whether there exists an $F \in \mathcal{F}^\dagger$ such that $(\theta, F) \in \mathcal{S}^\star$ and $\pi(\theta, F) = p$. The binary response model in Section 2, with the parameter $\theta = \beta$, had $\mathcal{U}_x(\beta) = \{\pm\infty, 0, \beta_0, \beta_0 + \beta_1\}$. The β_0 and $\beta_0 + \beta_1$ points were needed to satisfy both (U2.ω) and (U2.π), while 0 was included to satisfy either (U2.ρ) or U3, depending on which way one views the median independence restriction as entering Assumption A.

Condition U1 says that each $\mathcal{U}_x(\theta)$ must be rectangular, with each one-dimensional slice containing $\{\pm\infty\}$, so that they satisfy the conditions to be a domain of a subdistribution described in Definition 2. Condition U2 requires $\{\mathcal{U}_x(\theta) : x \in \overline{\mathcal{X}}\}$ to also be rich enough to evaluate the functionals corresponding to observational equivalence ($\omega_{y|x}(\theta, \cdot)$), the parameter of interest ($\pi(\theta, \cdot)$), and the a priori assumptions encoded using A4 of Assumption A ($\rho(\theta, \cdot)$). Similarly, Condition U3 requires each $\mathcal{U}_x(\theta)$ to be large enough to support classes of subdistribution functions that are extendible to (and reducible from) the classes of distribution functions specified in A3 of Assumption A. Conditions U4 and U5 require all or part of $\mathcal{U}_x(\theta)$ and $\mathcal{U}_{x'}(\theta)$ to be the same for values x and x' over which A1 and/or A2 are imposed. This is simply to ensure that it is possible to consider equating $\overline{F}(\cdot|x)$ and $\overline{F}(\cdot|x')$ as functions.

Conditions U1–U5 are trivially satisfied by taking $\mathcal{U}_x(\theta) = \overline{\mathbb{R}}^L$ for every $x \in \overline{\mathcal{X}}$ and any $\theta \in \Theta^\dagger$. This choice renders Theorem 1 useless, since determining whether there exist functions $\{\overline{F}(\cdot|x) : x \in \overline{\mathcal{X}}\}$ that satisfy (T1.1)–(T1.7) remains as difficult as determining whether there exists an $F \in \mathcal{F}^\dagger$ such that $(\theta, F) \in \mathcal{S}^\dagger$. The key to making Theorem 1 useful is to choose sets $\mathcal{U}_x(\theta)$ that are finite for every $x \in \overline{\mathcal{X}}$ and $\theta \in \Theta^\dagger$. In this case, Theorem 1 reduces the infinite-dimensional existence problem for F to a potentially finite-dimensional existence problem for $\{\overline{F}(\cdot|x) : x \in \overline{\mathcal{X}}\}$. The primary limitation of Theorem 1 comes from cases in which either the model or the assumptions require a choice of $\{\mathcal{U}_x(\theta) : x \in \overline{\mathcal{X}}\}$ that is infinite.

In this regard, Condition U1 is not by itself restrictive, since the rectangular form it dictates can always be constructed by including more points. The resulting rectangular set will still be finite if the initial non-rectangular set was finite. For the same reason, U4 and U5 are also not restrictive. The restrictiveness of Theorem 1 comes from being able to satisfy U2 and U3 while keeping $\mathcal{U}_x(\theta)$ finite. In particular, for $\mathcal{U}_x(\theta)$ to satisfy U2 and still have finite cardinality requires each of $\omega_{y|x}$ (for any $y \in \overline{\mathcal{Y}}$, $x \in \overline{\mathcal{X}}$), ρ, and π to depend on (θ, F) only through the value of F on a finite subset of its domain. This requires these functions to not depend on moments of U, which is a limitation in some models.[20]

[20]It is possible that one could circumvent this limitation through an approximation argument, for example by using a finite grid to approximate moments.

3.4 *Implementation and dimension reduction*

Implementing the general form of PIES in Theorem 1 requires two steps.[21]

The first step is to specify the system of equations in Theorem 1 and to find a collection of domain sets $\{\mathcal{U}_x(\theta) : x \in \overline{\mathcal{X}}\}$ that is sufficiently large to evaluate this system of equations. This requires deriving the distribution of Y, conditional on X, that would be generated by a given $S = (\theta, F)$, as in (5). This derivation yields the mapping $\omega_{y|x}$, which in turn determines a minimal set of points that must be included in the set $\mathcal{U}_x(\theta)$ for (U2.ω) to be satisfied. Examples of this derivation are given in Section 4, and Sections S.1–S.2 of the Supplemental Material (Torgovitsky (2019)). Next, one chooses a parameter of interest, applies the same type of derivation to obtain the mapping π, and enlarges $\mathcal{U}_x(\theta)$ sufficiently to ensure that (U2.π) is satisfied. Then, one casts their desired identifying assumptions in terms of Assumption A, and enlarges $\mathcal{U}_x(\theta)$ further so that (U2.ρ), U3, U4, and U5 are satisfied. At the end, the collected points are constructed into a final set $\mathcal{U}_x(\theta)$ that satisfies the rectangular property in Condition U1, adding $\{\pm\infty\}$ to the margins, if necessary. This concludes the first step of implementation.

To conduct this first step, the analyst needs to choose the sets $\overline{\mathcal{Y}}$ and $\overline{\mathcal{X}}$. If either of these sets is infinite, then, in general, the system of equations (T1.1)–(T1.7) cannot be solved exactly. An implication is that if Y and/or X are continuously distributed, then in practice one must take $\overline{\mathcal{Y}}$ and/or $\overline{\mathcal{X}}$ to be strict subsets of $\mathcal{Y}$ and/or $\mathcal{X}$. Applying Theorem 1 when this is the case leads to an outer set for $\mathcal{P}^\star$. With sufficient computing power, this outer identified set can be made arbitrarily close to the (sharp) identified set by taking $\overline{\mathcal{Y}}$ and $\overline{\mathcal{X}}$ to be arbitrarily large subsets of $\mathcal{Y}$ and $\mathcal{X}$. I demonstrate this point in Section 4 using simulated data in a case where X is a continuously distributed scalar random variable. The approach that I use there is to take $\overline{\mathcal{X}}$ to be a set of equally spaced quantiles. Intuitively, this helps ensure that different $x \in \overline{\mathcal{X}}$ provide non-redundant identifying information. When using actual data, these issues become less important, since the empirical distribution is always discrete, regardless of whether Y and/or X are modeled as continuous random variables.

The second step in implementing PIES is to solve the system of equations in Theorem 1 repeatedly over a grid of possible θ and p. An important consideration here is whether the system of equations (T1.1)–(T1.7) can be reliably solved for a given θ and p. An attractive feature of PIES is that the crucial subdistribution condition, (T1.2), always places a linear restriction on $\overline{F}$ given a fixed θ—recall Definition 2. The conditional independence restrictions (T1.3) and (T1.4) are also always linear for a fixed θ. As a result, the entire system in Theorem 1 will be linear for a fixed θ if each of $\overline{\omega}_{y|x}(\theta, \overline{F})$, $\overline{\rho}(\theta, \overline{F})$, and $\overline{\pi}(\theta, \overline{F})$ is linear in the values of $\overline{F}$. The same must also be true of whatever lower-level condition defines the classes of subdistributions in Condition U3.

Theorem 1 is still valid for cases in which this linearity does not hold, but it may be more difficult to implement due to the difficulties involved in reliably solving nonlinear systems of equations. Given this requirement, PIES is computationally more attractive in situations for which the system is linear in $\overline{F}$. Whether this is so depends on the model,

[21]Code for implementing the simulations discussed in the next section is available from the GitHub repository `https://github.com/a-torgovitsky/pies`.

assumptions, and parameters of interest, and therefore must be determined on a case-by-case basis. However, as I demonstrate in the next section and in the Supplemental Material, the requirement of linearity in $\overline{F}$ is less restrictive than it may initially appear.

A related consideration is whether $\mathcal{P}^\star$ can be characterized more efficiently by solving optimization problems, as in (8). As stated, Theorem 1 would suggest that one needs to determine the existence or non-existence of a solution to (T1.1) for every $\theta \in \Theta^\dagger$ and every p in the range of $\pi(\theta, \cdot)$. However, if π is scalar-valued and if $\overline{\omega}_{y|x}(\theta, \overline{F})$, $\overline{\rho}(\theta, \overline{F})$, and $\overline{\pi}(\theta, \overline{F})$ are all linear functions of $\overline{F}$ for a fixed $\theta \in \Theta^\dagger$, then one can determine $\mathcal{P}^\star$ by minimizing and maximizing $\overline{\pi}(\theta, \overline{F})$ over $\overline{F}$ subject to (T1.1)–(T1.6) once for each $\theta \in \Theta^\dagger$. The next corollary provides a formal and more general statement of this assertion.

PROPOSITION 1. *For any $\theta \in \Theta^\dagger$, let $\{\mathcal{U}_x(\theta) : x \in \overline{\mathcal{X}}\}$ satisfy Condition U, and define $\overline{\mathcal{F}}(\theta) \equiv \{\overline{F}(\cdot|x) : \mathcal{U}_x(\theta) \to [0,1], x \in \overline{\mathcal{X}}\}$. Suppose that $\overline{\mathcal{F}}(\theta)$ is finite-dimensional and that $\overline{\pi}(\theta, \cdot)$ is a continuous function on $\overline{\mathcal{F}}(\theta)$. Let $\overline{\mathcal{F}}^\star(\theta)$ denote the subset of $\overline{F} \in \overline{\mathcal{F}}(\theta)$ that satisfy (T1.1)–(T1.6), and suppose that $\overline{\mathcal{F}}^\star(\theta)$ is connected for every $\theta \in \Theta^\dagger$. Then $\mathcal{P}^\star \subseteq \bigcup_{\theta \in \Theta^\dagger} [\underline{p}^\star(\theta), \overline{p}^\star(\theta)]$, where*

$$\underline{p}^\star(\theta) \equiv \inf_{\overline{F} \in \overline{\mathcal{F}}(\theta)} \overline{\pi}(\theta, \overline{F}) \quad \textit{subject to } (T1.1)\text{–}(T1.6) \tag{13}$$

and $\overline{p}^\star(\theta)$ is the corresponding supremum.[22] *If $\overline{\mathcal{X}} = \mathcal{X}$ and $\overline{\mathcal{Y}} = \mathcal{Y}$, then $\mathcal{P}^\star = \bigcup_{\theta \in \Theta^\dagger} [\underline{p}^\star(\theta), \overline{p}^\star(\theta)]$.*

The procedure in Proposition 1 still requires one to create a grid on $\Theta^\dagger$ and check two optimization problems for each point $\theta \in \Theta^\dagger$. The difficulty of doing so increases in the size of the grid, which increases exponentially with the dimension of θ. This is a computational bottleneck, albeit one that is also faced by other general procedures for constructing sharp identified sets. However, a potentially useful feature of the PIES approach is that its structure suggests a way to ameliorate this problem. In particular, the next proposition records that in some cases, two choices of θ that yield the same "labeling" of $\mathcal{U}_x(\theta)$ also lead to the same optimization problems in Proposition 1. It follows that $\mathcal{P}^\star$ can be determined by partitioning $\Theta^\dagger$ into equivalence classes that yield the same labels, and then solving these problems once for a representative θ in each equivalence class.

PROPOSITION 2. *Suppose that the conditions and definitions of Theorem 1 and Proposition 1 are satisfied for $\theta, \theta' \in \Theta^\dagger$. If $\{\overline{\pi}(\theta, \overline{F}) : \overline{F} \in \overline{\mathcal{F}}^\star(\theta)\} = \{\overline{\pi}(\theta', \overline{F}) : \overline{F} \in \overline{\mathcal{F}}^\star(\theta')\}$, then $[\underline{p}^\star(\theta), \overline{p}^\star(\theta)] = [\underline{p}^\star(\theta'), \overline{p}^\star(\theta')]$.*

Typically, for the hypothesis of Proposition 2 to hold for distinct values θ, θ', one requires $\overline{\pi}(\theta, \overline{F})$ to not depend directly on θ. The dependence of $\overline{\pi}(\theta, \overline{F})$ on θ then arises solely through the set $\overline{\mathcal{F}}^\star(\theta)$ that $\overline{F}$ varies over. For example, this was the case for the average treatment effect in (3). In that model, the same property was also true for $\overline{\omega}_{y|x}$.

