Partial Identification by Extending Subdistributions

Alexander Torgovitsky*

July 7, 2015

Abstract

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 parameter of interest can be a structural function, but it can also be a more complicated feature of the model primitives, such as an average treatment effect. 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 that this characterization is sharp, I generalize a classic lemma in copula theory concerning the extendibility of subcopulas to show that related objects, termed subdistributions, can be extended to proper distribution functions. This result is then used to reduce the characterization of the identified set to the determination of the existence or non-existence of suitably-constrained subdistributions, which in turn is often equivalent to solving a linear system of equations. I describe this argument as partial identification by extending subdistributions, or PIES. I apply PIES to an ordered discrete response model and a two-sector Roy model. One product of the first application is a tractable characterization of the sharp identified set for the average treatment effect in the semiparametric binary response model considered by Manski (1975, 1985, 1988).

JEL classification: C14; C20; C51

Keywords: partial identification, nonlinear models, copulas, linear programming, discrete choice models, Roy models, semiparametrics, endogeneity

*Department of Economics, Northwestern University, a-torgovitsky@northwestern.edu. Chuck Manski provided helpful comments.

1 Introduction

A central and enduring obstacle in partial identification analysis is the difficulty of obtaining tractable, sharp characterizations of identified sets. Early work on partial identification approached this problem with a two-step argument. First, bounds for a given parameter are proposed. Second, the bounds are shown to be sharp by establishing the existence of an admissible choice of model parameters under which these bounds are achieved. While this type of argument has provided many useful results for nonparametric models (e.g., Manski (2003)), it is 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 commonly used fully parametric econometric models. An example of such a model is an ordered discrete choice model with a linear index function, but with no parametric distributional assumptions on the unobservable term.

This paper provides a new general procedure for constructing tractable, sharp characterizations of identified sets in semiparametric econometric models such as these. The method is based on a classic 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*—i.e. any function with the shape properties of an L-dimensional distribution function, but defined only on a subset of $\overline{\mathbb{R}}^L$—can be extended to a proper distribution function defined on the entirety of $\overline{\mathbb{R}}^L$.[1] The main contribution of this paper to develop an argument that shows 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.

PIES is intuitive and straightforward to describe. Suppose that an econometric model can be written as

$$Y = h(X, U), \tag{1}$$

where Y is a vector of outcome variables, X is a vector of explanatory variables, U is an L-dimensional vector of latent variables with conditional distribution $U|X = x$ given by

[1] Here I am using the standard notation for the extended real number system, $\overline{\mathbb{R}} \equiv \mathbb{R} \cup \{-\infty, +\infty\}$, and its L-fold Cartesian product $\overline{\mathbb{R}}^L$.

$F(\cdot|x)$, and h is the *structural function*. The researcher maintains certain assumptions on h and F, with choices of h and F that satisfy these assumptions described as being *admissible*. Fix an admissible choice of the structural function h. Then, by definition, h is in the identified set if and only if there exists an admissible F that generates the observed distribution of $Y|X = x$ through (1) when the structural function is h and $U|X = x$ is distributed according to $F(\cdot|x)$ for all supported x.

This requirement of *observational equivalence* constrains $F(\cdot|x)$ on a subset $\mathcal{U}_x(h)$ of $\overline{\mathbb{R}}^L$. Often, $\mathcal{U}_x(h)$ is a *finite* subset of $\overline{\mathbb{R}}^L$. The restriction of any $F(\cdot|x)$ to $\mathcal{U}_x(h)$ is what I refer to as a subdistribution, i.e. a function with the shape properties of a distribution function, but with a domain that may be a strict subset of $\overline{\mathbb{R}}^L$. Conversely, suppose that for each supported x there exists a subdistribution $\overline{F}(\cdot|x)$ defined on $\mathcal{U}_x(h)$ that satisfies the observational equivalence condition, and which could be the restriction of $F(\cdot|x)$ to $\mathcal{U}_x(h)$ for some admissible F. Then the subdistribution extension result developed in this paper implies that $\overline{F}(\cdot|x)$ can be extended to an $F(\cdot|x)$ that produces the observed distribution of $Y|X = x$ for each supported x, and that F is admissible. Hence, a given admissible choice of the structural function h is in the identified set *if and only if* for each supported x there exists a subdistribution $\overline{F}(\cdot|x)$ with domain $\mathcal{U}_x(h)$ that is constrained to satisfy the observational equivalence and admissibility conditions. The existence of such subdistributions therefore provides a sharp characterization of the identified set of h.

An important practical aspect of this argument is that the shape restrictions that determine whether a function is a subdistribution are always linear constraints on the values of that function. In addition, as I illustrate throughout the course of the paper, observational equivalence conditions tend to be linear (for a given h), and the restrictions associated with many commonly-used admissibility assumptions are also linear. In such cases, PIES involves determining the existence of a solution to a system of linear equations for each value of h.[2] This is a tractable computational problem as long as the linear system of equations is finite. The finiteness of the linear system depends primarily on the cardinality of the supports of Y and X. If Y and/or X are continuously distributed, so that the linear system is infinite, PIES can still be used to construct arbitrarily accurate outer sets by attempting to solve an arbitrarily large subset of the linear system. As a practical matter, the linear system is always finite for a given sample of data, since the empirical distribution of (Y, X) is necessarily discrete.

[2]As another benefit of PIES, one generally does not need to exhaust the entire space of admissible structural functions to trace out the identified set. This is because knowledge that h is or is not in the identified set also provides knowledge that any other h' with the same "ordering" is in the identified set—see Proposition 2.3.

A second important feature of PIES is that it can be modified to characterize sharp identified sets for parameters that depend on both h and F. This is especially important for nonlinear models, such as discrete choice models, in which causal parameters like the average treatment effect necessarily depend not only on the structural function h, but also on the distribution F of the unobservables.[3] Treating F as a nuisance parameter to be eliminated or avoided is not appropriate for these types of models.

I apply PIES to two widely used econometric models. The first is a semiparametric ordered discrete response model with a linear index function. The explanatory variable of interest in this model can be either exogenous or endogenous, and there can exist an instrumental variable that may or may not be related to the endogenous variable through a first stage equation. No parametric distributional assumptions need to be made about the latent terms in order to apply PIES. A special case of this model is the semiparametric binary response model with exogenous regressors considered by Manski (1975, 1985, 1988). The econometric analysis of such models is sometimes criticized on the grounds that it only identifies the index coefficients, and not causal parameters like the average treatment effect.[4] Using PIES, I show how to construct sharp identified sets for the average treatment effect using only Manski's conditional median independence assumption, thereby addressing this criticism without imposing stronger assumptions. I also show how PIES can be used to incorporate the identifying content of additional distributional shape restrictions, such as symmetry, in an effort to reduce the size of the identified set for the average treatment effect.

The second model is a two-sector Roy model of the sort considered by Heckman and Honoré (1990), Heckman and Vytlacil (2005), and Eisenhauer et al. (2015), among many others. A traditional approach to analyzing such models involves assuming a linear form for the structural function and placing parametric distributional assumptions (typically, normality or log-normality) on the latent variables. Heckman and Honoré (1990) argued that these parametric distributional assumptions are difficult to justify, and showed that they can be replaced by exogenous variables with extreme variation ("large support" in the current parlance). However, the existence of large support variables in actual empirical applications can be dubious. A partial identification approach recognizes the potential failure of large support, but a tractable, sharp characterization of the identified set in a semiparametric Roy model has not (to the best of my

[3]This point is emphasized in modern textbook treatments of nonlinear models, e.g. Wooldridge (2010).

[4]For example, Angrist and Pischke (2009, pg. 201) write "...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." See also Angrist (2001).

knowledge) been provided to date.[5] PIES provides a solution to this problem and, in particular, allows a researcher to construct sharp identified sets for causal parameters by solving linear programs.

The remainder of the paper is structured as follows. I begin in Section 2 by developing PIES for the ordered discrete response model. This analysis demonstrates the key ideas of PIES in a concrete, important and relatively straightforward model. It also shows how to construct identified sets for several types of empirically relevant semiparametric ordered discrete response models. I provide simulation evidence that shows that even under standard data generating processes, the identified sets for index coefficients in these models can be bizarrely shaped, with disconnected regions and other unusual non-convexities. In Section 3, I introduce the formal definition of a subdistribution, prove the subdistribution extension lemma, discuss the related concepts of extendibility and reducibility, and apply PIES to a general econometric model. I also show that the PIES characterization can be rephrased as a conditional moment equality model, thereby enabling the application of recent results on inference in such models, such as Andrews and Shi (2013). In Section 4, I show how PIES can be applied to the two-sector Roy model. Section 5 contains some concluding comments, including a brief and somewhat speculative comparison of the benefits and drawbacks of PIES relative to other general methods for characterizing identified sets based on random set theory (Beresteanu et al. (2011), Galichon and Henry (2011), Chesher and Rosen (2014a,b), Chesher et al. (2013), Aradillas-Lopez and Rosen (2014)) and entropy rankings (Schennach, 2014).

2 PIES for Ordered Discrete Response Models

Consider the ordered discrete response model

$$Y = \sum_{j=1}^{J} y_j \mathbb{1}[g_{j-1}(X) < U \leq g_j(X)], \tag{2}$$

where Y is an observed discrete random variable with support $\mathcal{Y} \equiv \{y_1, \ldots, y_J\}$ that is increasing in j, X is an observed discrete random variable with support $\mathcal{X} \equiv \{x_1, \ldots, x_K\}$, $g \equiv (g_0, g_1, \ldots, g_J)$ is a vector-valued function, and U is a scalar latent random variable. Let $\mathcal{F}$ denote the set of all conditional distribution functions

[5] Mourifié et al. (2015) have recently derived analytic expressions for identified sets in fully nonparametric two-sector Roy models. The analysis in this paper applies to both nonparametric Roy models and semiparametric Roy models with parametrically-specified structural functions, but provides a computational approach rather than analytic expressions.

$F : \overline{\mathbb{R}} \times \mathcal{X} \to [0,1]$. The researcher maintains some non-data identifying assumptions on g and F. For g, these assumptions might include linearity, exclusion restrictions, or certain normalizations, while for F they might represent location normalizations on U and/or the assumption that U and X are independent or median independent. Let $\mathcal{G}^\dagger$ and $\mathcal{F}^\dagger$ respectively denote the set of all functions g and all $F \in \mathcal{F}$ that satisfy these assumptions.

To begin the identification analysis, first consider the trivial case in which $\mathcal{F}^\dagger = \mathcal{F}$, i.e. suppose that the set of admissible conditional distribution functions for $U|X$ is unrestricted. This case is trivial because without any restrictions on F we should expect completely uninformative identified sets. However, the intuition of the PIES argument is clearest when beginning with this case. Consider any admissible $g \in \mathcal{G}^\dagger$. Then, using standard definitions of identification, g is an element of the identified set, $\mathcal{G}^\star$, if and only if there exists a conditional distribution function $F : \overline{\mathbb{R}} \times \mathcal{X} \to [0,1]$ such that when $U|X = x_k$ is distributed like $F(\cdot|x_k)$,

$$\mathbb{P}[Y \leq y_j | X = x_k] = \mathbb{P}[U \leq g_j(x_k) | X = x_k] = F(g_j(x_k)|x_k), \tag{OEQ}$$

for all j and k. This existence question is a complicated one, since conditional distribution functions are *infinite*-dimensional objects. Hence, while (OEQ) provides a sharp characterization of $\mathcal{G}^\star$, it is not clear how one could use this characterization, since it is not clear how one could verify or falsify the existence of an infinite-dimensional F that satisfies (OEQ).