[22] If the constraint sets are empty, then follow the usual convention of taking $\underline{p}^\star(\theta) = +\infty$ and $\overline{p}^\star(\theta) = -\infty$, so that $[\underline{p}^\star(\theta), \overline{p}^\star(\theta)] = \emptyset$.

Thus, the only channel through which θ could affect the set $\{\overline{\pi}(\theta, \overline{F}) : \overline{F} \in \overline{\mathcal{F}}^\star(\theta)\}$ was through its impact on the set $\mathcal{U}_x(\theta)$, which in that model was equal to $\{0, \beta_0, \beta_0 + \beta_1 x\}$. By itself, changes in this set caused by changes in θ have no direct impact on the functions $\overline{\mathcal{F}}(\theta)$, since they only serve to change the labeling of points in the domains $\{\mathcal{U}_x(\theta) : x \in \overline{\mathcal{X}}\}$. This change in labeling could cause the elements of $\overline{\mathcal{F}}^\star(\theta)$ to change, for example if it changes the ordering of points in the domains.[23] When it does not, θ and θ' will generate the same system of equations in Theorem 1 and the same optimization problem in Proposition 1. In this case, (13) and the analogous maximization problem would only need to be solved for either θ or θ', but not both points.

4. Bivariate binary response models

In this section, I apply Theorem 1 to the bivariate binary response model

$$Y_1 = \mathbb{1}\big[U_1 > g_1(Y_2, X)\big] \tag{14}$$

$$\text{and} \quad Y_2 = \mathbb{1}\big[U_2 > g_2(X)\big]. \tag{15}$$

In (14)–(15), $Y = (Y_1, Y_2)$ are observed binary random variables, $g \equiv (g_1, g_2)$ is a pair of unknown functions, X is a random vector with support $\mathcal{X}$, and $U \equiv (U_1, U_2)$ is a bivariate latent variable. Let $\mathcal{F}$ denote the set of all proper bivariate conditional distribution functions $F : \overline{\mathbb{R}}^2 \times \mathcal{X} \to [0, 1]$, and let $\mathcal{F}^\dagger$ denote the admissible subset of $\mathcal{F}$. In the notation of the previous section, the parameter θ is the pair of functions $g = (g_1, g_2)$, with admissible set $\mathcal{G}^\dagger$. The function $\omega_{y|x}$ defined in (11) is given by

$$\begin{aligned}\omega_{(y_1,y_2)|x}(g, F) &\equiv \mathbb{P}_S[Y_1 \le y_1, Y_2 \le y_2 | X = x] \\ &= \begin{cases} F\big(g_1(0, x), g_2(x)|x\big), & \text{if } (y_1, y_2) = (0, 0), \\ F\big(g_1(0, x), g_2(x)|x\big) + F\big(g_1(1, x), +\infty|x\big) & \\ \quad -F\big(g_1(1, x), g_2(x)|x\big), & \text{if } (y_1, y_2) = (0, 1), \\ F\big(+\infty, g_2(x)|x\big) - F\big(g_1(0, x), g_2(x)|x\big), & \text{if } (y_1, y_2) = (1, 0), \\ 1, & \text{if } (y_1, y_2) = (1, 1). \end{cases}\end{aligned} \tag{16}$$

From (16), one can see that to satisfy (U2.ω) in Theorem 1, $\mathcal{U}_x(g)$ must be chosen so that it contains the set

$$\big\{g_1(0, x), g_1(1, x), \pm\infty\big\} \times \big\{g_2(x), \pm\infty\big\} \tag{17}$$

for any fixed $g \in \mathcal{G}^\dagger$.[24]

In the remainder of this section, I demonstrate the application of Theorem 1 to this model under a variety of assumptions, represented primarily through different specifications of $\mathcal{F}^\dagger$. These different specifications correspond to different assumptions about

[23] For example, in the model in Section 2, this occurs when changes in $\theta = \beta$ move β_0 from larger than 0 to smaller than 0, or when changes in β_1 change the ordering of β_0 and $\beta_0 + \beta_1 x$.

[24] Including $-\infty$ in these sets is not necessary for (U2.ω), but is always required for U1.

the conditional distribution of (U_1, U_2) given X. Throughout the demonstration, the index function in (14) is assumed to be linear in Y_2 and a scalar component X_1 of X, that is, g_1 is restricted to be an element of

$$\mathcal{G}_1^\dagger \equiv \big\{g_1 : g_1(y_2, x_1) = \beta_0 + \beta_1 y_2 + x_1, \text{ for some } (\beta_0, \beta_1) \in \mathbb{R}^2\big\}, \tag{18}$$

where the coefficient of x_1 is set at 1 to impose a scale normalization.[25]

In one set of assumptions, the model for Y_2 is taken to be non-restrictive by setting $g_2(x) = \mathbb{P}[Y_2 = 0|X = x]$, and restricting $\mathcal{F}^\dagger$ to only admit conditional distributions F under which U_2 is uniformly distributed on $[0, 1]$ and independent of U_1 and X.[26] This effectively reduces the bivariate model (14)–(15) to a univariate model. In a second set of assumptions, the model for Y_2 is made restrictive by allowing U_2 to be dependent with U_1, but maintaining the nonparametric index function and uniform normalization on U_2. In a third set of assumptions, the model for Y_2 is specified as having a linear index in X_1 and another scalar component X_2 by requiring g_2 to be an element of

$$\mathcal{G}_2^\dagger \equiv \big\{g_2 : g_2(x_1, x_2) = \gamma_0 + x_1 + \gamma_2 x_2, \text{ for some } (\gamma_0, \gamma_2) \in \mathbb{R}^2\big\}, \tag{19}$$

where the coefficient on x_1 has again been normalized to 1. Under these assumptions and the additional requirement that (U_1, U_2) is jointly normal, this model becomes the well-known bivariate probit model (Heckman (1978)). Conditions for point identification in this and other parametric bivariate binary response models were recently developed by Han and Vytlacil (2017). Here, I use the PIES approach to study identification without imposing any parametric assumptions on the distribution of (U_1, U_2).

I consider several specifications for the distribution of (X_1, X_2), since this can have an important impact on the identified set. In the first set of simulations, both X_1 and X_2 are taken to have discrete support $\mathcal{X} = \mathcal{X}_{1,d_1} \times \mathcal{X}_{2,d_2}$ for $d_1 \in \{3, 5, 7\}$, $d_2 \in \{3, 5\}$, with

$$\begin{aligned}
&\mathcal{X}_{1,3} = \{-1, 0, 1\}, \qquad \mathcal{X}_{2,3} = \{-1, 0, 1\},\\
&\mathcal{X}_{1,5} = \left\{-1, -\frac{1}{2}, 0, \frac{1}{2}, 1\right\}, \qquad \mathcal{X}_{2,5} = \{-2, -1, 0, 1, 2\},\\
&\mathcal{X}_{1,7} = \left\{-1, -\frac{2}{3}, -\frac{1}{3}, 0, \frac{1}{3}, \frac{2}{3}, 1\right\}.
\end{aligned} \tag{20}$$

In these simulations, I always compute sharp identified sets. At the end of this section, I also demonstrate the construction of arbitrarily tight outer sets when X_2 has a continuous distribution. In all cases, I take (X_1, X_2) to be uniformly distributed over its support with X_1 independent of X_2. I take the distribution of Y conditional on X to be that generated through (14)–(15) with parameters $(\beta_0, \beta_1, \gamma_0, \gamma_2) = (0.5, -0.75, 0.3, 0.2)$ when (U_1, U_2) follows a bivariate standard normal distribution.

Table 1 details nine specifications of assumptions that will be considered in sequence. All of the assumptions can be shown to yield systems of equations in Theorem 1

[25]The normalization could be achieved by the weaker condition that $|X_1| = 1$ without assuming that the sign is known a priori. This could be implemented by repeating each simulation for $X_1 = 1$ and $X_1 = -1$.

[26]Alternatively (and equivalently), (15) could be ignored, and Y_2 could be treated as a component of X.

TABLE 1. Assumptions Maintained in the Different Specifications

	Exogenous Y_2			Endogenous Y_2					
Assumption	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]
$\text{med}(U_1\|Y_2, X_1) = 0$	✓	✓	✓						
$U_1\|Y_2, X_1$ symmetric around 0		✓	✓						
$U_1 \perp\!\!\!\perp (Y_2, X_1)$			✓						
$\text{med}(U_1\|X_1, X_2) = 0$				✓	✓	✓	✓	✓	✓
$U_1\|X_1, X_2$ symmetric around 0						✓			✓
$U_1 \perp\!\!\!\perp (X_1, X_2)$					✓	✓	✓	✓	✓
$Y_2 = \mathbb{1}[g_{21}(X_1, X_2) > U_2]$							✓		
$(U_1, U_2) \perp\!\!\!\perp (X_1, X_2)$							✓	✓	✓
$Y_2 = \mathbb{1}[\pi_0 + X_1 + \pi_2 X_2 > U_2]$								✓	✓
$\text{med}(U_2\|X_1, X_2) = 0$								✓	✓
$U_2\|X_1, X_2$ symmetric around 0									✓

Note: Specifications [1]–[6] are single equation models of Y_1 only. In specifications [1]–[3], Y_2 and X_1 are assumed to be exogenous, using increasingly strong concepts of exogeneity. In specifications [4]–[6], Y_2 is potentially endogenous, and restrictions are placed on the relationship between (X_1, X_2) and U_1. These specifications can be seen as a single equation semiparametric instrumental variables model with X_2 as an excluded instrument. In specification [7], a nonparametric first stage equation is added, meaning that g_2 is specified nonparametrically but with (U_1, U_2) assumed to be independent of (X_1, X_2). In specification [8], the nonparametric first stage is replaced by the linear-in-parameters parametric form (19). In specification [9], conditional symmetry is added for both U_1 and U_2.

that are linear in $\overline{F}$, in the sense discussed in Section 3.4.[27] Specification [1] is the semiparametric binary response model discussed in Section 2, except with two explanatory variables and a different normalization on the index coefficients. Figures 2a–2c are plots of the identified set for the index coefficients, (β_0, β_1), for different values of d_1. Manski's (1975, 1985, 1988) conditions for point identification of these parameters require X_1 to be continuously distributed, which is not the case here. As a result, the index coefficients are partially identified, although the sizes of the identified sets decrease with the number of support points of X_1 (Horowitz (2009), Komarova (2013)). The first row of Table 2 reports identified sets for the ATE of Y_2 on Y_1 under different values of d_1 and d_2. While this model has been widely studied, it does not appear that any prior work has shown how to construct these identified sets for the ATE.

If U_1 were assumed to be normally distributed and independent of (Y_2, X_1), then the identified set for the index coefficients would be the singleton given by the "+" mark at $(0.5, -0.75)$, while the identified set for the ATE would be the singleton (either $\{0.218\}$ or $\{0.234\}$) indicated in the header of Table 2. The difference between these singleton sets and the identified sets for [1] represents the combined identifying content of independence and normality. These effects can be separated by adding assumptions to [1].

Specification [2] strengthens the median independence assumption to the assumption that the distribution of $U_1|Y_2, X_1$ is symmetric around 0. Manski (1988) reasoned that this would have no effect on the identified set for the index coefficients, an assertion which is supported by comparing Figures 2a–2c and 2d–2f. However, comparing

[27] As a result, the PIES procedure reduces to solving linear programs. I wrote these linear programs in AMPL (Fourer, Gay, and Kernighan (2002)) and solved them using Gurobi (Gurobi Optimization LLC (2015)).