However, closer inspection reveals that determining the existence of an entire conditional distribution F is more difficult than necessary. Observe that for a fixed g, (OEQ) only constrains a candidate conditional distribution function F at the *finite* number of points determined by $g_j(x_k)$ at different combinations of j and k. Hence, to establish that a given admissible g is in the identified set, it suffices to show that a function defined on this finite set of points can be *extended* to a proper conditional distribution function defined on the entirety of $\overline{\mathbb{R}}$. The infinite-dimensional problem can therefore be turned into a finite-dimensional problem, as long as a suitable extension argument can be provided. The necessity of the existence of a solution to the finite-dimensional problem is trivial, since if $g \in \mathcal{G}^\star$, then (OEQ) is satisfied for an infinite-dimensional object F, and hence is also satisfied by the restriction of F to a finite domain that includes the points of evaluation in (OEQ).

For the current model, such an extension argument is straightforward, due to the fact that U is scalar and that $\mathcal{F}^\dagger = \mathcal{F}$ is unrestricted. In particular, suppose that $\overline{F}(\cdot|x_k)$ is a function with domain $\{g_j(x_k)\}_{j=0}^J$ and range contained in $[0,1]$, where

$g \in \mathcal{G}^\dagger$ is arbitrary and fixed. Intuitively, it should be clear that as long as $\overline{F}(\cdot|x_k)$ is weakly increasing on this finite domain, then it can be extended by interpolation to a proper distribution function $F(\cdot|x_k)$ defined on the entirety of $\overline{\mathbb{R}}$.[6] As an extension of $\overline{F}(\cdot|x_k)$, this distribution function $F(\cdot|x_k)$ satisfies $F(g_j(x_k)|x_k) = \overline{F}(g_j(x_k)|x_k)$ for every j. If for each k, $\overline{F}(\cdot|x_k)$ can be chosen to be not only weakly increasing, but also to satisfy (OEQ) in the sense that

$$\mathbb{P}[Y \leq y_j | X = x_k] = \overline{F}(g_j(x_k)|x_k) \quad \forall j, \tag{3}$$

then the extensions $\{F(\cdot|x_k) : k \in 1, \ldots, K\}$ together comprise a conditional distribution function $F \in \mathcal{F}$ that satisfies (OEQ), which implies that g is in the identified set $\mathcal{G}^\star$. This observation is formalized in the following proposition.

Proposition 2.1. *Fix an admissible $g \in \mathcal{G}^\dagger$. For each $k = 1, \ldots, K$, let*

$$\mathcal{U}_k(g) \equiv \{g_j(x_k) : j = 0, 1, \ldots, J\} \cup \{-\infty, +\infty\}. \tag{4}$$

Then $g \in \mathcal{G}^\star$ if and only if for each $k = 1, \ldots, K$ there exists a real-valued function $\overline{F}(\cdot|x_k)$ with domain $\mathcal{U}_k(g)$ such that for every $k = 1, \ldots, K$

$$\overline{F}(g_j(x_k)|x_k) = \mathbb{P}[Y \leq y_j | X = x_k] \text{ for all } j = 0, 1, \ldots, J, \tag{P1.1}$$

$$1 \geq \overline{F}(u|x_k) \geq 0 \text{ for all } u \in \mathcal{U}_k(g), \tag{P1.2}$$

$$\overline{F}(-\infty|x_k) = 0, \tag{P1.3}$$

$$\overline{F}(+\infty|x_k) = 1, \tag{P1.4}$$

$$\overline{F}(u'|x_k) \geq \overline{F}(u|x_k) \text{ for all } u, u' \in \mathcal{U}_k(g) \text{ such that } u' \geq u. \tag{P1.5}$$

Proof of Proposition 2.1. If $g \in \mathcal{G}^\star$ then there exists a conditional distribution function $F : \overline{\mathbb{R}} \times \mathcal{X} \to [0, 1]$ that satisfies (OEQ) for all j and k. Letting $\overline{F}(u|x_k) = F(u|x_k)$ for each $k = 1, \ldots, K$ and all $u \in \mathcal{U}_k(g)$, conditions (P1.1)–(P1.5) follow immediately from (OEQ) and the properties of conditional distribution functions.

Conversely, suppose that $g \in \mathcal{G}^\dagger$ and that for each $k = 1, \ldots, K$ there exists an $\overline{F}(\cdot|x_k)$ with domain $\mathcal{U}_k(g)$ that satisfies (P1.1)–(P1.5). By (P1.1), $\overline{F}(\cdot|x_k)$ satisfies (OEQ) for all j and k. By (P1.2)–(P1.5) and Lemma 3.2, each $\overline{F}(\cdot|x_k)$ can be extended to a proper conditional distribution function $F(\cdot|x_k)$ that also satisfies (OEQ). Hence $g \in \mathcal{G}^\star$. *Q.E.D.*

[6] This one-dimensional interpolation argument is also shown formally as part of Lemma 3.2 ahead. A graphical depiction of the argument is provided in Figure 10.

Proposition 2.1 is an example of what is described in this paper as partial identification by extending subdistributions (PIES). A function $\overline{F}(\cdot|x_k)$ that satisfies (P1.2)–(P1.5) is called a subdistribution because it satisfies all of the properties of a distribution function, except that its domain $\mathcal{U}_k(g)$ is a proper subset of $\overline{\mathbb{R}}$.[7] The subdistribution extension lemma, discussed in Section 3.1, shows that for any subdistribution there exists a proper distribution that extends it, in the sense that the two functions are equal on the smaller domain of the subdistribution. Limiting attention to subdistributions is useful for characterizing identified sets because, given a $g \in \mathcal{G}^\dagger$, the observational equivalence condition (OEQ) only restricts a proper conditional distribution function at a finite number of points. As a result, the domain $\mathcal{U}_k(g)$ of the relevant subdistribution $\overline{F}(\cdot|x_k)$ is finite, at least in models of discrete response. Proposition 2.1 shows that if there exist subdistributions that satisfy (OEQ), then there also exist proper distributions that satisfy (OEQ) and, conversely, that if there exist proper distributions that satisfy (OEQ), then their restrictions to smaller domains are subdistributions that satisfy (OEQ). Hence, the existence of subdistributions satisfying (OEQ) is equivalent to the existence of distributions satisfying (OEQ), which is in turn sufficient and necessary for an admissible g to be in the identified set. This is why Proposition 2.1 provides a sharp characterization of the identified set $\mathcal{G}^\star$.

The real power of PIES comes in the models discussed ahead, in which there are more than one unobservable. The intuition of Proposition 2.1 carries over to these models, but the result that any subdistribution can be extended to a proper distribution is more subtle. Sklar (1959) established an analogous extension result for subcopulas in the course of proving his celebrated eponymous theorem.[8] His argument is based on multilinear interpolation. Lemma 3.2 builds on Sklar's result to provide a companion extension result for subdistributions. This is discussed more fully in Section 3.1.

A crucial practical aspect of Proposition 2.1 is that the restrictions on the functions $\overline{F}(\cdot|x_k)$ are linear for a fixed g. Hence, for a given $g \in \mathcal{G}^\dagger$, Proposition 2.1 shows that determining whether g is in the identified set is simply a matter of determining the existence of a solution to a system of linear equations. This is a relatively straightforward computational problem. This linearity property of Proposition 2.1 will be shared by all of the low-level characterizations discussed in this paper, and is the primary reason that PIES is methodologically attractive.[9]

[7] This terminology is not fully standard but makes sense given the standard and closely related concept of a subcopula. These definitions are discussed formally in Section 3.1 and Appendix A.

[8] Sklar's result is reported in Appendix A.

[9] However, linearity is not the defining property of PIES, as shown in the general development in Section 3.2.

Continuing with the analysis of the ordered discrete response model, note that up to this point no assumptions have been placed on the set of admissible functions $g \in \mathcal{G}^\dagger$ or the set of admissible conditional distribution functions $F \in \mathcal{F}^\dagger$. As a consequence, the identified set $\mathcal{G}^\star$ is completely uninformative. I show next that the characterization in Proposition 2.1 can be modified to impose various additional identifying restrictions on g and/or F, thereby leading to informative identified sets.

Restrictions on g are easy to impose by simply defining the admissible set $\mathcal{G}^\dagger$ to only contain functions with those restrictions. For example, if $J = 2$ so that Y is a binary outcome, then one might assume that $\mathcal{G}^\dagger = \{g = (g_0, g_1, g_2) : g_0(x) = -\infty, g_2(x) = +\infty, g_1(x) = \beta_0 + \beta_1 x, \text{ some } (\beta_0, \beta_1) \in \mathbb{R}^2\}$ to impose a standard linear index form. In this case, fixing a $g \in \mathcal{G}^\dagger$ would be equivalent to fixing a $(\beta_0, \beta_1) \in \mathbb{R}^2$. Imposing restrictions on F requires more subtlety. By way of explanation, consider the following proposition, which shows how independence, median independence, and conditional symmetry conditions can be incorporated into the characterization of Proposition 2.1.

Proposition 2.2. *Fix an admissible $g \in \mathcal{G}^\dagger$.*

1. *Suppose that $\mathcal{F}^\dagger = \mathcal{F}^{ind}$, where $\mathcal{F}^{ind}$ is the set of all $F \in \mathcal{F}$ such that $F(u|x_k) = F(u|x_{k'})$ for all $k, k' \in \{1, \ldots, K\}$ and all $u \in \overline{\mathbb{R}}$. For each $k = 1, \ldots, K$, let $\mathcal{U}_k(g)$ be defined as*

$$\mathcal{U}_k(g) \equiv \mathcal{U}(g) = \bigcup_{k'=1}^{K} \{g_j(x_{k'}) : j = 0, 1, \ldots, J\} \cup \{-\infty, +\infty\}. \tag{U1.I}$$

Then $g \in \mathcal{G}^\star$ if and only if for each $k = 1, \ldots, K$ there exists a function $\overline{F}(\cdot|x_k)$ with domain $\mathcal{U}_k(g)$ that satisfies (P1.1)–(P1.5)*, and if these functions can be taken so that*

$$\overline{F}(u|x_k) = \overline{F}(u|x_{k'}) \text{ for all } k, k' \text{ and } u \in \mathcal{U}(g). \tag{P1.I}$$

2. *Suppose that $\mathcal{F}^\dagger = \mathcal{F}^{med}$, where $\mathcal{F}^{med}$ is the set of all $F \in \mathcal{F}$ such that $F(0|x_k) = \frac{1}{2}$ for all $k = 1, \ldots, K$. For each $k = 1, \ldots, K$, let*

$$\mathcal{U}_k(g) \equiv \{g_j(x_k) : j = 0, 1, \ldots, J\} \cup \{0\} \cup \{-\infty, +\infty\}. \tag{U1.M}$$

Then $g \in \mathcal{G}^\star$ if and only if for each $k = 1, \ldots, K$ there exists a function $\overline{F}(\cdot|x_k)$ with domain $\mathcal{U}_k(g)$ that satisfies (P1.1)–(P1.5)*, and also satisfies*

$$\overline{F}(0|x_k) = \frac{1}{2} \text{ for every } k = 1, \ldots, K. \tag{P1.M}$$