TABLE 2. Sharp Identified Sets for the Specifications Listed in Table 1, for Different Values of (d_1, d_2)

(d_1, d_2)	(3, 3)	(5, 3)	(3, 5)	(5, 5)
DGP ATE	0.218	0.234	0.218	0.234
[1]	[0.052, 0.481]	[0.031, 0.491]	[0.052, 0.480]	[0.031, 0.490]
[2]	[0.065, 0.429]	[0.039, 0.427]	[0.065, 0.428]	[0.039, 0.426]
[3]	{0.218}	{0.234}	{0.218}	{0.234}
[4]	[−0.243, 0.756]	[−0.254, 0.735]	[−0.219, 0.754]	[−0.238, 0.733]
[5]	[−0.173, −0.027]	[−0.189, −0.030]	[−0.109, −0.053]	[−0.119, −0.060]
	∪ [0.027, 0.685]	∪ [0.030, 0.660]	∪ [0.053, 0.638]	∪ [0.060, 0.615]
[6]	[−0.173, −0.027]	[−0.189, −0.030]	[−0.109, −0.053]	[−0.119, −0.060]
	∪ [0.027, 0.666]	∪ [0.030, 0.629]	∪ [0.053, 0.616]	∪ [0.060, 0.583]
[7]	[0.027, 0.658]	[0.030, 0.569]	[0.053, 0.569]	[0.060, 0.521]
[8]	[0.027, 0.658]	[0.030, 0.569]	[0.053, 0.569]	[0.060, 0.521]
[9]	[0.027, 0.629]	[0.030, 0.544]	[0.053, 0.543]	[0.060, 0.499]

Note: The supports corresponding to different values of (d_1, d_2) are given in (20). Note that the second header row of the table lists the true value of the average treatment effect, which varies with d_1. As a result, the sets in the second and fourth columns are not necessarily strict subsets of those in (respectively) the first and third columns.

rows [1] and [2] in Table 2 shows that the symmetry assumption *does* have an effect on the size of the identified set for the ATE.

In specification [3], both symmetry and full independence are maintained, so that the identified set for the ATE collapses to a singleton, as expected. Nevertheless, Figures 2g–2i show that the index coefficients are still partially identified. A practical consequence is that using a binary response model for extrapolation to values of the explanatory variables not observed in the data will still lead to partial identification, even under full independence, unless one imposes a parametric assumption on the distribution of the unobservable term.[28]

In specification [4], Y_2 is allowed to be an endogenous variable. In particular, this specification imposes the assumption that the median of U_1 is 0, conditional on (X_1, X_2), but it does not restrict dependence between U_1 and Y_2. This is like an instrumental variables version of Manski's (1975, 1985, 1988) semiparametric binary response model with the instrument here being X_2. Hong and Tamer (2003) established point identification of this model under a support condition on (X_1, X_2) that is similar to the one required in the exogenous regressor case. In particular, their assumption is not satisfied when all components of (X_1, X_2) are discrete, as in this simulation. The sharp identified set for (β_0, β_1) is displayed in Figures 3a–3c. It is unbounded and much larger than the corresponding sets for [1] in Figures 2a–2c. This reflects the loss in identifying power of assuming exogeneity of the excluded instrument, X_2, rather than the variable

[28] One could also consider imposing full independence without symmetry, as in Cosslett (1983), Matzkin (1992), Klein and Spady (1993), and certain cases of Han (1987). The point identification results in these papers require support conditions on the explanatory variables, which are not satisfied in this simulation. Since the index parameters are partially identified under specification [3], the implication is that the support conditions are crucial for ensuring point identification. I am grateful to an anonymous referee for suggesting this point.

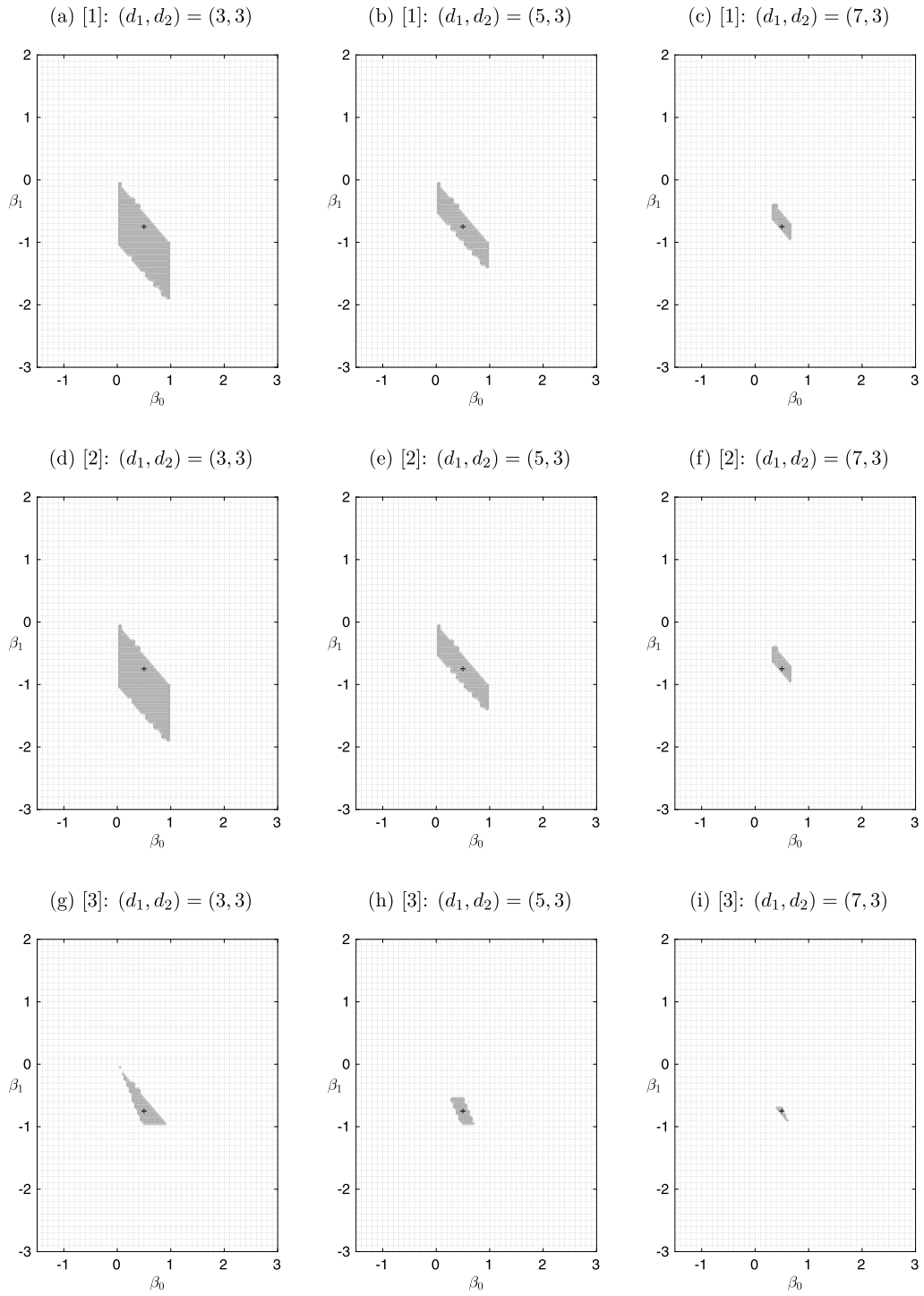

FIGURE 2. Sharp identified sets for (β_0, β_1) for specifications [1]–[3] in Table 1 for different values of (d_1, d_2). The shaded area indicates the identified set. The black "+" mark is placed at $(\beta_0, \beta_1) = (0.5, -0.75)$, which is the value in the data generating process.

of interest, Y_2. As in the exogenous case, the size of the identified set shrinks with the number points in $\mathcal{X}$.

In specification [5], the median independence condition in [4] is strengthened to full independence, that is, $U_1 \perp\!\!\!\perp (X_1, X_2)$. Sharp identified sets for the index coefficients under specification [5] are displayed in Figures 3d–3f. These identified sets are not only non-convex, but disconnected. As a result, the sharp identified set for the ATE is also disconnected.[29] Specification [6] adds the condition that U_1 is symmetrically distributed around 0, conditional on (X_1, X_2). This has a small effect on the identified set for the ATE. However, Figures 3g–3i show that, unlike in the exogenous case, here symmetry can also greatly reduce the size of the identified set for the index coefficients.

Specifications [7], [8], and [9] impose the full triangular model structure. In [7], it is only assumed that there exists some first stage equation with an unobservable term U_2 such that (U_1, U_2) is independent of (X_1, X_2). Chesher (2005), Shaikh and Vytlacil (2011), Jun, Pinkse, and Xu (2011, 2012), and Mourifié (2015) derived analytic expressions for fully nonparametric triangular bivariate binary response models.[30] In contrast, the linear index model considered here is semiparametric, which makes analytically characterizing its sharp identified set quite difficult. Intuitively, the source of this difficulty is that to establish sharpness in the linear index model, one needs to also fully exhaust the additional information carried in the linearity assumption. PIES provides a tractable computational framework under which this can be achieved.

Comparing specifications [5] and [7] in Figures 3d–3f and 4a–4c reveals that the assumption that (X_1, X_2) is exogenous with respect to both the first stage and outcome equations has an impact on the identified set for both the index coefficients and the ATE. As it turns out, imposing a linear index structure on the first stage, as in [8], has no additional impact on the identified sets for either the index coefficients or the ATE. Specification [9] also requires both U_1 and U_2 to be symmetrically distributed around 0. This substantially tightens the identified sets of the index coefficients relative to both [8] and [6], but has only a modest effect on the identified set of the ATE.

Specification [9] is about as close as one can get to a bivariate probit model without imposing the joint normality assumption. The size of the identified sets in [9] can be interpreted as the cost of removing a parametric assumption on the distribution of unobservables. That these identified sets are relatively uninformative implies that identification in the bivariate probit model is primarily driven by the normality assumption, at least in this data generating process.

As noted in Section 3, Theorem 1 applies regardless of the cardinality of $\mathcal{X}$. However, in practice the system of equations in Theorem 1 and the optimization problem in Proposition 1 can only be solved with a finite number of variables and constraints. An implication is that if X is continuously distributed, then one must choose a finite subset $\overline{\mathcal{X}}$ on which to solve these problems. As shown in Theorem 1, the resulting identified sets may be (non-sharp) outer identified sets. These outer identified sets become closer to

[29] This aligns with similar findings by Chesher (2010, 2013), Chiburis (2010), and Chesher and Smolinski (2012).

[30] See also Vytlacil and Yıldız (2007) and Yıldız (2013), who established point identification by using exogenous variation in the outcome equation under some additional support conditions.