3. *Suppose that* $\mathcal{F}^\dagger = \mathcal{F}^{sym}$*, where* $\mathcal{F}^{sym}$ *is the set of all* $F \in \mathcal{F}$ *for which* $U|X = x_k$ *would be distributed symmetrically around* 0*, i.e. those* F *such that* $F(u|x_k) = 1 - F(-u|x_k)$ *for all* $u \in \mathbb{R}$*. For each* $k = 1, \ldots, K$*, let*

$$\mathcal{U}_k(g) \equiv \{\pm g_j(x_k) : j = 0, 1, \ldots, J\} \cup \{0\} \cup \{-\infty, +\infty\}. \tag{U1.S}$$

Then $g \in \mathcal{G}^\star$ *if and only if for each* $k = 1, \ldots, K$ *there exists a function* $\overline{F}(\cdot|x_k)$ *with domain* $\mathcal{U}_k(g)$ *that satisfies* (P1.1)–(P1.5)*, and also satisfies*

$$\overline{F}(u|x_k) = 1 - \overline{F}(-u|x_k) \text{ for every } u \in \mathcal{U}_k(g). \tag{P1.S}$$

Proof of Proposition 2.2. Suppose that $g \in \mathcal{G}^\star$. Then for the first, second and third cases there exist, respectively, an $F \in \mathcal{F}^{\text{med}}$, an $F \in \mathcal{F}^{\text{ind}}$, and an $F \in \mathcal{F}^{\text{sym}}$ for which (OEQ) holds. In all of these cases, the result follows by taking $\overline{F}(u|x_k) = F(u|x_k)$ and observing the definitions of $\mathcal{F}^{\text{med}}$, $\mathcal{F}^{\text{ind}}$ or $\mathcal{F}^{\text{sym}}$.

Conversely, suppose that $g \in \mathcal{G}^\dagger$ and that the first set of conditions hold for some functions $\overline{F}(\cdot|x_k)$ with domains $\mathcal{U}(g)$ as defined in (U1.I). Let $\overline{\overline{F}} : \mathcal{U}(g) \to [0,1]$ be defined by $\overline{\overline{F}}(u) = \overline{F}(u|\overline{x})$ for any arbitrary $\overline{x} \in \mathcal{X}$. Then given (P1.2)–(P1.5), Lemma 3.2 shows that $\overline{\overline{F}}$ can be extended to a proper distribution function $\widetilde{F} : \overline{\mathbb{R}} \to [0,1]$ such that $\widetilde{F}(u) = \overline{\overline{F}}(u)$ for all $u \in \mathcal{U}(g)$. Let $F : \overline{\mathbb{R}} \times \mathcal{X} \to [0,1]$ be defined by $F(u|x_k) = \widetilde{F}(u)$ for all k, so that $F \in \mathcal{F}^{\text{ind}}$. From (P1.I) it follows that $F(u|x_k) = \overline{F}(u|\overline{x}) = \overline{F}(\cdot|x_k)$ for all k. As a result of this and (P1.1), F also satisfies (OEQ) for each k. Hence $g \in \mathcal{G}^\star$.

Suppose instead that $g \in \mathcal{G}^\dagger$ and that the second set of conditions hold for some functions $\overline{F}(\cdot|x_k)$ with domains $\mathcal{U}_k(g)$ as defined as in (U1.M). The same extension argument as in Proposition 2.1 shows that there exists an $F \in \mathcal{F}$ such that $F(u|x_k) = \overline{F}(u|x_k)$ for all $u \in \mathcal{U}_k(g)$ and each $k = 1, \ldots, K$, so that F satisfies (OEQ). In addition, since $0 \in \mathcal{U}_k(g)$ and $\overline{F}$ satisfies (P1.M), $F(0|x_k) = \overline{F}(0|x_k) = \frac{1}{2}$ for each $k = 1, \ldots, K$, so that $F \in \mathcal{F}^{\text{med}}$. Hence $g \in \mathcal{G}^\star$.

The proof of necessity for the third set of conditions follows the same essential argument as the second set of conditions, but uses a more involved construction discussed in Example 3.4 ahead. *Q.E.D.*

The first and second parts of Proposition 2.2 provide sharp characterizations of $\mathcal{G}^\star$ under, respectively, the assumption that U is independent of X, and the assumption that U has median 0, conditional on X. The third part strengthens the assumption of the second part to also impose symmetry on the distribution of U, conditional on X. Notice that in all three cases, the characterization of $\mathcal{G}^\star$ still amounts to determining the existence of solutions to systems of linear equations, since each of (P1.I), (P1.M)

and (P1.S) are linear in $\{\overline{F}(u|x_k) : u \in \mathcal{U}_k(g), k = 1, \ldots, K\}$. It is straightforward to combine the first and second assumptions so that U is not only independent of X, but also has the median 0 location restriction. To do this, start with the first statement of the proposition, but modify the definition of $\mathcal{U}_k(g) \equiv \mathcal{U}(g)$ in (U1.I) so that the set also includes 0, as in (U1.M), and then impose both (P1.I) and (P1.M). An immediate modification of the proof shows that, for a fixed $g \in \mathcal{G}^\dagger$, the existence of $\overline{F}(\cdot|x_k)$ satisfying these constraints together with (P1.1)–(P1.5) both implies and is implied by $g \in \mathcal{G}^\star$. A similar argument could be used to combine the first and third assumptions. The general model developed in Section 3.2 provides a formal framework for combining assumptions in this way.

At this point it may be helpful to illustrate Propositions 2.1 and 2.2 through a numerical illustration. Consider the binary response model

$$Y_1 = \mathbb{1}[\beta_0 + \beta_1 X_1 + \beta_2 X_2 < U_1], \tag{5}$$

with two included explanatory variables, X_1 and X_2. This model is a special case of the ordered discrete response model with $J = 2$, $y_1 = 0$, $y_2 = 1$, and

$$\begin{aligned}\mathcal{G}^\dagger = \{g = (g_0, g_1, g_2) : g_0(x) = -\infty, g_2(x) = +\infty,& \\ g_1(x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2, \text{ some } (\beta_0, \beta_1, \beta_2) \in \mathbb{R}^3\}.&\end{aligned} \tag{6}$$

The explanatory variable X_1 is binary and generated as

$$X_1 = \mathbb{1}[\pi_0 + \pi_2 X_2 + \pi_3 X_3 < U_2], \tag{7}$$

where X_3 is a third explanatory variable that is excluded from g. The data was generated with parameters values $(\beta_0^\star, \beta_1^\star, \beta_2^\star) = (.5, -.75, 1)$, $(\pi_0^\star, \pi_2^\star, \pi_3^\star) = (.3, 1, .2)$. The distribution of X_2 and X_3 were taken to be independent with marginal distributions uniform over one of the following supports

$$\begin{aligned}\mathcal{X}_{2,K_2=3} &= \{-1, 0, 1\} & \mathcal{X}_{3,K_3=3} &= \{-1, 0, 1\} \\ \mathcal{X}_{2,K_2=5} &= \left\{-1, -\frac{1}{2}, 0, \frac{1}{2}, 1\right\} & \mathcal{X}_{3,K_2=5} &= \{-2, -1, 0, 1, 2\} \\ \mathcal{X}_{2,K_2=7} &= \left\{-1, -\frac{2}{3}, -\frac{1}{3}, 0, \frac{1}{3}, \frac{2}{3}, 1\right\}\end{aligned} \tag{8}$$

where the integers K_2 and K_3 indicate the number of support points. The distribution of (U_1, U_2) was taken to be bivariate normal independent of (X_2, X_3), with independent components that each have mean 0 and variance 1.

It is well-known that informative identification in binary response models is only possible up to scale. In these simulations I imposed a scale normalization by assuming that $\beta_2 = 1$. Since this normalized value matches $\beta_2^\star$, the identified sets that will be constructed can be interpreted as those for the levels of (β_0, β_1). For cases in which $\beta_2^\star$ has a value different than 1, the constructed identified sets can be interpreted as those for $(\beta_0/|\beta_2^\star|, \beta_1/|\beta_2^\star|)$, i.e. they would be identified sets for (β_0, β_1) up to scale.[10]

The first row of Figure 1 displays the identified set for (β_0, β_1) under the assumption that $\mathcal{F}^\dagger = \mathcal{F}^{\text{med}}$, i.e. that the median of $U_1|X_1, X_2$ is 0.[11] If it were further assumed that U_1 is normally distributed and fully independent of (X_1, X_2), then these identified sets would collapse to the point $(\beta_0^\star, \beta_1^\star)$ indicated on the plots by a hash mark, as in the textbook probit model. The effect of symmetry and independence versus normality can be seen in the third row of Figure 1 which displays the identified set for (β_0, β_1) under the stronger restriction that $\mathcal{F}^\dagger = \mathcal{F}^{\text{med}} \cap \mathcal{F}^{\text{ind}} \cap \mathcal{F}^{\text{sym}}$.

The assumption that $\mathcal{F}^\dagger = \mathcal{F}^{\text{med}}$ was also employed by Manski (1975, 1985, 1988) to establish point identification in semiparametric binary response models without maintaining a parametric distributional assumption on U_1. However, Manski's conditions for point identification rely crucially on at least one component of (X_1, X_2) having a continuous distribution. In general, point identification is not possible when (X_1, X_2) are discretely distributed as in this simulation (Manski, 1988). The columns in the first row of Figure 1 show that the size of the identified set decreases with the size of the support of X_2. This intuitive property can be seen from Proposition 2.2, which shows that additional points of support add equations to a linear system, thereby imposing more stringent restrictions for a given (β_0, β_1) pair to be in the identified set.

Proposition 2.2 provides a new way to compute the sharp identified set for β in semiparametric binary response models, but not the first; Horowitz (2009, pp. 100-108) suggested a computational approach and Komarova (2013) developed Horowitz's approach into a more analytic argument.[12] However, as I show below, an extension of Proposition 2.2 also provides a method for computing the identified set of causal parameters such as the average treatment effect under the maximum score assumptions. No method for doing this has (to the best of my knowledge) been proposed before. In addition, the PIES methodology can also be applied to models with two (or more) equations. As demonstrated ahead, PIES can therefore be used to construct identi-

[10] These issues are exactly the same as in other parametric and semiparametric binary response models; I am just restating them here for clarity.

[11] All simulations in this paper were written in AMPL (Fourer et al., 2002) and the linear programs were solved with CPLEX (IBM, 2010).

[12] See also Blevins (2015) who considers set estimation in this framework.

fied sets that exploit variation in instrumental variables, such as X_3, with or without maintaining the assumption that X_1 is determined through the relationship given in (7).

Each panel in Figure 1 was computed by searching across a grid of candidate $\beta = (\beta_0, \beta_1)$ pairs, determining the existence or nonexistence of a solution to the linear systems posed in Proposition 2.2, and then correspondingly classifying β as either inside or outside of the identified set. This is more computationally intensive than necessary. In fact, if g and g' induce the same orderings of the elements of $\mathcal{U}_k(g)$ and $\mathcal{U}_k(g')$ for every k, then both or neither of g and g' are in $\mathcal{G}^\star$. This reduces the computational problem from one of searching over a grid to one of considering all possible orderings, which in many cases will dramatically decrease the computational intensity.[13] The next proposition formalizes this argument.

Proposition 2.3. *Fix any $g, g' \in \mathcal{G}^\dagger$. Maintain the assumptions of any part of Proposition 2.2 and let $\mathcal{U}_k(g)$ and $\mathcal{U}_k(g')$ be defined accordingly. Consider the mappings*

$$\psi_k : \mathcal{U}_k(g') \to \mathcal{U}_k(g) : \psi_k(u) \equiv \begin{cases} g_j(x) & \text{if } u = g'_j(x) \text{ for some } j = 0, 1, \ldots, J \\ u & \text{if otherwise and } u \in \mathcal{U}_k(g) \cap \mathcal{U}_k(g'). \end{cases} \tag{9}$$

Suppose that each ψ_k is well-defined through (9), and that each ψ_k is order-preserving, in the sense that $u' \leq u''$ for $u', u'' \in \mathcal{U}_k(g')$ if and only if $\psi_k(u') \leq \psi_k(u'')$. Then $g \in \mathcal{G}^\star$ if and only if $g' \in \mathcal{G}^\star$.

Proof of Proposition 2.3. Suppose that ψ_k as defined by (9) is well-defined and order-preserving for each k, and that $g \in \mathcal{G}^\star$. For concreteness, suppose that $\mathcal{F} = \mathcal{F}^{\text{med}}$ as in the second part of Proposition 2.2 and that $\mathcal{U}_k(g)$ is defined as in (U1.M). Then for each k there exists an $\overline{F}(\cdot|x_k) : \mathcal{U}_k(g) \to [0,1]$ such that conditions (P1.1)–(P1.5) and (P1.M) are satisfied. Define the functions $\overline{F}'(\cdot|x_k) : \mathcal{U}_k(g') \to [0,1]$ by $\overline{F}'(u|x_k) = \overline{F}(\psi_k(u)|x_k)$. Notice that each $\overline{F}'(\cdot|x_k)$ is well-defined, since each ψ_k is well-defined with codomain $\mathcal{U}_k(g)$. By the definition of ψ_k, each $\overline{F}'(\cdot|x_k)$ satisfies (P1.1)–(P1.4) (with $\mathcal{U}_k(g)$ replaced by $\mathcal{U}_k(g')$) and (P1.M), since $\overline{F}(\cdot|x_k)$ satisfies these conditions. Since each ψ_k is order-preserving, each $\overline{F}'(\cdot|x_k)$ also satisfies (P1.5) (again, with $\mathcal{U}_k(g)$ replaced by $\mathcal{U}_k(g')$). Given these properties of each $\overline{F}'(\cdot|x_k)$, Proposition 2.2 implies that $g' \in \mathcal{G}^\star$. The converse follows by exchanging the roles of g and g'. *Q.E.D.*

Consider the linear specification of $\mathcal{G}^\dagger$ in (6) with β_2 normalized to 1. Under the independence condition in the first part of Proposition 2.2, with $\mathcal{U}_k(g)$ specified as (4),

[13] This ordering property appears to be related to similar findings by Chesher (2010, 2013), Chiburis (2010), and Chesher and Smolinski (2012).

the mapping ψ_k is order-preserving if and only if $g(x) = \beta_0 + \beta_1 x_1 + x_2$ and $g'(x) = \beta_0' + \beta_1' x_1 + x_2$ are such that β_1 and β_1' have the same sign. If the median zero condition in the second part of Proposition 2.2 is also imposed, so that $0 \in \mathcal{U}_k(g) \cap \mathcal{U}_k(g')$, then ψ_k is order-preserving if and only if β_1 and β_1' have the same sign and the set of supported (x_1, x_2) for which $\beta_0 + \beta_1 x_1 + x_2 \leq 0$ are the same as those for which $\beta_0' + \beta_1' x_1 + x_2 \leq 0$. In this example, independence by itself ensures identification of at most the sign of β_1, while some information on (β_0, β_1) can potentially be obtained with the additional median 0 restriction.

The goal of many empirical studies in economics is to ascribe a causal interpretation to the threshold-crossing model (2). In such analyses, the object of interest is not only g but a causal parameter, such as an average treatment effect, that depends on both g and F. For example, in the threshold-crossing model (2), a researcher may be interested in the average structural function at x_k. Suppose that Y is binary with all $g \in \mathcal{G}^\dagger$ having $g_0(x) = -\infty$ and $g_2(x) = +\infty$. Then the average structural function at a fixed value $x^\star$ (not necessarily in the support of X) when the data is generated by (g, F) is defined as

$$\text{ASF}(g, F) \equiv \mathbb{P}[U > g_1(x^\star)] = 1 - \sum_{k=1}^{K} F(g_1(x^\star)|x_k)\, \mathbb{P}[X = x_k], \tag{10}$$

see e.g. Blundell and Powell (2003). Let $\mathcal{P}^\star$ denote the identified set for the average structural function at $x^\star$, i.e. the set of all $p \in [0, 1]$ for which there exists a $g \in \mathcal{G}^\dagger$ and an $F \in \mathcal{F}^\dagger$ that satisfy both (OEQ) and $p = \text{ASF}(g, F)$. The next proposition shows how one can modify Propositions 2.2 and 2.3 to compute $\mathcal{P}^\star$ by solving linear programs.

Proposition 2.4. *Fix an admissible $g \in \mathcal{G}^\dagger$. Suppose that $\mathcal{F}^\dagger$ is as in one of the parts of Proposition 2.2. For each $k = 1, \ldots, K$, let $\mathcal{U}_k(g)$ be defined as in Proposition 2.2 according to the choice of $\mathcal{F}^\dagger$, but also include the point $g_1(x^\star)$ in each $\mathcal{U}_k(g)$. Define*

$$\underline{p}^\star(g) \equiv \min_{\substack{\overline{F}(\cdot|x_k):\mathcal{U}_k(g) \to [0,1] \\ k=1,\ldots,K}} 1 - \sum_{k=1}^{K} \overline{F}(g_1(x^\star)|x_k)\, \mathbb{P}[X = x_k] \tag{11}$$

$$\textit{subject to } (\text{P1.1})\text{–}(\text{P1.5}) \textit{ and } (\text{P1.I}), (\text{P1.M}), \textit{ or } (\text{P1.S}),$$

and define $\overline{p}^\star(g)$ to be the analogous maximum.[14] *Then $\mathcal{P}^\star \equiv \bigcup_{g \in \mathcal{G}^\star} [\underline{p}^\star(g), \overline{p}^\star(g)]$.*

Moreover, $[\underline{p}^\star(g), \overline{p}^\star(g)] = [\underline{p}^\star(g'), \overline{p}^\star(g')]$ for any $g, g' \in \mathcal{G}^\star$ for which the mappings

[14] If the constraint set is empty, then follow the standard convention of setting $\underline{p}^\star(g) = +\infty$, $\overline{p}^\star(g) = -\infty$ and $[\underline{p}^\star(g), \overline{p}^\star(g)] = \emptyset$.

ψ_k in Proposition 2.3 are well-defined and order-preserving.

Proof of Proposition 2.4. The result follows trivially if $\mathcal{G}^\star$ is the empty set. So assume that $\mathcal{G}^\star$ is not empty.

Suppose that $p \in \mathcal{P}^\star$. By definition, there exists a $g' \in \mathcal{G}^\star$ and an $F \in \mathcal{F}^\dagger$ that satisfy (OEQ) and $p = \text{ASF}(g', F)$. Taking $\overline{F}(u|x_k) = F(u|x_k)$ for each k and all $u \in \mathcal{U}_k(g')$ as in Propositions 2.1 and 2.2, shows that the constraint set in the program defining $\underline{p}^\star(g')$ and $\overline{p}^\star(g')$ is non-empty. Hence, since $p = 1 - \sum_{k=1}^K \overline{F}(g_1'(x^\star)|x_k)\,\mathbb{P}[X = x_k]$, it follows that $p \in [\underline{p}^\star(g'), \overline{p}^\star(g')]$, and hence that $p \in \bigcup_{g \in \mathcal{G}^\star} [\underline{p}^\star(g), \overline{p}^\star(g)]$.

Conversely, suppose that there exists a $g \in \mathcal{G}^\star$ for which $p \in [\underline{p}^\star(g), \overline{p}^\star(g)]$. By Proposition 2.2, there exist functions $\overline{F}(\cdot|x_k) : \mathcal{U}_k(g) \to [0,1]$ satisfying the constraints in (11).[15] As noted previously, these constraints are all linear in the values of $\overline{F}(\cdot|x_k)$, so the set of all such $\{\overline{F}(u|x_k) : u \in \mathcal{U}_k(g), k = 1, \ldots, K\}$ is closed and convex. Since the objective function in (11) is continuous (in fact, linear) in the values of $\overline{F}(u|x_k)$, the image of the objective function over the constraint set is the closed interval $[\underline{p}^\star(g), \overline{p}^\star(g)]$. Hence, since $p \in [\underline{p}^\star(g), \overline{p}^\star(g)]$, there exists some choice $\{\overline{\overline{F}}(u|x_k) : u \in \mathcal{U}_k(g), k = 1, \ldots, K\}$ in the constraint set for which $p = 1 - \sum_{k=1}^K \overline{\overline{F}}(g_1(x^\star)|x_k)\,\mathbb{P}[X = x_k]$. The result follows as in the previous two propositions by extending $\overline{\overline{F}}$ to a proper conditional distribution function F.

For the second statement, suppose that $g, g' \in \mathcal{G}^\star$ and that $p \in [\underline{p}^\star(g), \overline{p}^\star(g)]$. Then there exist $\{\overline{F}(u|x_k) : u \in \mathcal{U}_k(g), k = 1, \ldots, K\}$ in the constraint set of (11) for which $p = 1 - \sum_{k=1}^K \overline{F}(g_1(x^\star)|x_k)\,\mathbb{P}[X = x_k]$. Define $\overline{F}'(\cdot|x_k) : \mathcal{U}_k(g') \to [0,1]$ by $\overline{F}'(u|x_k) = \overline{F}(\psi_k(u)|x_k)$, as in the proof of Proposition 2.3. As shown in that proof, each $\overline{F}'(\cdot|x_k)$ also satisfies the constraints of (11) (with $\mathcal{U}_k(g')$ replacing $\mathcal{U}_k(g)$). Moreover, given the definition of ψ_k,

$$\begin{aligned} 1 - \sum_{k=1}^K \overline{F}'\left(g_1'(x^\star)|x_k\right)\mathbb{P}[X = x_k] &= 1 - \sum_{k=1}^K \overline{F}\left(\psi_k(g_1'(x^\star))|x_k\right)\mathbb{P}[X = x_k] \\ &= 1 - \sum_{k=1}^K \overline{F}\left(g_1(x^\star)|x_k\right)\mathbb{P}[X = x_k] = p, \end{aligned}$$

so that $p \in [\underline{p}^\star(g'), \overline{p}^\star(g')]$. The opposite inclusion follows by exchanging the roles of g and g'. *Q.E.D.*

Proposition 2.4 justifies the following computational procedure for determining sharp identified sets for causal parameters like the average structural function. First, fix a $g \in \mathcal{G}^\dagger$. Next, attempt to solve the minimizing linear program in (11). There

[15] Note that adding the point $g_1(x^\star)$ to $\mathcal{U}_k(g)$ for each k does not affect this observation.

exists a feasible solution to this program if and only if $g \in \mathcal{G}^\star$. If in fact $g \in \mathcal{G}^\star$, then solve the maximizing program. The resulting interval $[\underline{p}^\star(g), \overline{p}^\star(g)]$ represents the values of the average structural function that are consistent with $g \in \mathcal{G}^\star$. Repeat this procedure over the entire space $\mathcal{G}^\dagger$ to determine $\mathcal{G}^\star$ and $\mathcal{P}^\star = \cup_{g \in \mathcal{G}^\star}[\underline{p}^\star(g), \overline{p}^\star(g)]$. To be more efficient computationally, one can partition $\mathcal{G}^\dagger$ into subsets such that every g and g' in a given subset induces a mapping ψ_k in Proposition 2.3 that is well-defined and order-preserving. Then, the minimizing and maximizing linear programs need only be solved once for a representative g from each subset. Chiburis (2010) used a similar computational procedure; Proposition 2.4 provides a theoretical justification for this procedure and, more importantly, suggests that it applies to a much more general class of models than considered by Chiburis (2010). In Section 3, I develop this class of models and also describe a wider choice of functions of (g, F) to which a similar procedure can be applied to determine the identified set. In particular, it is straightforward to compute identified sets for an average treatment effect (the difference of two average structural functions) by simply adjusting the objective function in (11).