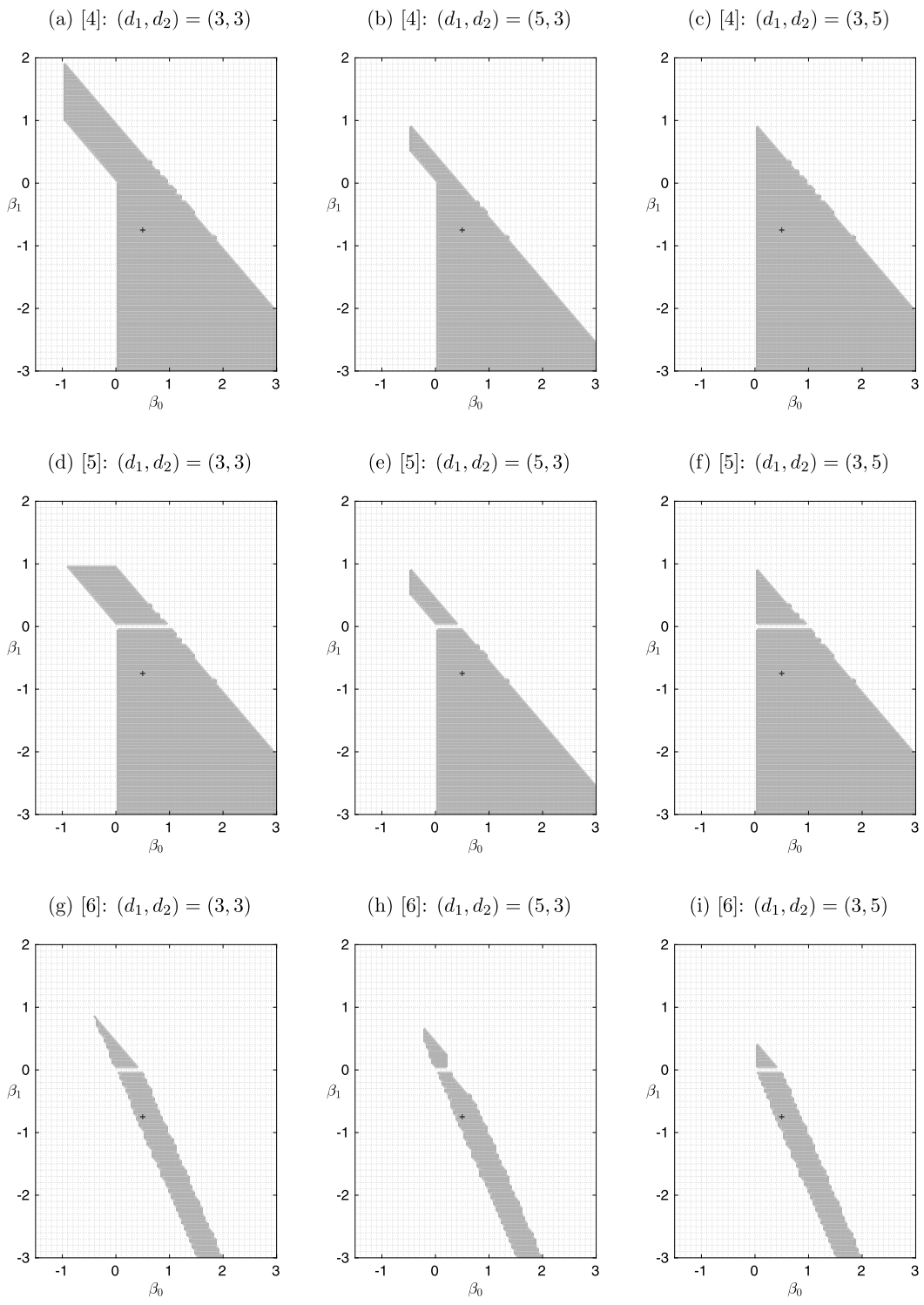

FIGURE 3. Sharp identified sets for (β_0, β_1) for specifications [4]–[6] in Table 1.

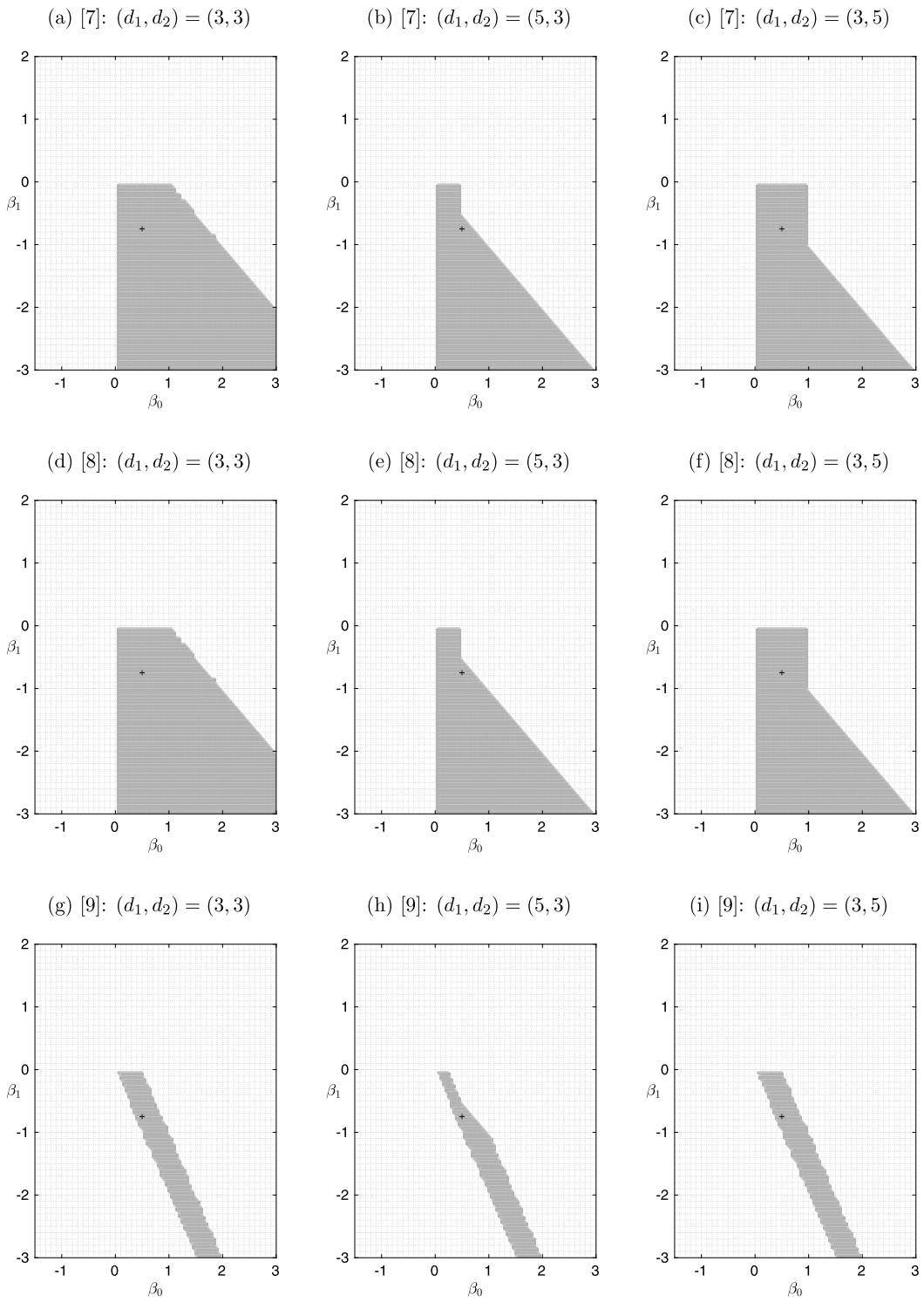

FIGURE 4. Sharp identified sets for (β_0, β_1) for specifications [7]–[9] in Table 1.

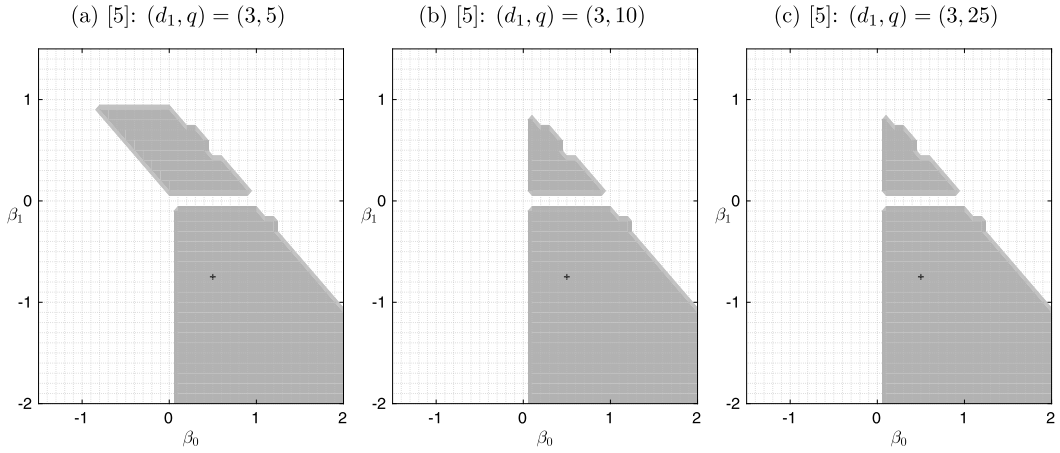

FIGURE 5. Outer identified sets for (β_0, β_1) in specification [5] when X_2 is continuously distributed and $\overline{\mathcal{X}} = \mathcal{X}_{1,d_1} \times \overline{\mathcal{X}}_{2,q}$ for $d_1 = 3$ and three values of q.

TABLE 3. Outer identified sets for the average treatment effect of Y_2 on Y_1, conditional on $X = (0, 0)$ in specification [5] when X_2 is continuously distributed and $\overline{\mathcal{X}} = \mathcal{X}_{1,d_1} \times \overline{\mathcal{X}}_{2,q}$ for $d_1 = 3$ and increasing values of q

q	Outer Set	Length
5	$[-0.227, -0.053] \cup [0.053, 0.591]$	$0.175 + 0.539 = 0.713$
10	$[-0.189, -0.070] \cup [0.070, 0.570]$	$0.120 + 0.500 = 0.620$
15	$[-0.177, -0.075] \cup [0.075, 0.563]$	$0.102 + 0.488 = 0.590$
20	$[-0.171, -0.078] \cup [0.078, 0.560]$	$0.093 + 0.482 = 0.574$
25	$[-0.167, -0.080] \cup [0.080, 0.558]$	$0.087 + 0.478 = 0.565$

the sharp identified set as $\overline{\mathcal{X}}$ becomes closer to $\mathcal{X}$. If the population is known, as in these simulations, then the only drawback of increasing $\overline{\mathcal{X}}$ is increased computation.

To demonstrate this point, I return to specification [5], but now I take X_2 to be uniformly distributed over the interval $[-2, 2]$. I then construct outer identified sets using $\overline{\mathcal{X}} = \mathcal{X}_{1,d_1} \times \overline{\mathcal{X}}_{2,q}$, where $\mathcal{X}_{1,d_1}$ remains as in (20), but $\overline{\mathcal{X}}_{2,q}$ is now a set of $(q-1)$ equally spaced quantiles of X_2. I repeat the simulation for $q \in \{5, 10, 15, 20, 25\}$.[31] Outer identified sets for the index coefficients are shown in Figure 5 for three choices of q, while those for the ATE of Y_2 on Y_1, conditional on $(X_1, X_2) = (0, 0)$, are shown in Table 3. As expected, the sets become smaller as q increases at a rate that also decreases with q.

The simulations in this section were conducted by forming grids over the index coefficients and then applying Proposition 1. This type of grid search approach is commonly used in general procedures for constructing sharp identified sets. However, using a grid search introduces a curse of dimensionality which severely limits the number of index coefficients that can be used in practice. For the PIES argument, Proposition 2 suggested that a grid search could be needlessly inefficient.

[31] So, for example, $\overline{\mathcal{X}}_{2,10} = \{0.1, 0.2, \ldots, 0.9\}$ and $\overline{\mathcal{X}}_{2,25} = \{0.04, 0.08, \ldots, 0.96\}$.

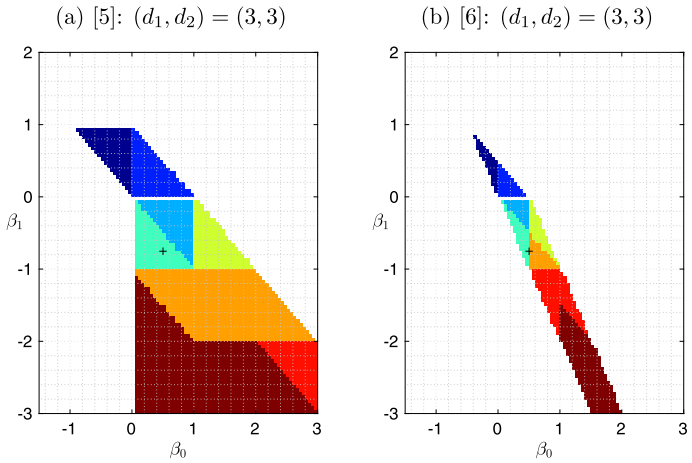

FIGURE 6. The partitioning property discussed in Proposition 2 for specifications [5] and [6].