Table 2 reports sharp identified sets for the average treatment effect of X_1 on Y_1 for the data generating process defined by (5), (6), (7), and the surrounding discussion. Specifically, the parameter of interest here is $\text{ATE} \equiv \text{ASF}_1 - \text{ASF}_0$, where

$$\text{ASF}_0 \equiv \mathbb{P}[U_1 > \beta_0 + \beta_2 X_2] = 1 - \sum_{k=1}^{K_2} F(\beta_0 + \beta_2 x_{2,k} | x_{2,k}) \, \mathbb{P}[X_2 = x_{2,k}], \text{ and}$$

$$\text{ASF}_1 \equiv \mathbb{P}[U_1 > \beta_0 + \beta_1 + \beta_2 X_2] = 1 - \sum_{k=1}^{K_2} F(\beta_0 + \beta_1 + \beta_2 x_{2,k} | x_{2,k}) \, \mathbb{P}[X_2 = x_{2,k}].$$

The row corresponding to specification [1] reports the sharp identified set for the ATE under the assumption that the conditional median of U_1 is 0. There does not appear to be any other method in the literature for obtaining sharp identified sets of this quantity.[16] Proposition 2.4 addresses the criticism of (e.g.) Angrist and Pischke (2009, pg. 201) that the focus of binary response models should be average treatment effects rather than index coefficients. The PIES method makes it possible to address identifi-

[16] It is important to reiterate here that the ATE involves *counterfactual* choice probabilities that have no direct analog in the observed data. Khan (2013) proves that semiparametric binary response models with the conditional median 0 assumption are observationally equivalent to parametric response models with latent terms that have parametrically-specified unconditional distributions with multiplicative heteroskedasticity of an unknown form. He uses this result to construct estimators of the *factual* choice probabilities ($\mathbb{P}[Y = 1|X = x]$) for conditional median 0 binary response models without making parametric assumptions on the distribution of the latent term. An implication of Khan's (2013) result that is relevant for the current discussion is that the counterfactual choice probabilities are not point identified under only a conditional median 0 assumption.

cation of both the index coefficients and average treatment effects, without resorting to a parametric distributional assumption on the error term U_1. The fact that the identified set is not a singleton represents the cost of being agnostic on the shape of the distribution of U_1. The assertion made by Wooldridge (2010, pg. 606) that "... we can only learn about relative sizes of the coefficients using the semiparametric approach..." turns out to be overly pessimistic. Informative bounds on causal parameters can also be obtained using the PIES methodology.

In specification [2], the conditional median 0 assumption is strengthened into full symmetry around 0 of the conditional distribution of U_1. Manski (1988) reasoned that this has no effect on the identified set for the index coefficients, an assertion which is supported by the first and second rows of Figure 1. However, comparing rows [1] and [2] in Table 1 shows that the symmetry assumption *does* have an effect on the size of the identified set for the ATE. Under the full independence condition in specification [3], the identified set for the ATE collapses to a singleton, as expected. Nevertheless, the index coefficients are still partially identified under full independence. A practical consequence is that using a binary response model for extrapolation to covariate values not observed in a given data set would still lead to partial identification, even under the strong condition of full independence.[17]

The analysis thus far has been concerned with the single equation model given by (2). The sharp characterizations of identified sets given in Propositions 2.2 and 2.4 utilized the intuitive result that a one-dimensional subdistribution function can be extended to a proper one-dimensional distribution function with domain $\overline{\mathbb{R}}$. Similar but less intuitive results also hold in higher dimensions. Some additional definitions are needed for the general case. These are discussed in full detail in Section 3.1. For now, I simply state and use the following result for two dimensions, which requires no new definitions.[18]

Lemma 2.1. ***(Subdistribution extension in two dimensions)*** *Suppose that $\mathcal{U} \subseteq \overline{\mathbb{R}}^2$ can be written as $\mathcal{U} = \mathcal{U}_1 \times \mathcal{U}_2$ where each of $\mathcal{U}_1$ and $\mathcal{U}_2$ is a closed subset of $\overline{\mathbb{R}}$ that contains $\{-\infty, +\infty\}$. Let $\overline{F} : \mathcal{U} \to [0,1]$ be a function for which (i) $\overline{F}(u_1, u_2) = 0$ if either u_1 or u_2 is equal to $-\infty$, (ii) $\overline{F}(+\infty, +\infty) = 1$, and (iii)*

$$\overline{F}(u_1', u_2') - \overline{F}(u_1'', u_2') - \overline{F}(u_1', u_2'') + \overline{F}(u_1'', u_2'') \geq 0 \tag{12}$$

[17] Such an extrapolation exercise would correspond to setting $x^\star$ to a point not in $\mathcal{X}$ in Proposition 2.4. Extrapolation like this is often of interest in discrete choice analysis, for example in forecasting demand for a new or hypothetical product on which no data is observed.

[18] The proof of Lemma 2.1 is a special case of Lemma 3.2 and Corollary 3.1 below.

for any $u', u'' \in \mathcal{U}$ such that $u'' \leq u'$ component-wise. Then there exists a proper 2-dimensional joint distribution function F with domain $\overline{\mathbb{R}}^2$ such that $F(u) = \overline{F}(u)$ for all $u \in \mathcal{U}$.[19]

Suppose additionally that the margins of $\overline{F}$, i.e.

$$\begin{aligned} \overline{F}_1 : \mathcal{U}_1 \to [0,1] : \overline{F}_1(u_1) = \overline{F}(u_1, +\infty) \\ \text{and } \overline{F}_2 : \mathcal{U}_2 \to [0,1] : \overline{F}_2(u_2) = \overline{F}(+\infty, u_2), \end{aligned}$$

are such that for each $j = 1, 2$ there exists an F_j in some class of proper one-dimensional distribution functions $\mathcal{F}_j$ such that $F_j(u_j) = \overline{F}_j(u_j)$ for all $u_j \in \mathcal{U}_j$. Then there exists a proper, 2-dimensional joint distribution function F with domain $\overline{\mathbb{R}}^2$ such that $F(u) = \overline{F}(u)$ for all $u \in \mathcal{U}$, and such that $F_j \in \mathcal{F}_j$ for $j = 1, 2$, where

$$\begin{aligned} F_1 : \mathcal{U}_1 \to [0,1] : F_1(u_1) = F(u_1, +\infty) \\ \text{and } F_2 : \mathcal{U}_2 \to [0,1] : F_2(u_2) = F(+\infty, u_2) \end{aligned}$$

are the margins of F.

To see how Lemma 2.1 can be used to characterize identified sets for models with multiple unobservable variables, consider adding a first stage equation to (2), i.e. assume that

$$Y_1 = \sum_{j=1}^{J} y_{1j} \mathbb{1}[g_{1(j-1)}(Y_2, X) < U_1 \leq g_{1j}(Y_2, X)] \tag{13}$$

$$\text{and } \; Y_2 = \sum_{k=1}^{K} y_{2k} \mathbb{1}[g_{2(k-1)}(X) < U_2 \leq g_{2k}(X)], \tag{14}$$

where Y_1 and Y_2 are observed variables with supports $\mathcal{Y}_1 \equiv \{y_{11}, \ldots, y_{1J}\}$ and $\mathcal{Y}_2 \equiv \{y_{21}, \ldots, y_{2K}\}$ ordered to be increasing,

$$g \equiv (g_1, g_2) \equiv (g_{10}, g_{11}, \ldots, g_{1J}, g_{20}, g_{21}, \ldots, g_{2K})$$

is an unknown vector of functions, X is an observed variable with finite support $\mathcal{X}$, and $U \equiv (U_1, U_2)$ is a bivariate latent random variable. Relative to the single equation model, X has been replaced notationally by Y_2, an equation (14) for Y_2 has been added to (13), and a new variable X—the determination of which is not modeled—has been introduced. This change in notation is made in order to be consistent with

[19] The technical qualifier that F is non-defective can also be added; this is addressed in the formal discussion in Section 3.1.

the general formulation in Section 3, where a distinction is made between internal (modeled) variables Y and external (unmodeled) variables X.

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 subset of $\mathcal{F}$ deemed admissible by the researcher. In this two-equation model, $g \in \mathcal{G}^\star$ if and only if $g \in \mathcal{G}^\dagger$ and there exists an $F \in \mathcal{F}^\dagger$ such that when $U|X = x$ is distributed like $F(\cdot|x)$,

$$\begin{aligned}
&\mathbb{P}[Y_1 \leq y_{1j}, Y_2 \leq y_{2k}|X = x] \\
&\quad = \sum_{k'=1}^{k} \mathbb{P}\left[U_1 \leq g_{1j}(y_{2k'}, x), Y_2 = y_{2k'}|X = x\right] \\
&\quad = \sum_{k'=1}^{k} \mathbb{P}\left[U_1 \leq g_{1j}(y_{2k'}, x), U_2 \in (g_{2(k'-1)}(x), g_{2k'}(x)]|X = x\right] \\
&\quad = \sum_{k'=1}^{k} F\left(g_{1j}(y_{2k'}, x), g_{2k'}(x)|x\right) - F\left(g_{1j}(y_{2k'}, x), g_{2(k'-1)}(x)|x\right) \qquad \text{(OEQ2)}
\end{aligned}$$

for all $j = 1, \ldots, J$, $k = 1, \ldots, K$ and $x \in \mathcal{X}$.

The following proposition shows how to apply Lemma 2.1 to characterize $\mathcal{G}^\star$ under the assumption that X is an exogenous instrument in the sense that $(U_1, U_2) \perp\!\!\!\perp X$. It also provides a characterization of $\mathcal{G}^\star$ under the weaker exogeneity assumption that $U_1 \perp\!\!\!\perp X$, in which case (14) is without loss of generality.[20] The difference between the identified sets in the two cases reflects the identifying content of the first stage equation (14) under the joint independence assumption $(U_1, U_2) \perp\!\!\!\perp X$.[21]

Proposition 2.5. *Fix an admissible $g \in \mathcal{G}^\dagger$.*

1. *Suppose that $F \in \mathcal{F}^{marg}$, where $\mathcal{F}^{marg}$ is the set of all $F \in \mathcal{F}$ such that $F(u_1, +\infty|x) = F(u_1, +\infty|x')$ for all $x, x' \in \mathcal{X}$ and all $u_1 \in \overline{\mathbb{R}}$. For each $x \in \mathcal{X}$, let*

$$\begin{aligned}
&\mathcal{U}_{1,x}(g) = \mathcal{U}_1(g) \equiv \{g_{1j}(y_{2k}, x) : j = 0, 1, \ldots, J, k = 1, \ldots, K, x \in \mathcal{X}\} \cup \{-\infty, +\infty\}, \\
&\quad \text{and } \mathcal{U}_{2,x}(g) \equiv \{g_{2k}(x) : k = 0, 1, \ldots, K\} \cup \{-\infty, +\infty\},
\end{aligned}$$

and let $\mathcal{U}_x(g) \equiv \mathcal{U}_1(g) \times \mathcal{U}_{2,x}(g)$. Then $g \in \mathcal{G}^\star$ if and only if for each $x \in \mathcal{X}$ there

[20] To see this, note that if U_2 is uniformly distributed on $[0,1]$ and g is taken so that $g_{2,0}(x) = 0$, $g_{2,K}(x) = 1$ and $g_{2,k}(x) = \mathbb{P}[Y_2 \leq y_{2,k}|X = x]$, then (14) always generates the distribution of $Y_2|X$. In fact, (14) is still without loss of generality if both $U_1 \perp\!\!\!\perp X$ and $U_2 \perp\!\!\!\perp X$. The content of (14) comes from the assumption that (U_1, U_2) are *jointly* independent of X, which is not implied by marginal independence between X and each of U_1 and U_2.

[21] In both cases, it is common to also impose the exclusion restriction that g_1 not depend directly on X. This can be incorporated by appropriately defining $\mathcal{G}^\dagger$, as illustrated in the simulations below. Proposition 2.5 is valid whether or not one imposes this type of exclusion restriction.

exists a function $\overline{F}(\cdot|x)$ *with domain* $\mathcal{U}_x(g)$ *such that*

$$\sum_{k'=1}^{k} \overline{F}(g_{1j}(y_{2k'}, x), g_{2k'}(x)|x) - \overline{F}(g_{1j}(y_{2k'}, x), g_{2(k'-1)}(x)|x) \tag{P2.1}$$
$$= \mathbb{P}[Y_1 \leq y_{1j}, Y_2 \leq y_{2k}|X = x] \text{ for all } j = 1, \ldots, J \text{ and } k = 1, \ldots, K,$$

$$1 \geq \overline{F}(u|x) \geq 0 \text{ for all } u \in \mathcal{U}_x(g), \tag{P2.2}$$

$$\overline{F}(u_1, -\infty|x) = 0 \text{ and } \overline{F}(-\infty, u_2|x) = 0 \text{ for all } u_1 \in \mathcal{U}_1(g),\ u_2 \in \mathcal{U}_{2,x}(g), \tag{P2.3}$$

$$\overline{F}(+\infty, +\infty|x) = 1, \tag{P2.4}$$

$$\overline{F}(u_1', u_2'|x) - \overline{F}(u_1', u_2''|x) - \overline{F}(u_1'', u_2'|x) + \overline{F}(u_1'', u_2''|x) \geq 0$$
$$\text{for all } u', u'' \in \mathcal{U}_x(g) \text{ such that } u'' \leq u' \text{ component-wise}, \tag{P2.5}$$

together with

$$\overline{F}(u_1, +\infty|x) = \overline{F}(u_1, +\infty|x') \text{ for all } x, x' \in \mathcal{X} \text{ and } u_1 \in \mathcal{U}_1(g). \tag{P2.MI}$$

2. *Suppose that* $F \in \mathcal{F}^{joint}$, *where* $\mathcal{F}^{joint}$ *is the set of all* $F \in \mathcal{F}$ *such that* $F(u|x) = F(u|x')$ *for all* $x, x' \in \mathcal{X}$ *and all* $u \in \overline{\mathbb{R}}^2$. *Let*

$$\mathcal{U}_1(g) \equiv \{g_{1j}(y_{2k}, x) : j = 0, 1, \ldots, J, k = 1, \ldots, K, x \in \mathcal{X}\} \cup \{-\infty, +\infty\},$$
$$\text{and } \mathcal{U}_2(g) \equiv \{g_{2k}(x) : k = 0, 1, \ldots, K, x \in \mathcal{X}\} \cup \{-\infty, +\infty\},$$

and for each $x \in \mathcal{X}$ *let* $\mathcal{U}_x(g) \equiv \mathcal{U}(g) \equiv \mathcal{U}_1(g) \times \mathcal{U}_2(g)$. *Then* $g \in \mathcal{G}^\star$ *if and only if for each* $x \in \mathcal{X}$ *there exists a function* $\overline{F}(\cdot|x)$ *with domain* $\mathcal{U}_x(g) \equiv \mathcal{U}(g)$ *that satisfies* (P2.1)–(P2.5), *together with*

$$\overline{F}(u|x) = \overline{F}(u|x') \text{ for all } x, x' \in \mathcal{X} \text{ and } u \in \mathcal{U}(g). \tag{P2.JI}$$

Proof of Proposition 2.5. The necessity statement follows from analogous arguments to those in Propositions 2.1 and 2.2.

The sufficiency statement is more straightforward to prove in the second case, so suppose first that $g \in \mathcal{G}^\dagger$ and that there exist functions $\overline{F}(\cdot|x)$ with domains $\mathcal{U}_x(g) \equiv \mathcal{U}(g)$ that satisfy (P2.1)–(P2.5), and (P2.JI). Given (P2.2)–(P2.5), Lemma 2.1 shows that for any arbitrary $\overline{x} \in \mathcal{X}$, $\overline{F}(\cdot|\overline{x})$ can be extended to a proper bivariate distribution function $\widetilde{F} : \overline{\mathbb{R}}^2 \to [0, 1]$ for which $\widetilde{F}(u) = \overline{F}(u|\overline{x})$ for all $u \in \mathcal{U}(g)$. Given (P2.1) and (P2.JI), letting $F \in \mathcal{F}^{\text{joint}}$ be defined by $F(u|x) = \widetilde{F}(u) = \overline{F}(u|\overline{x})$ shows that there exists an $F \in \mathcal{F}^{\text{joint}}$ satisfying (OEQ2), and hence that $g \in \mathcal{G}^\star$.

Now suppose that the functions $\overline{F}(\cdot|x)$ satisfy the conditions in the first case. Define

$\overline{\overline{F}}_1 : \mathcal{U}_1(g) \to [0,1]$ by $\overline{\overline{F}}_1(u_1) = \overline{F}(u_1, +\infty|\overline{x})$ for any arbitrary $\overline{x} \in \mathcal{X}$. Then (P2.3) and (P2.5) imply that $\overline{\overline{F}}_1$ is weakly increasing,[22] so there exists a proper one-dimensional distribution function $\widetilde{F}_1$ that extends $\overline{\overline{F}}_1$. Letting $F_1 : \overline{\mathbb{R}} \times \mathcal{X} \to [0,1]$ be defined as $F_1(u_1|x) = \widetilde{F}_1(u_1)$ for all $x \in \mathcal{X}$, it follows that $F_1(u_1|x) = \widetilde{F}_1(u_1) = \overline{\overline{F}}_1(u_1) \equiv \overline{F}(u_1, +\infty|\overline{x})$ for all $x \in \mathcal{X}$, so that $F_1(\cdot|x)$ is an extension of $\overline{F}(\cdot, +\infty|x)$ that also does not vary with x. Hence, (P2.1)–(P2.5) with the second part of Lemma 2.1 implies that there exists an $F \in \mathcal{F}^{\text{marg}}$ satisfying (OEQ2), so that $g \in \mathcal{G}^\star$. *Q.E.D.*

The primary change in Proposition 2.5 from the one-dimensional cases is represented by (P2.5). This condition says that $\overline{F}(\cdot|x)$ is 2-increasing.[23] An additional difference evident in the first part of Proposition 2.5 is that, when dealing with subdistribution functions of higher dimensions, it is possible to impose conditions on the margins separately. Aside from these details, however, the central idea of the PIES argument is the same, in that it shows that the existence of a properly constrained subdistribution is both sufficient and necessary for an admissible $g \in \mathcal{G}^\dagger$ to be in the identified set $\mathcal{G}^\star$. Notice in particular that (P2.1)–(P2.5), (P2.MI) and (P2.JI) are all still linear in $\{\overline{F}(u|x) : u \in \mathcal{U}(g), x \in \mathcal{X}\}$, so that the practical determination of whether $g \in \mathcal{G}^\star$ remains a straightforward computational problem. The adaptation of Propositions 2.3 and 2.4 to the triangular model is immediate. It is also immediate to combine the instrumental variable restrictions in Proposition 2.5 with the location and symmetry restrictions in Proposition 2.2 by simply adding those constraints to the linear system. A unified discussion of this type of general framework is contained in Section 3.2.

The simulated data generating process used above to illustrate the single equation case can also be used to illustrate the triangular case. To stay consistent with notation, I relabel X_1 as Y_2 so that equations (5) and (7) become

$$Y_1 = \mathbb{1}[\beta_0 + \beta_1 Y_2 + \beta_2 X_2 < U_1] \tag{15}$$

$$\text{and} \quad Y_2 = \mathbb{1}[\pi_0 + \pi_2 X_2 + \pi_3 X_3 < U_2]. \tag{16}$$

Aside from this change in notation, the data generating process remains the same as before. This model is essentially the bivariate probit model considered originally by Heckman (1978), and more recently by Han and Vytlacil (2015), but the PIES methodology does not require the researcher to maintain a parametric assumption on the distribution of (U_1, U_2). With access to X_3, the researcher can entertain instrumental variables type assumptions that impose various forms of independence between X_3

[22] See Lemma 3.1 in Appendix A.

[23] See Section 3.1 for the general definition of an L-increasing function.

and U_1 and/or U_2. As shown in Proposition 2.5, these assumptions can either be agnostic about the determination of Y_2, or can explicitly impose the triangular first stage equation (16). This first stage equation can either be parametrically or nonparametrically specified. The simulation results discussed ahead shed light on the identifying content of such assumptions.

Specification [4] listed in Table 1 imposes only the assumption that the median of $U_1|X_2, X_3$ is equal to 0, which is like an instrumental variables version of Manski's (1988) semiparametric binary response model.[24] Hong and Tamer (2003) establish point identification of this model under a support condition on (X_2, X_3) that is similar to the one required in the exogenous regressor case, and in particular does not hold when all components of the instruments are discrete, as in this simulation. The sharp identified set for (β_0, β_1) is displayed in the top row of Figure 2. It is unbounded and much larger than the corresponding set for [1] in Figure 1, which reflects the identifying content of assuming exogeneity of the regressor of interest rather than an instrument. As seen by comparing the columns of the first row of Figure 2, the size of the identified set shrinks with the number of support points of the exogenous variables (X_2, X_3) for the same reason as in the exogenous regressor case. This property persists throughout all of the simulations discussed ahead. Row [4] of Table 2 reports the sharp identified set for the average treatment effect of Y_2 on Y_1 under this median independence instrumental variables assumption using an appropriate modification of Proposition 2.4.