Figures 6a and 6b provide supporting evidence of this inefficiency. Each colored region in the figures represents a piece of the partition discussed in connection with Proposition 2. Solving the programs in Proposition 1 more than once per colored region is inefficient, since Proposition 2 implies that the programs for any other (β_0, β_1) in the region will be equivalent. An algorithm that could determine the partition ex ante would reap enormous computational gains, perhaps breaking the curse of dimensionality associated with a grid search. Constructing such an algorithm is not a straightforward problem, even in the specific bivariate binary response model considered in these simulations, but may be a good avenue for future research.

5. CONCLUSION

This paper contains the development of a general procedure for constructing sharp identified sets called partial identification by extending subdistributions, or PIES. The PIES method is based on copula theory, and in particular uses a generalization of Sklar's (1959) multilinear interpolation lemma to determine when an admissible, observationally equivalent distribution function can be reconstructed as an extension of a lower-dimensional subdistribution function. This procedure is natural and intuitive, since it works directly with distribution functions, the properties of which are widely understood by economists. Its primary advantage is that it enables the construction of sharp identified sets for low-dimensional functionals of infinite-dimensional parameters without requiring one to first construct a sharp identified set for the infinite-dimensional parameters. I used PIES to study identification in a bivariate binary response model with a linear index function but without any parametric assumptions for the distribution of unobservables. Simulation results from that analysis suggest that point identification in the bivariate probit model can be largely driven by the assumption that the latent variables are normally distributed.

Appendix A: Copula theory

In this appendix, I record some definitions and results from copula theory that are used in the proof of Lemma 2. See Schweizer and Sklar (1983) and Nelsen (2006) for further discussion.

An L-dimensional copula is similar to a distribution function, but with domain $[0,1]^L$ and margins equal to the identity function. A subcopula is like a copula, but not necessarily defined on the entirety of $[0,1]^L$.

Definition C. Let $\mathcal{U}$ be a subset of $[0,1]^L$ such that $\mathcal{U} = \mathcal{U}_1 \times \cdots \times \mathcal{U}_L$, where $\mathcal{U}_l \subseteq [0,1]$ is such that $\{0,1\} \subseteq \mathcal{U}_l$ for each l. An L-dimensional *subcopula* is a function C with domain $\mathcal{U}$ such that

C1. C is L-increasing.

C2. $C(u) = 0$ for any $u \in \mathcal{U}$ that has at least one component equal to 0.

C3. $C(u) = u_l$ for any $u = (u_1, \ldots, u_L) \in \mathcal{U}$ that has all components except the lth equal to 1.

An L-dimensional *copula* (or, for emphasis, a *proper* L-dimensional copula) is an L-dimensional subcopula for which $\mathcal{U} = [0,1]^L$.

The fundamental result in copula theory is a two-part theorem due to Sklar (1959); see, for example, Nelsen (2006) for a modern treatment. The first part of Sklar's theorem shows that any distribution function can be decomposed into its marginal distributions and a possibly non-unique copula. The second part shows that, conversely, a copula combined with a collection of L one-dimensional distribution functions generates an L-dimensional distribution function.

Sklar's Theorem. 1. Let F be a proper L-dimensional distribution function with margins $F_l : \overline{\mathbb{R}} \to [0,1]$ defined as in Definition 3. Then there exists a proper L-dimensional copula C such that $F(u) = C(F_1(u_1), \ldots, F_L(u_L))$ for all $u \equiv (u_1, \ldots, u_L) \in \overline{\mathbb{R}}^L$. If F_l is continuous on $\overline{\mathbb{R}}$ for every $l = 1, \ldots, L$, then C is unique; otherwise, C is uniquely determined on $\{F_1(u_1) : u_1 \in \overline{\mathbb{R}}\} \times \cdots \times \{F_L(u_L) : u_L \in \overline{\mathbb{R}}\}$.

2. If C is a proper L-dimensional copula and F_l are proper one-dimensional distribution functions for each $l = 1, \ldots, L$, then the function $F : \overline{\mathbb{R}}^L \to [0,1] : F(u) \equiv C(F_1(u_1), \ldots, F_L(u_L))$ is a proper L-dimensional distribution function with margins F_l for each $l = 1, \ldots, L$.

The analysis in this paper uses the second part of Sklar's theorem, which is easy to prove. It does not use the first part of Sklar's theorem as stated, but it does make important use of the main lemma employed in Sklar's original proof of the first part. The proof of this lemma for $L = 2$ can be found in Nelsen's (2006) monograph on copula theory (see Lemma 2.3.5). Sklar (1996) provided a proof for higher dimensions; see also the discussion of Theorem 6.2.6 in Schweizer and Sklar (1983).

SKLAR'S LEMMA. Let $\overline{C}$ be an L-dimensional subcopula with domain $\mathcal{U}$. Then there exists a (typically non-unique) proper L-dimensional copula C such that $C(u) = \overline{C}(u)$ for all $u \in \mathcal{U}$.

APPENDIX B: PROOFS FOR SECTION 3.1

PROOF OF LEMMA 1. Conditions 2b and 2c follow immediately from the fact that F is an L-dimensional subdistribution. Condition 2a, which in the one-dimensional case reduces to the usual notion of weakly increasing, is implied by Lemma 6.1.5 of Schweizer and Sklar (1983). □

The proof of Lemma 2 is characteristically different in the $L = 1$ and $L > 1$ cases. Since the $L > 1$ case makes use of the $L = 1$ case, I begin with the latter.

PROOF OF LEMMA 2 (CASE $L = 1$). Suppose that $\overline{F}$ is a one-dimensional subdistribution with domain $\mathcal{U}$ that is closed and contains $\{\pm\infty\}$. A proper one-dimensional distribution function that extends $\overline{F}$ can be constructed as follows, although many other constructions are possible.

First, let $\overline{\mathcal{U}}$ denote the set $\mathcal{U} \setminus \{\pm\infty\}$, and let $\underline{u} \equiv \inf\overline{\mathcal{U}}$ and $\overline{u} \equiv \sup\overline{\mathcal{U}}$, noting that $\underline{u}, \overline{u} \in \mathcal{U}$, since $\mathcal{U}$ is closed. Partition the complement of $\overline{\mathcal{U}}$, that is, $\overline{\mathcal{U}}^c \equiv \mathbb{R} \setminus \overline{\mathcal{U}}$, into three sets

$$\overline{\mathcal{U}}^c = \overline{\mathcal{U}}^c_- \cup \overline{\mathcal{U}}^c_0 \cup \overline{\mathcal{U}}^c_+, \tag{21}$$

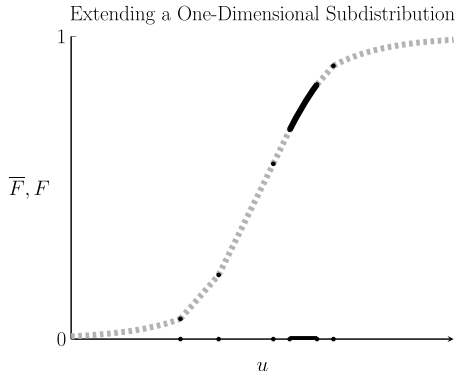

FIGURE 7. An example of the construction used in Lemma 2. The set $\overline{\mathcal{U}}$ is shown in bold on the horizontal axis and the subdistribution $\overline{F}$ is plotted in bold on its domain $\overline{\mathcal{U}}$. On $\overline{\mathcal{U}}$, F must be chosen to match $\overline{F}$. In between the smallest and largest elements of $\overline{\mathcal{U}}$, F is chosen to linearly interpolate between values of $\overline{F}$. Outside of this range, F is chosen to have the shape of an exponential distribution. Many other constructions are possible.

where $\overline{\mathcal{U}}^c_- \equiv \{u \in \overline{\mathcal{U}}^c : u < \underline{u}\}$, $\overline{\mathcal{U}}^c_+ \equiv \{u \in \overline{\mathcal{U}}^c : u > \overline{u}\}$, and $\overline{\mathcal{U}}^c_0 = \{u \in \overline{\mathcal{U}}^c : \underline{u} \leq u \leq \overline{u}\}$. Then define $F : \overline{\mathbb{R}} \to [0,1]$ as

$$F(u) = \begin{cases} 0, & \text{if } u = -\infty, \\ \overline{F}(\underline{u})e^{u-\underline{u}}, & \text{if } u \in \overline{\mathcal{U}}^c_-, \\ \ell\big(u; [a_u, b_u]\big), & \text{if } u \in \overline{\mathcal{U}}^c_0, \\ \overline{F}(u), & \text{if } u \in \overline{\mathcal{U}}, \\ \overline{F}(\overline{u}) + \big(1 - \overline{F}(\overline{u})\big)\big(1 - e^{\overline{u}-u}\big), & \text{if } u \in \overline{\mathcal{U}}^c_+, \\ 1, & \text{if } u = +\infty, \end{cases} \tag{22}$$

where $a_u \equiv \sup\{u' \in \overline{\mathcal{U}} : u' \leq u\}$, $b_u \equiv \inf\{u' \in \overline{\mathcal{U}} : u' \geq u\}$, and $\ell(\cdot; [a_u, b_u])$ is the linear function that interpolates $\overline{F}$ between a_u and b_u, that is,

$$\ell\big(u; [a_u, b_u]\big) \equiv \overline{F}(a_u) + \left(\frac{\overline{F}(b_u) - \overline{F}(a_u)}{b_u - a_u}\right)(u - a_u), \tag{23}$$

noting that $a_u < b_u$ for all $u \in \overline{\mathcal{U}}^c_0$. In words, F is equal to $\overline{F}$ on $\overline{\mathcal{U}}$, it has the shape of an exponential distribution for $u \in \overline{\mathcal{U}}^c_-$ and $u \in \overline{\mathcal{U}}^c_+$, and it linearly interpolates $\overline{F}$ on $\overline{\mathcal{U}}^c_0$. See Figure 7. It is straightforward to verify that F is a proper one-dimensional distribution function and that, by construction, $F(u) = \overline{F}(u)$ for every $u \in \mathcal{U}$. It is also straightforward to verify that if $\overline{F}$ is continuous on $\mathcal{U}$, then F is continuous on its entire domain.

Alternatively, suppose that $\mathcal{U}$ is a finite set and let u^+ be any number such that $u^+ \geq \max\overline{\mathcal{U}}$. Enumerate the elements of $\mathcal{U} \cup \{u^+\}$ as $\{u_j\}_{j=0}^J$, so that $u_0 = -\infty$, $u_{J-1} = u^+$, and $u_J = +\infty$. Instead of (22), define

$$F : \overline{\mathbb{R}} \to [0,1] : \quad F(u) = \sum_{j=0}^{J-2} \overline{F}(u_j)\mathbb{1}\big[u \in [u_j, u_{j+1})\big] + \mathbb{1}[u \geq u_{J-1}].$$

Then, by construction, F is a proper one-dimensional distribution function with support $\overline{\mathcal{U}} \cup \{u^+\}$, and such that $F(u) = \overline{F}(u)$ for all $u \in \mathcal{U}$. □

Proof of Lemma 2 (case $L > 1$). Suppose that $L > 1$ and let $\mathcal{U}$ and $\overline{F}$ be as in the statement of the lemma. For $l = 1, \ldots, L$, let $\overline{F}_l$ be the lth margin of $\overline{F}$, that is,

$$\overline{F}_l : \mathcal{U}_l \to [0,1] : \quad \overline{F}_l(u_l) \equiv \overline{F}(+\infty, \ldots, +\infty, u_l, +\infty, \ldots, +\infty).$$

Lemma 1 shows that each $\overline{F}_l$ is itself a one-dimensional subdistribution function.