Specification [5] strengthens the median independence condition in [4] to full independence, i.e. $U_1 \perp\!\!\!\perp (X_2, X_3)$. This model is similar to that studied by Chesher (2010, 2013) and Chesher and Smolinski (2012), although the results of those authors only cover the case in which $U_1 \perp\!\!\!\perp X_3$, i.e. they do not show how to exploit exogeneity assumptions for included explanatory variables. It is straightforward to treat both cases using the PIES methodology. Those authors also considered a fully nonparametric specification, under which some functional normalization is needed to achieve non-trivial identified sets. A common choice, maintained by those authors, is that U_1 is uniformly distributed on $[0, 1]$ (Matzkin, 1994, 2003, 2007). Such a restriction is not a normalization if g is parametrically specified, as it is here. On the other hand, Chesher (2010, 2013) and Chesher and Smolinski (2012) established analytic expres-

[24] While this case is not covered explicitly under Proposition 2.5, it should be clear how the argument of the first part of Proposition 2.2 can be combined with Lemma 2.1 to provide the appropriate result. In particular, 0 should be included in $\mathcal{U}_x(g)$ for all x, and $\overline{F}$ should be constrained so that $\overline{F}(0|x_2, x_3) = 0$ for all (x_2, x_3). The repetition of enumerating all of these particular cases motivates the general formulation discussed in Section 3.

sions for the sharp identified set of g, whereas the PIES approach is computational. While analytic expressions are helpful for understanding the structure of the identified set, the sharp constructions used in these papers are subtle to derive and difficult to modify for different choices of assumptions.

Sharp identified sets for (β_0, β_1) under [5] are displayed in the second row of Figure 2. Like Chesher (2010, 2013), Chiburis (2010), and Chesher and Smolinski (2012), I find that the sharp identified set for the structural function (here parameterized by (β_0, β_1) is not only non-convex, but disconnected. As a result, the sharp identified set for the ATE is also disconnected, see Table 2. Specification [6] adds to [5] the condition that $U_1|X_2, X_3$ is symmetrically distributed around 0. This has a small effect on the identified set for the ATE, but, unlike in the exogenous Y_2 case, also greatly reduces the size of the identified set of (β_0, β_1).

Some insight into why symmetry has a much larger effect on the identified sets for (β_0, β_1) than it does on the ATE can be seen in the heat maps plotted in Figures 4–9. These graphs show the sharp bounds on the ATE for fixed values of (β_0, β_1). Darker colors indicate better bounds. Comparing the second and third rows of Figure 6, one can see that even though the sharp identified set for (β_0, β_1) is much smaller in [6] than in [5], points in the small disconnected region, which are the ones that achieve the worst fixed-(β_0, β_1) lower bounds, remain in the sharp identified set in both specifications. This explains why the lower bound does not change between [5] and [6]. On the other hand, Figure 7 shows that the upper bound improves somewhat from [5] to [6] because the additional assumptions reduce the size of the larger disconnected region, which contains the (β_0, β_1) points associated with the poorest upper bounds.

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 that is independent of (X_2, X_3) jointly with U_1. Chesher (2005), Shaikh and Vytlacil (2011), Jun et al. (2011, 2012), and Mourifié (2015) derived analytic expressions for completely nonparametric binary response models with first stage equations.[25] As in the single equation literature, the nonparametric outcome equation allows the distribution of U_1 to be normalized, which greatly reduces the difficulty of finding sharp characterizations of identified sets.[26] In contrast, when g is assumed to satisfy a parametric index structure, such a normalization is no longer without loss of generality and a more flexible approach, such as PIES, is required. This has important

[25] 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.

[26] Even in this nonparametric case, deriving the sharp identified set under limited support restrictions requires a subtle construction; see for example the argument in Mourifié (2015).

implications for empirical work, because the curse of dimensionality often renders a fully nonparametric analysis too imprecise to be useful. The analysis of this section shows how one can characterize the empirical content of discrete response—and other nonlinear models—that maintain the linear-in-parameters index, but relax the more objectionable parametric distributional assumptions.

Comparing the results for specifications [5] and [7] in Figures 2 and 3 shows that the assumption that (X_2, X_3) are exogenous with respect to both the first stage and outcome equations does have an impact on the identified set for both the index coefficients and the ATE. Interestingly, 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, which 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.

3 The General Theory of PIES

3.1 The Subdistribution Extension Lemma

This section develops a formal statement of the subdistribution extension lemma alluded to in the previous section. This requires some definitions and intermediate results.[27] The first key concept is that of an *L-increasing* function.

Definition 3.1. *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,*

$$Vol_F(u', u'') \equiv \sum_{u \in Vrt(u', u'')} \text{sgn}_{(u', u'')}(u) F(u) \geq 0, \tag{17}$$

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

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

*The quantity $Vol_F(u', u'')$ 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 $Vrt(u', u'')$ are the* vertices *of the box.*

[27] This discussion follows treatments by Schweizer and Sklar (1983) and Nelsen (2006).

L-increasingness reduces to the standard definition of weakly increasing for $L = 1$.[28] For $L = 2$, it is condition (12) given in Lemma 2.1. Intuitively, the property of L-increasingness requires a function F to assign a non-negative volume to every L-box with vertices in $\mathcal{U}$. This is a key property of distribution functions, for which volumes become interpretable as probabilities of events.

The following definition of a distribution function is standard.[29] The concept of a subdistribution—which, to the best of my knowledge, is novel to this paper—is a natural counterpart to the concept of a subcopula, which is discussed in Appendix A and used in the proofs in Appendix B.[30]

Definition D. *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 $\{-\infty, +\infty\} \in \mathcal{U}_l$ for each l. An L-dimensional* subdistribution function *is a function F with domain $\mathcal{U}$ such that*

D1. *F is L-increasing.*

D2. *$F(u) = 0$ for any $u \in \mathcal{U}$ that has at least one component equal to $-\infty$.*

D3. *$F(+\infty, \ldots, +\infty) = 1$.*

An L-dimensional distribution function *(or, for emphasis, a* proper *L-dimensional distribution function) is an L-dimensional subdistribution function for which $\mathcal{U} = \overline{\mathbb{R}}^L$. A proper L-dimensional distribution function is* non-defective *if $F(u_n) \to 1$ for any sequence $\{u_n\}_{n=1}^{\infty} \subset \overline{\mathbb{R}}^L$ with all components diverging to $+\infty$ and $F(u_n) \to 0$ for any sequence $\{u_n\}_{n=1}^{\infty} \subset \overline{\mathbb{R}}^L$ with at least one component diverging to $-\infty$.*

Corresponding to every L-dimensional subdistribution function is a collection of L *margins*.

Definition 3.2. *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*

$$F_l : \mathcal{U}_l \to [0, 1] : F_l(u_l) = F(+\infty, \ldots, u_l, \ldots, +\infty). \tag{18}$$

The next lemma records the fact that each margin is itself a one-dimensional subdistribution function.

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

[29] Although, note that the normalization of left- or right-continuity for each margin (see below) 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 e.g. Section 2.2 of Schweizer and Sklar (1983).

[30] 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.

Lemma 3.1. *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 $\mathcal{U}_l$.*

Proof of Lemma 3.1. Conditions D2 and D3 follow immediately from the fact that F is an L-dimensional subdistribution. Condition D1, 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). *Q.E.D.*

With these definitions, the general subdistribution extension lemma can be stated as follows.

Lemma 3.2. ***(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 $\{-\infty, +\infty\}$. Let $\overline{F} : \mathcal{U} \to [0,1]$ be an L-dimensional subdistribution. Then there exists a proper, non-defective, L-dimensional joint distribution function F defined on $\overline{\mathbb{R}}^L$ such that $F(u) = \overline{F}(u)$ for all $u \in \mathcal{U}$. Furthermore, if $\overline{F}$ is continuous on $\mathcal{U}$, then there exists such an F that is continuous on $\overline{\mathbb{R}}^L$.*[31]

A proof of Lemma 3.2 is provided in Appendix B. The proof makes use of some fundamental results in copula theory, in particular, Sklar's Theorem and a key intermediate result (referred to here as Sklar's Lemma) used in the classical proof of Sklar's Theorem.[32] Sklar's Lemma establishes a result analogous to Lemma 3.2 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 3.2 can be viewed as an extension of Sklar's Lemma. The main contribution of this paper is to show how Lemma 3.2 (and its corollary below) can be applied to provide tractable, sharp characterizations of identified sets in a wide variety of econometric models.[33]

The following corollary to Lemma 3.2 provides some additional flexibility in choosing the extension F for a given subdistribution $\overline{F}$ for situations in which it is known

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

[32] Appendix A provides a brief review of the relevant concepts and results in copula theory.

[33] Chiburis (2010, pg. 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 did not appear to realize that the argument could be generalized to apply to the much broader class of 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 same nonparametric binary response model.

that the margins of $\overline{F}$ satisfy certain properties. This is useful in the partial identification analysis because it allows a researcher to maintain assumptions directly on the margins of admissible F. To state the corollary, I employ the following definition.

Definition 3.3. *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 $F|_{\mathcal{U}} = \overline{F}$.*[34]

The corollary can now be stated as follows.

Corollary 3.1. *Suppose that $\mathcal{U}$ and $\overline{F}$ are as in Lemma 3.2 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, non-defective, one-dimensional distribution functions. Then there exists a proper, non-defective, 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 but, as the examples below show, it is typically easier to establish than extendibility.

Definition 3.4. *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 $\mathcal{F}$ is* reducible *to $\overline{\mathcal{F}}$ if $F \in \mathcal{F}$ implies that $F|_{\mathcal{U}} \in \overline{\mathcal{F}}$.*

The following are some concrete examples of extendibility and reducibility that will be useful for the partial identification analysis of the next section. 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 nonlinear parametric econometric models, and is also often imposed as a normalization in nonparametric models (e.g. Matzkin (2003), Chernozhukov and Hansen (2005) and Chesher (2010)).

Example 3.1. *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 $\overline{\mathcal{F}}$.*

[34] 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 $\mathcal{B}$.

The next example observes that subdistributions with compact domains (leaving aside $\{-\infty, +\infty\}$) are extendible to the set of distributions with mean 0, and that the converse reducibility property also holds.

Example 3.2. *Suppose that $\mathcal{F}$ is the set of all proper one-dimensional distribution functions such that $\int u \, dF(u) = 0$. Suppose that $\overline{\mathcal{U}} \subseteq \overline{\mathbb{R}}$ is compact and that $\overline{\mathcal{F}}$ is any collection of subdistributions $\overline{F}$ with domains $\mathcal{U} \equiv \overline{\mathcal{U}} \cup \{-\infty, +\infty\}$ 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 regardless of whether F has mean zero or not, its restriction to $\mathcal{U}$ is still a subdistribution. 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 select an $\mathcal{F}$ 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 of a single equation semiparametric binary response model, such as the one considered in the simulations in Section 2. Manski's intuition is essentially what is contained in Example 3.2. In the next section it will become clear that, because of the extendible/reducible property of mean 0 distribution functions, conditional mean zero assumptions do not, by themselves, have identifying content for standard parameters of interest in any ordered discrete response model.[35] This is because, in the absence of any additional identifying assumptions, the subdistribution of interest for these models always has a bounded domain (leaving aside $\{-\infty, +\infty\}$). As a result, restricting the admissible 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. On the other hand, if $\mathcal{F}$ satisfies a median (or other quantile) restriction, then it is only reducible to subdistribution classes that satisfy a similar condition.