Next, define the set

$$\mathcal{T} \equiv \mathcal{T}_1 \times \cdots \times \mathcal{T}_L \equiv \big\{\overline{F}_1(u_1) : u_1 \in \mathcal{U}_1\big\} \times \cdots \times \big\{\overline{F}_L(u_L) : u_L \in \mathcal{U}_L\big\},$$

and define the function $\overline{C} : \mathcal{T} \to [0,1]$ by

$$\overline{C}\big(\overline{F}_1(u_1), \ldots, \overline{F}_L(u_L)\big) \equiv \overline{F}(u_1, \ldots, u_L).$$

Note that $\overline{C}$ is well-defined, because if u, u' are such that $\overline{F}_l(u_l) = \overline{F}_l(u'_l)$ for all l, then $\overline{F}(u) = \overline{F}(u')$; see Lemma 2.10.4 of Nelsen (2006) or Lemma 6.1.9 of Schweizer and Sklar (1983).

I claim that $\overline{C}$ is a subcopula on $\mathcal{T}$. To see this, first note that $\mathcal{T} \subseteq [0,1]^L$ and that $\{0,1\} \subseteq \mathcal{T}_l$ for each l, since by assumption $\{\pm\infty\} \subset \mathcal{U}_l$ with $\overline{F}_l(-\infty) = 0$ and $\overline{F}_l(+\infty) = 1$. Next, notice that

$$\begin{aligned}\overline{C}\big(0, \overline{F}_2(u_2), \ldots, \overline{F}_L(u_L)\big) &= \overline{C}\big(\overline{F}_1(-\infty), \overline{F}_2(u_2), \ldots, \overline{F}_L(u_L)\big) \\ &\equiv \overline{F}(-\infty, u_2, \ldots, u_L) = 0,\end{aligned}$$

and similarly if any other collection of the arguments of $\overline{C}$ are 0. This shows that Condition C2 is satisfied. Condition C3 is satisfied because, by construction,

$$\begin{aligned}\overline{C}\big(\overline{F}_1(u_1), 1, \ldots, 1\big) &= \overline{C}\big(\overline{F}_1(u_1), \overline{F}_2(+\infty), \ldots, \overline{F}_L(+\infty)\big) \\ &\equiv \overline{F}(u_1, +\infty, \ldots, +\infty) \equiv \overline{F}_1(u_1),\end{aligned}$$

and similarly for any other index $l = 2, \ldots, L$.

To see that $\overline{C}$ satisfies Condition C1, consider any $t', t'' \in \mathcal{T}$ such that $t' \leq t''$. Then there exist $u', u'' \in \mathcal{U}$ such that $u' \leq u''$ with $t' = (\overline{F}_1(u'_1), \ldots, \overline{F}_L(u'_L))$ and $t'' = (\overline{F}_1(u''_1), \ldots, \overline{F}_L(u''_L))$.[32] Consider the sets of vertices $\mathrm{Vrt}(t', t'')$ and $\mathrm{Vrt}(u', u'')$, and the function $\xi : \mathrm{Vrt}(t', t'') \to \mathrm{Vrt}(u', u'')$ defined by $\xi(t) = (\xi_1(t_1), \ldots, \xi_L(t_L))$, where for each l,

$$\xi_l : \{t'_l, t''_l\} \to \{u'_l, u''_l\} : \quad \xi_l(t_l) \equiv \begin{cases} u'_l, & \text{if } t_l = t'_l, \\ u''_l, & \text{if } t_l = t''_l. \end{cases}$$

Then, by construction, ξ is bijective and $\mathrm{sgn}_{(t',t'')}(t) = \mathrm{sgn}_{(u',u'')}(\xi(t))$ for every $t \in \mathrm{Vrt}(t', t'')$. Moreover, $\overline{C}(t) = \overline{F}(\xi(t))$ for any $t \in \mathrm{Vrt}(t', t'')$. Thus,

$$\begin{aligned}\mathrm{Vol}_{\overline{C}}(t', t'') &\equiv \sum_{t \in \mathrm{Vrt}(t', t'')} \mathrm{sgn}_{(t',t'')}(t)\overline{C}(t) \\ &= \sum_{t \in \mathrm{Vrt}(t', t'')} \mathrm{sgn}_{(u',u'')}\big(\xi(t)\big)\overline{F}\big(\xi(t)\big) \\ &= \sum_{u \in \mathrm{Vrt}(u', u'')} \mathrm{sgn}_{(u',u'')}(u)\overline{F}(u) \equiv \mathrm{Vol}_{\overline{F}}(u', u''),\end{aligned} \tag{24}$$

where the first equality is by definition, the second equality imposes the above observations, the third equation changes the indexing variable from t to $u = \xi(t)$ by using the bijectivity of ξ and $\mathrm{Vrt}(t', t'') = \xi^{-1}(\mathrm{Vrt}(u', u''))$, while the final equality is by definition. By assumption, $\mathrm{Vol}_{\overline{F}}(u', u'') \geq 0$ since $u', u'' \in \mathcal{U}$ with $u' \leq u''$. Thus, by (24), $\mathrm{Vol}_{\overline{C}}(t', t'') \geq 0$ as well. Since t', t'' were arbitrary elements of $\mathcal{T}$ with $t' \leq t''$, this shows that $\overline{C}$ satisfies C1.

[32] Specifically, note that by the definition of $\mathcal{T}_l$, there exists at least one u'_l such that $t'_l = \overline{F}_l(u'_l)$. If $t''_l = t'_l$, then take $u''_l = u'_l$. Otherwise, if $t''_l > t'_l$, then let u''_l be any such that $t''_l = \overline{F}_l(u''_l)$. This choice ensures that $u'_l \leq u''_l$ for each l, since each $\overline{F}_l$ is weakly increasing.

Since $\overline{C}$ satisfies C1–C3, it is an L-dimensional subcopula with domain $\mathcal{T}$. By Sklar's lemma, $\overline{C}$ can be extended (perhaps non-uniquely) to a proper L-dimensional copula C with domain $[0,1]^L$ such that $C(t) = \overline{C}(t)$ for every $t \in \mathcal{T}$. Moreover, the $L = 1$ case of the current lemma shows that, for each l, there exists a proper one-dimensional distribution function $F_l : \overline{\mathbb{R}} \to [0,1]$ such that $F_l(u) = \overline{F}_l(u)$ for all $u_l \in \mathcal{U}_l$. Define the function

$$F : \overline{\mathbb{R}}^L \to [0,1] : \quad F(u) = C\big(F_1(u_1), \ldots, F_L(u_L)\big). \tag{25}$$

Since each F_l is a proper one-dimensional distribution function and C is a proper copula, Sklar's theorem shows that F is a proper L-dimensional joint distribution function. Moreover, F is an extension of $\overline{F}$, since for any $u \in \mathcal{U}$,

$$\begin{aligned} F(u) &\equiv C\big(F_1(u_1), \ldots, F_L(u_L)\big) \\ &= C\big(\overline{F}_1(u_1), \ldots, \overline{F}_L(u_L)\big) \\ &= \overline{C}\big(\overline{F}_1(u_1), \ldots, \overline{F}_L(u_L)\big) = \overline{F}(u). \end{aligned}$$

Finally, note that copulas are continuous (Theorem 2.2.4 of Nelsen (2006)). As a consequence, if $\overline{F}_l$ is continuous on $\mathcal{U}_l$ for each l, so that each F_l can be taken to be continuous, then the continuity of C implies that F is also continuous. Alternatively, if $\mathcal{U}$ is a finite set, so that each $\mathcal{U}_l$ is also a finite set, then each F_l can be taken to have support $(\mathcal{U}_l \setminus \{\pm\infty\}) \cup \{u_l^+\}$ for any $u_l^+ \in \mathbb{R}$ with $u_l^+ \geq \max(\mathcal{U}_l \setminus \{+\infty\})$, so that F has support contained in the Cartesian product of these sets. □

Proof of Corollary 1. The case where $L = 1$ follows tautologically from the assumption and definition of extendibility, so assume that $L > 1$. Follow the proof of Lemma 2 up to (25). By assumption, for each $l = 1, \ldots, L$, there exists an $F_l \in \mathcal{F}_l$ such that $F_l(u_l) = \overline{F}_l(u_l)$ for each $u_l \in \mathcal{U}_l$. Then defining F as in (25) with these margins F_l for $l = 1, \ldots, L$ and appealing to Sklar's theorem shows that F is a proper L-dimensional distribution function with margins $F_l \in \mathcal{F}_l$ for each $l = 1, \ldots, L$. □

Proof for Example 4. Reducibility follows immediately by the properties of $\mathcal{F}$, $\overline{\mathcal{F}}$, and $\mathcal{U}$. To see that $\overline{\mathcal{F}}$ is extendible to $\mathcal{F}$, let $\overline{F}$ be an arbitrary element of $\overline{\mathcal{F}}$ and define a function $\overline{F}^-$ on $\mathcal{U}$ by

$$\overline{F}^- : \mathcal{U} \to [0,1] : \quad \overline{F}^-(u) = \begin{cases} 2\overline{F}(u), & \text{if } u \leq 0, \\ 1, & \text{if } u > 0. \end{cases}$$

Observe that $\overline{F}^-$ is also a one-dimensional subdistribution, and let $F^- : \overline{\mathbb{R}} \to [0,1]$ be an extension of $\overline{F}^-$, which exists by Lemma 2. Then define $F : \overline{\mathbb{R}} \to [0,1]$ by

$$F(u) = \begin{cases} \dfrac{1}{2}F^-(u), & \text{if } u \leq 0, \\ 1 - \dfrac{1}{2}F^-(-u), & \text{if } u > 0, \end{cases}$$

so that $F(u) = 1 - F(-u)$ for all u by construction, that is, $F \in \mathcal{F}$. In addition, F extends $\overline{F}$ because if $u \in \mathcal{U}$ and $u \leq 0$, then

$$F(u) = \frac{1}{2}F^{-}(u) = \frac{1}{2}\overline{F}^{-}(u) = \overline{F}(u),$$

while if $u \in \mathcal{U}$ and $u \geq 0$, then

$$F(u) = 1 - \frac{1}{2}F^{-}(-u) = 1 - \frac{1}{2}\overline{F}^{-}(-u) = 1 - \overline{F}(-u) = \overline{F}(u).$$

Thus, $\overline{\mathcal{F}}$ is extendible to $\mathcal{F}$. □

Appendix C: Proofs for Section 3

Proof of Theorem 1. Suppose that $p \in \mathcal{P}^{\star}$. By definition, there exists an $S = (\theta, F) \in \mathcal{S}^{\star}$ such that $\pi(\theta, F) = p$. For every $x \in \overline{\mathcal{X}}$ and $u \in \mathcal{U}_x(\theta)$, let $\overline{F}(u|x) = F(u|x)$. Then

1. Condition U1 implies that (T1.2) is satisfied, since the restriction of any distribution function to a subset satisfying the properties of $\mathcal{U}_x(\theta)$ is a subdistribution on that subset.
2. Condition U2, $(\theta, F) \in \mathcal{S}^{\star}$, and A4 imply that (T1.1), (T1.6), and (T1.7) are satisfied.
3. Condition U3 and A3 imply that (T1.5) is satisfied.
4. Condition U4 and A1 imply that (T1.3) is satisfied.
5. Condition U5 and A2 imply that (T1.4) is satisfied.

Thus, there is a $\theta \in \Theta^{\dagger}$ and functions $\overline{F}(\cdot|x) : \mathcal{U}_x(\theta) \to [0,1]$ for $x \in \overline{\mathcal{X}}$ that satisfy (T1.1)–(T1.7).

Conversely, suppose that $\overline{\mathcal{X}} = \mathcal{X}$, $\overline{\mathcal{Y}} = \mathcal{Y}$ and that there is a $\theta \in \Theta^{\dagger}$ and functions $\overline{F}(\cdot|x) : \mathcal{U}_x(\theta) \to [0,1]$ for $x \in \mathcal{X}$ such that (T1.1)–(T1.7) are satisfied. The proof that $p \in \mathcal{P}^{\star}$ will use an extension argument based on Lemma 2 and Corollary 1. The extension argument will be conducted for each $x \in \mathcal{X}$, while respecting the requirements of Assumptions A. To do this, I begin by partitioning $\mathcal{X}$ into different subsets defined by the sets $\mathcal{X}^{\dagger}_{0,m_0}$, $\mathcal{X}^{\dagger}_{l,m_l}$ in A1 and A2.

To this end, first define $\mathcal{X}^{\dagger}_{l,0} \equiv \mathcal{X} \setminus \bigcup_{m_l=1}^{M_l} \mathcal{X}^{\dagger}_{l,m_l}$ for each $l = 0, 1, \ldots, L$. Then $\{\mathcal{X}^{\dagger}_{l,m_l}\}_{m_l=0}^{M_l}$ forms a partition of $\mathcal{X}$ for each l, given the stipulation that $\{\mathcal{X}^{\dagger}_{l,m_l}\}_{m_l=1}^{M_l}$ are disjoint sets. Next, let $\mathcal{M} \equiv \bigotimes_{l=0}^{L}\{0, 1, \ldots, M_l\}$. For each $m \equiv (m_0, m_1, \ldots, m_L) \in \mathcal{M}$, define

$$\mathcal{X}^m \equiv \left\{x \in \mathcal{X} : x \in \mathcal{X}^{\dagger}_{l,m_l} \text{ for each } l = 0, 1, \ldots, L\right\}.$$

Then $\{\mathcal{X}^m : m \in \mathcal{M}\}$ forms a partition of $\mathcal{X}$.

There are three types of sets in this partition, and the extension argument will differ slightly for each:

(i) full independence regions, that is, those with any m such that $m_0 > 0$;

(ii) marginal independence regions, that is, those with $m_0 = 0$ and $m_l > 0$ for at least one $l \geq 1$;

(iii) the region with no independence conditions, that is, $m_l = 0$ for all l.

I consider the three regions in turn. On each region, I will construct proper conditional distribution functions that satisfy the requirements of Assumption A. Then, I will use the fact that $\{\mathcal{X}^m : m \in \mathcal{M}\}$ forms a partition to link these conditional distribution functions together into a single element of $\mathcal{F}^\dagger$.

(i) Fix an $m \in \mathcal{M}$ with $m_0 > 0$, and fix an arbitrary $\overline{x} \in \mathcal{X}^m$. By (T1.2), $\overline{F}(\cdot|\overline{x})$ is a subdistribution with domain $\mathcal{U}_{\overline{x}}(\theta)$. By (T1.5), $\overline{F}_l(\cdot|\overline{x}) \in \overline{\mathcal{F}}^\dagger_{l,\overline{x}}$ for each $l = 1, \ldots, L$. Thus, given U3, Corollary 1 implies that there exists a proper L-dimensional distribution function $\widetilde{F}^m$ such that $\widetilde{F}^m(u) = \overline{F}(u|\overline{x})$ for all $u \in \mathcal{U}_{\overline{x}}(\theta)$, and such that the lth margin of $\widetilde{F}^m$ is in $\mathcal{F}^\dagger_{l,\overline{x}}$ for every $l = 1, \ldots, L$. Define $F^m : \overline{\mathbb{R}}^L \times \mathcal{X}^m \to [0,1] : F^m(u|x) = \widetilde{F}^m(u)$ to be a conditional distribution function that is constant as a function of x over $\mathcal{X}^m$, and equal to the unconditional distribution function $\widetilde{F}^m$ that was just defined. Observe that A1 and A3 together require that $\mathcal{F}^\dagger_{l,x} = \mathcal{F}^\dagger_{l,x'}$ for all $x, x' \in \mathcal{X}^m$ and all $l = 1, \ldots, L$. Thus, the lth margin of $F^m(\cdot|x)$ is an element of $\mathcal{F}^\dagger_{l,\overline{x}} = \mathcal{F}^\dagger_{l,x}$ for all $x \in \mathcal{X}^m$ and every l. In addition, U4 and (T1.3) imply that $F^m(u|x) = \widetilde{F}^m(u) = \overline{F}(u|\overline{x}) = \overline{F}(u|x)$ for all $x \in \mathcal{X}^m$ and $u \in \mathcal{U}_x(\theta)$.

(ii) Fix an $m \in \mathcal{M}$ with $m_0 = 0$ and $m_l > 0$ for some $l \geq 1$, and fix an arbitrary $\overline{x} \in \mathcal{X}^m$. Let $\mathcal{L}^m \equiv \{l : m_l > 0\}$. For each $l \in \mathcal{L}^m$, let $\overline{F}_l(\cdot|\overline{x})$ be the lth margin of $\overline{F}(\cdot|\overline{x})$. By (T1.2) and Lemma 1, each $\overline{F}_l(\cdot|\overline{x})$ is a one-dimensional subdistribution function. By (T1.5), $\overline{F}_l(\cdot|\overline{x}) \in \overline{\mathcal{F}}^\dagger_{l,\overline{x}}$ for each $l \in \mathcal{L}^m$. Thus, given U3, Corollary 1 implies that, for each $l \in \mathcal{L}^m$, there exists a proper distribution function $\widetilde{F}^m_l : \overline{\mathbb{R}} \to [0,1]$ such that $\widetilde{F}^m_l(u) = \overline{F}_l(u|\overline{x})$ for every $u \in \mathcal{U}_{l,\overline{x}}(\theta)$ and such that $\widetilde{F}^m_l \in \mathcal{F}^\dagger_{l,\overline{x}}$. Observe that A2 and A3 together require that $\mathcal{F}^\dagger_{l,x} = \mathcal{F}^\dagger_{l,x'}$ for all $x, x' \in \mathcal{X}^m$ and every $l \in \mathcal{L}^m$. Thus, $\widetilde{F}^m_l$ is an element of $\mathcal{F}^\dagger_{l,\overline{x}} = \mathcal{F}^\dagger_{l,x}$ for all $l \in \mathcal{L}^m$ and $x \in \mathcal{X}^m$. In addition, (T1.4) implies that $\widetilde{F}^m_l(u) = \overline{F}_l(u|\overline{x}) = \overline{F}_l(u|x)$ for all $x \in \mathcal{X}^m$, all $u \in \mathcal{U}_{l,x}(\theta)$, and each $l \in \mathcal{L}^m$.

These arguments show that for each $x \in \mathcal{X}^m$ and every $l \in \mathcal{L}^m$, the singleton set $\{\overline{F}_l(\cdot|x)\}$ is extendible to the singleton set $\{\widetilde{F}^m_l\}$. On the other hand, for each $x \in \mathcal{X}^m$ and every $l \notin \mathcal{L}^m$, (T1.5) and U3 imply that the singleton set $\{\overline{F}_l(\cdot|x)\}$ is extendible to $\mathcal{F}^\dagger_{l,x}$. Given (T1.2), Corollary 1 then implies that for each $x \in \mathcal{X}^m$, there exists a proper L-dimensional distribution function $F^m(\cdot|x)$ such that $F^m(u|x) = \overline{F}(u|x)$ for all $u \in \overline{\mathbb{R}}^L$. Moreover, the corollary establishes that the lth margins of $F^m(\cdot|x)$—call them $F^m_l(\cdot|x)$—are equal to $\widetilde{F}^m_l$ for every $l \in \mathcal{L}^m$, and are elements of $\mathcal{F}^\dagger_{l,x}$ for every $l \notin \mathcal{L}^m$. It follows that $F^m_l(\cdot|x) \in \mathcal{F}^\dagger_{l,x}$ for every $l = 1, \ldots, L$ and all $x \in \mathcal{X}^m$, and that $F^m_l(u|x) = \widetilde{F}_l(u) = F^m_l(u|x')$ for every $l \in \mathcal{L}^m$ and all $x, x' \in \mathcal{X}^m$.

(iii) Finally, consider the set $\mathcal{X}^m$ with $m = (0, 0, \ldots, 0)$. For each $x \in \mathcal{X}^m$, (T1.2) and (T1.5) with U3, together with an application of Corollary 1, establish the existence of a proper L-dimensional distribution function $F^m(\cdot|x)$ such that $F^m(u|x) = \overline{F}(u|x)$ for all $u \in \overline{\mathbb{R}}^L$. Moreover, the lth margins of this distribution are elements of $\mathcal{F}^\dagger_{l,x}$.

Combine the constructions in (i), (ii), and (iii) into

$$F : \overline{\mathbb{R}}^L \times \mathcal{X} \to [0,1] : \quad F(u|x) = \sum_{m \in \mathcal{M}} \mathbb{1}\big[x \in \mathcal{X}^m\big] F^m(u|x). \tag{26}$$

Observe that F is well-defined, since $\{\mathcal{X}^m : m \in \mathcal{M}\}$ forms a partition of $\mathcal{X}$. Given the previous discussion, it is known that for every $x \in \mathcal{X}$, $F(\cdot|x)$ is a proper L-dimensional

distribution function such that $F(u|x) = \overline{F}(u|x)$ for all $u \in \mathcal{U}_x(\theta)$. Moreover, the previous discussion established that F satisfies A1–A3. Given (T1.6), U2, and the established fact that each $F(\cdot|x)$ extends each $\overline{F}(\cdot|x)$, it is also known that $\rho(\theta, F) \geq 0$, so that A4 is satisfied. Thus, $F \in \mathcal{F}^\dagger$. Moreover, U2, (T1.1), and the assumption that $\overline{\mathcal{Y}} = \mathcal{Y}$, $\overline{\mathcal{X}} = \mathcal{X}$ imply that $\omega_{y|x}(\theta, F) = \mathbb{P}[Y \leq y|X = x]$ for all $y \in \mathcal{Y}$ and $x \in \mathcal{X}$, so that $(\theta, F) \in \mathcal{S}^\star$. Finally, U2 and (T1.7) imply that $\pi(\theta, F) = p$ which, since $(\theta, F) \in \mathcal{S}^\star$ and $\theta \in \Theta^\dagger$, implies that $p \in \mathcal{P}^\star$. □

Proof of Proposition 1. Since $\overline{\pi}(\theta, \cdot)$ is assumed to be continuous on $\overline{\mathcal{F}}^\star(\theta)$, and since $\overline{\mathcal{F}}^\star(\theta)$ is assumed to be connected, it follows that the image of $\overline{\mathcal{F}}^\star(\theta)$ under $\overline{\pi}(\theta, \cdot)$ is $[\underline{p}^\star(\theta), \overline{p}^\star(\theta)]$ for any $\theta \in \Theta^\dagger$; see, for example, Theorem 4.22 of Rudin (1976).

Now, suppose that $p \in \mathcal{P}^\star$. By Theorem 1, there exist a $\theta' \in \Theta^\dagger$ and an $\overline{F} \in \overline{\mathcal{F}}^\star(\theta')$ for which $p = \overline{\pi}(\theta', \overline{F})$. Thus, $p \in [\underline{p}^\star(\theta'), \overline{p}^\star(\theta')]$, so $p \in \bigcup_{\theta \in \Theta^\dagger} [\underline{p}^\star(\theta), \overline{p}^\star(\theta)]$, as well. Conversely, let $p \in \bigcup_{\theta \in \Theta^\dagger} [\underline{p}^\star(\theta), \overline{p}^\star(\theta)]$. Then there is a $\theta' \in \Theta^\dagger$ for which $[\underline{p}^\star(\theta'), \overline{p}^\star(\theta')]$ is nonempty with $p \in [\underline{p}^\star(\theta'), \overline{p}^\star(\theta')]$. Thus, there exists an $\overline{F} \in \overline{\mathcal{F}}^\star(\theta')$ such that $\overline{\pi}(\theta', \overline{F}) = p$. By Theorem 1, this implies that $p \in \mathcal{P}^\star$, given $\overline{\mathcal{X}} = \mathcal{X}$ and $\overline{\mathcal{Y}} = \mathcal{Y}$. □

Proof of Proposition 2. The result follows immediately from the hypothesis and definitions in Proposition 1. □

References

Alesina, A. and N. Fuchs-Schündeln (2007), "Good-bye Lenin (or not?): The effect of communism on people's preferences." *American Economic Review*, 97, 1507–1528. [106]

Altonji, J. G., T. E. Elder, and C. R. Taber (2005), "An evaluation of instrumental variable strategies for estimating the effects of catholic schooling." *The Journal of Human Resources*, 40, 791–821. [106]

Angrist, J. D. (2001), "Estimation of limited dependent variable models with dummy endogenous regressors." *Journal of Business & Economic Statistics*, 19, 2–28. [106]

Angrist, J. D. and J.-S. Pischke (2009), *Mostly harmless econometrics: An empiricist's companion*. Princeton University Press. [106, 108]

Ashraf, N., E. Field, and J. Lee (2014), "Household bargaining and excess fertility: An experimental study in Zambia." *American Economic Review*, 104, 2210–2237. [106]

Beresteanu, A., I. Molchanov, and F. Molinari (2011), "Sharp identification regions in models with convex moment predictions." *Econometrica*, 79, 1785–1821. [107]

Blevins, J. R. (2015), "Non-standard rates of convergence of criterion-function-based set estimators for binary response models." *The Econometrics Journal*, 18, 172–199. [108]

Chernozhukov, V. and C. Hansen (2005), "An IV model of quantile treatment effects." *Econometrica*, 73, 245–261. [114]

Chesher, A. (2005), "Nonparametric identification under discrete variation." *Econometrica*, 73, 1525–1550. [128]

Chesher, A. (2010), "Instrumental variable models for discrete outcomes." *Econometrica*, 78, 575–601. [107, 108, 114, 128]

Chesher, A. (2013), "Semiparametric structural models of binary response: Shape restrictions and partial identification." *Econometric Theory*, 29, 231–266. [108, 128]

Chesher, A. and A. Rosen (2012), "Simultaneous equations models for discrete outcomes: Coherence, completeness and identification." cemmap working paper 21/12. [107]

Chesher, A. and A. Rosen (2014), "An instrumental variable random coefficients model for binary outcomes." Econometrics Journal. [107]

Chesher, A. and A. M. Rosen (2013), "What do instrumental variable models deliver with discrete dependent variables?" *American Economic Review*, 103, 557–562. [107]

Chesher, A. and A. M. Rosen (2017), "Generalized instrumental variable models." *Econometrica*, 85, 959–989. [107]

Chesher, A. and K. Smolinski (2012), "IV models of ordered choice." *Journal of Econometrics*, 166, 33–48. [128]

Chiburis, R. C. (2010), "Semiparametric bounds on treatment effects." *Journal of Econometrics*, 159, 267–275. [113, 128]

Ciliberto, F. and E. Tamer (2009), "Market structure and multiple equilibria in airline markets." *Econometrica*, 77, 1791–1828. [117]

Cosslett, S. R. (1983), "Distribution-free maximum likelihood estimator of the binary choice model." *Econometrica*, 51, 765–782. [108, 126]

Demuynck, T. (2015), "Bounding average treatment effects: A linear programming approach." *Economics Letters*, 137, 75–77. [107]

Ekeland, I., A. Galichon, and M. Henry (2010), "Optimal transportation and the falsifiability of incompletely specified economic models." *Economic Theory*, 42, 355–374. [107]

Fourer, R., D. M. Gay, and B. W. Kernighan (2002), "AMPL: A modeling language for mathematical programming." Cengage Learning. [125]

Freyberger, J. and J. L. Horowitz (2015), "Identification and shape restrictions in nonparametric instrumental variables estimation." *Journal of Econometrics*, 189, 41–53. [107]

Galichon, A. and M. Henry (2009), "A test of non-identifying restrictions and confidence regions for partially identified parameters." *Journal of Econometrics*, 152, 186–196. [107]

Galichon, A. and M. Henry (2011), "Set identification in models with multiple equilibria." *The Review of Economic Studies*, 78, 1264–1298. [107]

Galichon, A. and M. Henry (2013), "Dilation bootstrap." *Journal of Econometrics*, 177, 109–115. [107]

Grogger, J. and D. Neal (2000), "Further evidence on the effects of catholic secondary schooling." Brookings-Wharton Papers on Urban Affairs, 151–201. [106]

Gurobi Optimization LLC (2015), "Gurobi optimizer reference manual." [125]

Han, A. K. (1987), "Non-parametric analysis of a generalized regression model: The maximum rank correlation estimator." *Journal of Econometrics*, 35, 303–316. [126]

Han, S. and E. Vytlacil (2017), "Identification in a generalization of bivariate probit models with dummy endogenous regressors." *Journal of Econometrics*, 199, 63–73. [124]

Hansen, L. P., J. Heaton, and E. G. J. Luttmer (1995), "Econometric evaluation of asset pricing models." *The Review of Financial Studies*, 8, 237–274. [107]

Hansen, L. P. and R. Jagannathan (1991), "Implications of security market data for models of dynamic economies." *Journal of Political Economy*, 99, 225–262. [106]

Heckman, J. J. (1978), "Dummy endogenous variables in a simultaneous equation system." *Econometrica*, 46, 931–959. [124]

Henry, M., R. Méango, and M. Queyranne (2015), "Combinatorial approach to inference in partially identified incomplete structural models." *Quantitative Economics*, 6, 499–529. [107]

Hong, H. and E. Tamer (2003), "Endogenous binary choice model with median restrictions." *Economics Letters*, 80, 219–225. [126]

Honoré, B. E. and A. Lleras-Muney (2006), "Bounds in competing risks models and the war on cancer." *Econometrica*, 74, 1675–1698. [107]

Honoré, B. E. and E. Tamer (2006), "Bounds on parameters in panel dynamic discrete choice models." *Econometrica*, 74, 611–629. [107]

Horowitz, J. L. (2009), *Semiparametric and nonparametric methods in econometrics*. Springer. [108, 125]

Jun, S. J., J. Pinkse, and H. Xu (2011), "Tighter bounds in triangular systems." *Journal of Econometrics*, 161, 122–128. [128]

Jun, S. J., J. Pinkse, and H. Xu (2012), "Discrete endogenous variables in weakly separable models." *The Econometrics Journal*, 15, 288–303. [128]

Karlan, D. S. and J. Zinman (2008), "Credit elasticities in less-developed economies: Implications for microfinance." *American Economic Review*, 98, 1040–1068. [106]

Kitamura, Y. and J. Stoye (2013), "Nonparametric Analysis of Random Utility Models: Testing." cemmap working paper CWP36/13. [107]

Klein, R. W. and R. H. Spady (1993), "An efficient semiparametric estimator for binary response models." *Econometrica*, 61, 387–421. [126]

Komarova, T. (2013), "Binary choice models with discrete regressors: Identification and misspecification." *Journal of Econometrics*, 177, 14–33. [108, 125]

Lafférs, L. (2013), "Identification in models with discrete variables." NHH Dept. of Economics Discussion Paper No. 01/2013. [107]

Lafférs, L. (2015), "Bounding average treatment effects using linear programming." Working paper. [107]

Lehmann, E. L. (1966), "Some concepts of dependence." *The Annals of Mathematical Statistics*, 37, 1137–1153. [118]

Manski, C. (1994), "The selection problem." In *Advances in Econometrics, Sixth World Congress*, Vol. 1, 143–170. [106]

Manski, C. F. (1975), "Maximum score estimation of the stochastic utility model of choice." *Journal of Econometrics*, 3, 205–228. [105, 106, 107, 108, 125, 126]

Manski, C. F. (1985), "Semiparametric analysis of discrete response: Asymptotic properties of the maximum score estimator." *Journal of Econometrics*, 27, 313–333. [105, 107, 108, 125, 126]

Manski, C. F. (1988), "Identification of binary response models." *Journal of the American Statistical Association*, 83, 729–738. [105, 107, 108, 115, 125, 126]

Manski, C. F. (1989), "Anatomy of the selection problem." *The Journal of Human Resources*, 24, 343–360. [106]

Manski, C. F. (2007), "Partial indentification of counterfactual choice probabilities." *International Economic Review*, 48, 1393–1410. [107]

Manski, C. F. (2014), "Identification of income-leisure preferences and evaluation of income tax policy." *Quantitative Economics*, 5, 145–174. [107]

Masten, M. A. and A. Poirier (2016), "Partial Independence in Nonseparable Models." cemmap working paper CWP26/16. [115]

Matzkin, C. F. (2003), "Nonparametric estimation of nonadditive random functions." *Econometrica*, 71, 1339–1375. [114]

Matzkin, C. F. (2007), "Chapter 73 Nonparametric identification." In *Handbook of Econometrics*, Vol. Volume 6, Part 2 (J. J. Heckman and E. E. Leamer, eds.), 5307–5368, Elsevier. [114]

Matzkin, R. L. (1992), "Nonparametric and distribution-free estimation of the binary threshold crossing and the binary choice models." *Econometrica*, 60, 239–270. [126]

Mogstad, M., A. Santos, and A. Torgovitsky (2017), "Using instrumental variables for inference about policy relevant treatment parameters." NBER Working Paper. DOI: 10.3982/ECTA15463. [107]

Molinari, F. (2008), "Partial identification of probability distributions with misclassified data." *Journal of Econometrics*, 144, 81–117. [107]

Mourifié, I. (2015), "Sharp bounds on treatment effects in a binary triangular system." *Journal of Econometrics*, 187, 74–81. [113, 128]

Neal, D. (1997), "The effects of catholic secondary schooling on educational achievement." *Journal of Labor Economics*, 15, 98–123. [106]

Nelsen, R. (2006), *An Introduction to Copulas*. Springer, New York. [111, 133, 136, 137]

Rudin, W. (1976), *Principles of Mathematical Analysis*. McGraw-Hill, New York. [140]

Schennach, S. M. (2014), "Entropic latent variable integration via simulation." *Econometrica*, 82, 345–385. [107]

Schweizer, B. and A. Sklar (1983), *Probabilistic Metric Spaces*. North-Holland, New York. [111, 112, 133, 134, 136]

Shaikh, A. M. and E. J. Vytlacil (2011), "Partial identification in triangular systems of equations with binary dependent variables." *Econometrica*, 79, 949–955. [108, 113, 128]

Shorack, G. R. (2000), *Probability for Statisticians*. Springer, New York. [113]

Sklar, A. (1959), "Fonctions de répartition à n dimensions et leurs marges." *Publ. Inst. Statist. Univ. Paris*, 8, 1. [106, 132, 133]

Sklar, A. (1996), "Random variables, distribution functions, and copulas: A personal look backward and forward." *Institute of Mathematical Statistics Lecture Notes-Monograph Series*, 28, 1–14. [106, 133]

Tamer, E. (2003), "Incomplete simultaneous discrete response model with multiple equilibria." *The Review of Economic Studies*, 70, 147–165. [117]

Thornton, R. L. (2008), "The demand for, and impact of, learning HIV status." *American Economic Review*, 98, 1829–1863. [106]

Torgovitsky, A. (2015), "Identification of nonseparable models using instruments with small support." *Econometrica*, 83, 1185–1197. [114]

Torgovitsky, A. (2016), "Nonparametric inference on state dependence with applications to employment dynamics." SSRN Working Paper 2564305. [107]

Torgovitsky, A. (2017), "Minimum distance from independence estimation of nonseparable instrumental variables models." *Journal of Econometrics*, 199, 35–48. [114]

Torgovitsky, A. (2019), "Supplement to 'Partial identification by extending subdistributions'." *Quantitative Economics Supplemental Material*, 10, https://doi.org/10.3982/QE634. [121]

Vytlacil, E. and N. Yıldız (2007), "Dummy endogenous variables in weakly separable models." *Econometrica*, 75, 757–779. [128]

Wooldridge, J. M. (2010), *Econometric Analysis of Cross Section and Panel Data*. MIT Press. [111]

Yıldız, N. (2013), "Estimation of binary choice models with linear index and dummy endogenous variables." *Econometric Theory*, 29, 354–392. [128]

Co-editor Rosa L. Matzkin handled this manuscript.

Manuscript received 22 October, 2015; final version accepted 6 April, 2018; available online 1 May, 2018.