Example 3.3. *Suppose that $\mathcal{F}$ is the set of all proper one-dimensional distribution functions F such that $F(0) = 1/2$. 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 $\overline{F}(0) = 1/2$. Then $\overline{\mathcal{F}}$ is extendible to $\mathcal{F}$ and $\mathcal{F}$ is reducible to $\overline{\mathcal{F}}$.*

Notice in particular that the restriction that $0 \in \mathcal{U}$ and $\overline{F}(0) = 1/2$ for all $\overline{F} \in \overline{\mathcal{F}}$ is required for extendibility. For clearly, if $\overline{F}(0) \neq 1/2$ then any proper conditional distribution function F that extends $\overline{F}$ does not satisfy $F(0) = 1/2$, and so is not an element of $\mathcal{F}$.

[35] However, these types of moment conditions can have content when combined with additional independence and support assumptions, see e.g. Lewbel (2000) or Magnac and Maurin (2007).

The next example considers symmetry restrictions of the sort imposed in the third part of Proposition 2.2. The proof of the third part of that proposition was delayed with the claim that it depended on a more involved construction. That construction is presented in this example. It shows how more complicated shape restrictions, such as symmetry, can be embedded in the concepts of reducibility and extendibility. As the general analysis in the next section will make clear, these concepts and Corollary 3.1 will allow a researcher to characterize sharp identified sets for models with multiple unobservables and shape restrictions on the marginal distribution of each of these unobservables.

Example 3.4. *Suppose that $\mathcal{F}$ is the set of all proper one-dimensional distribution functions that are symmetric around 0, i.e. 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 $\overline{\mathcal{F}}$.*

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

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

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 3.2. Then define $F : \overline{\mathbb{R}} \to [0,1]$ by

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

so that $F(u) = 1 - F(-u)$ for all u by construction, i.e. $F \in \mathcal{F}$. 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

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

Hence $\overline{\mathcal{F}}$ is extendible to $\mathcal{F}$.

3.2 PIES for a General Econometric Model

In this section, I use the subdistribution extension lemma and its corollary to develop a sharp characterization of identified sets in a general model of the distribution of an *internal* random vector Y with support $\mathcal{Y} \subseteq \mathbb{R}^{d_Y}$, conditional on an *external* random vector X with support $\mathcal{X} \subseteq \mathbb{R}^{d_X}$. The model is assumed to be describable by a structure $S = (h, F)$, where $F : \overline{\mathbb{R}}^L \times \mathcal{X} \to [0,1]$ is the conditional-on-X distribution function for a latent vector $U \in \mathbb{R}^L$ with known dimension L, and h is a vector-valued function mapping X and U into Y, i.e. $Y = h(X,U)$. This model is quite general, but does not allow for models that are incomplete, i.e. models for which some values of (x,u) lead to more than one value of Y. The distinction between Y and X is that the determination of Y is modeled, while the determination of X is not.[36]

A structure $S = (h, F)$ generates a conditional distribution of Y through the relationship

$$\mathbb{P}^S[Y \leq y | X = x] = \mathbb{P}^S[h(x,U) \leq y | X = x], \tag{19}$$

where $\mathbb{P}^S$ denotes the probability measure induced by the assumption that $Y = h(X,U)$ with $U|X = x$ having conditional distribution following $F(\cdot|x)$ for each $x \in \mathcal{X}$. For characterizing the identified set, it will be convenient to represent the right-hand side of (19) notationally as

$$\mathbb{P}^S[h(x,U) \leq y | X = x] \equiv \omega_{y|x}(h, F), \tag{20}$$

for some function $\omega_{y|x}$ of h and F.

Remark 3.1. *Consider the single equation model in Section 2. In that model, h is given by* (2)*, i.e.*

$$h(X,U) = \sum_{j=1}^{J} y_j \mathbb{1}[g_{j-1}(X) < U \leq g_j(X)], \tag{2}$$

so that h is fully determined by the vector-valued function g. The mapping $\omega_{y|x}$ is given by the right-hand side of (OEQ)*. Observe that $\omega_{y|x}$ does not depend on the entirety of F and h, but only on the values of $F(u|x_k)$ at $u \in \{g_j(x_k) : j = 0, 1, \ldots, J\}$. Similarly, for the triangular model in Section 2, h is defined through* (13)–(14)*, with*

[36] The purpose of using the terminology "internal" and "external" is to reserve the phrases "endogenous" and "exogenous" for describing the relationship between observed and unobserved variables. The latter usage of these terms seems more common in the recent literature.

$g \equiv (g_1, g_2)$ fully determining h, and $\omega_{y|x}$ is given by the right-hand side of (OEQ2). *In this model too, $\omega_{y|x}$ only depends on F and h through the values of $F(u_1, u_2|x)$ with $u_1 \in \{g_{1j}(y_{2k}, x) : j = 0, 1, \ldots, J, k = 1, \ldots, K\}$ and $u_2 \in \{g_{2k}(x) : k = 0, 1, \ldots, K\}$. The limited dependence of $\omega_{y|x}$ on (h, F) played an important role in the results of Section 2 and will be captured in the general Theorem 1 below.*

The notational distinction between Y and X for the ordered discrete response models analyzed in Section 2 is consistent with that developed here. Notice, however, that there is often some flexibility with whether one classifies a variable as internal or external. For example, under the assumptions imposed in the second part of Proposition 2.2, (14) *was without loss of generality, so Y_2 could have alternatively been classified as an external variable without affecting the model.*

The F component of the structure $S = (h, F)$ is a proper conditional joint distribution function $F : \overline{\mathbb{R}}^L \times \mathcal{X} \to [0, 1]$ for the latent variables $U \in \mathbb{R}^L$, conditional on the external variables X. The dimension of the latent variables, L, is assumed to be known by the researcher, but is otherwise unrestricted in relationship to the dimension of the observed variables. The space of all such proper conditional L-dimensional distribution functions is denoted by $\mathcal{F}$. The researcher only considers structures with F components lying in some admissible set $\mathcal{F}^\dagger \subseteq \mathcal{F}$, which is typically a proper subset of $\mathcal{F}$. The admissible set contains only those distributions which satisfy the a priori (or "identifying") assumptions maintained by the researcher. The next assumption places abstract restrictions on the types of identifying assumptions that can be maintained.

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

A1. *$\rho(h, F) \geq \vec{0}$ for some known function ρ, where $\vec{0}$ is the 0–vector of finite dimension d_ρ, and the inequality is interpreted component-wise.*

A2. *For each $l = 1, \ldots, L$ and $x \in \mathcal{X}$, $F_l(\cdot|x) \in \mathcal{F}_{l,x}^\dagger$ where $\mathcal{F}_{l,x}^\dagger$ is a known collection of proper one-dimensional distribution functions.*

A3. *For all $u \in \overline{\mathbb{R}}^L$, $F(u|x) = F(u|x')$ for all $x, x' \in \mathcal{X}_0^\dagger$, where $\mathcal{X}_0^\dagger$ is a (possibly empty) known subset of $\mathcal{X}$.*

A4. *For each $l = 1, \ldots, L$ and all $u_l \in \overline{\mathbb{R}}$, $F_l(u_l|x) = F_l(u_l|x')$ for all $x, x' \in \mathcal{X}_l^\dagger$, where $\mathcal{X}_l^\dagger$ is a (possibly empty) known subset of $\mathcal{X}$.*

Remark 3.2. *The different forms of A1–A4 provide flexibility in specifying the admissible set. The independence condition in Proposition 2.2 falls under both A3 and*

A4, since $L = 1$ in that case. The median independence condition in Proposition 2.2 could be included as either A1 or A2—see Example 3.3. The symmetry condition in Proposition 2.2 can also be included as A2—see Example 3.4. In Proposition 2.5, the marginal conditional independence restriction would be classified under A4, while the joint conditional independence restriction would be classified under A3. The statement of A3 and A4 allows the global conditional independence in those propositions to be weakened to local (for $x \in \mathcal{X}_0^\dagger$ and/or $\mathcal{X}_l^\dagger$) conditional independence statements in the manner considered by Chesher (2003, 2005, 2007).

An example of an assumption that one might wish to consider 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 is satisfied (by definition) if and only if $F(u_1, u_2) \geq F_1(u_1)F_2(u_2)$ for all $(u_1, u_2) \in \overline{\mathbb{R}}^2$, where I have suppressed covariates simply for notational ease. 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 A1–A4 are general, they are not vacuous.*

The h component of $S = (h, F)$ is a real-valued function (the "structural function") with domain contained in $\mathcal{X} \times \mathbb{R}^L$. As a function, $h(x, \cdot)$ maps every u in its domain to exactly one value of y. However, the domain of $h(x, \cdot)$ could be a proper subset of $\mathbb{R}^L$ so that there are values of u such that $h(x, u)$ does not map to any value of y, i.e. the model can be incoherent in the sense of Chesher and Rosen (2012). To satisfy observational equivalence, F must necessarily place 0 probability on sets of such u. On the other hand, the assumption that h is a well-defined function requires the model to be complete, in the sense that h generates a unique value of y for each (x, u), which rules out models with multiple equilibria. The researcher only considers structures with h components lying in some admissible set denoted by $\mathcal{H}^\dagger$.

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

$$\mathcal{S}^\star \equiv \Big\{ S = (h, F) : F \in \mathcal{F}^\dagger, h \in \mathcal{H}^\dagger \\ \text{and } \omega_{y|x}(h, F) = \mathbb{P}[Y \leq y | X = x] \text{ for all } y \in \mathcal{Y},\, x \in \mathcal{X} \Big\}. \tag{21}$$

For a fixed $h \in \mathcal{H}^\dagger$, let $\mathcal{F}^\star(h) \equiv \{F : (h, F) \in \mathcal{S}^\star\}$ denote the profiled identified set for F, given h. Then the profiled identified set for h can be written as $\mathcal{H}^\star \equiv \{h \in \mathcal{H}^\dagger : \mathcal{F}^\star(h) \text{ is nonempty}\}$. The researcher's object of interest is a function $\pi(h, F)$ of the entire structure, which is assumed to take values in a finite dimensional space $\overline{\mathbb{R}}^{d_\pi}$. A

simple example of π is $\pi(h, F) = h$, if h is finitely parameterized. However, π can also represent causal parameters that depend not only on h but also on F.

Remark 3.3. *For the ordered discrete response model of Section 2, the average structural function at $x^\star$ given in* (10) *is an example of a function π of (h, F) that depends on both h (fully determined by g in this case) and F. Observe that, like $\omega_{y|x}$, ASF only depends on F for a fixed h through the values of F on a subset of its domain, in particular $\{F(g_1(x^\star)|x_k) : k \in 1, \ldots, K\}$. This limited dependence of π on F is important, and will be incorporated into Theorem 1 ahead.*

The identified set associated with a given parameter π is denoted by $\mathcal{P}^\star = \{\pi(h, F) : (h, F) \in \mathcal{S}^\star\}$. The next theorem shows how the subdistribution extension lemma can be applied to characterize $\mathcal{P}^\star$ for the general model developed in this section.

Theorem 1 (**The general method of 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 $h \in \mathcal{H}^\dagger$, let $\{\mathcal{U}_x(h) : x \in \overline{\mathcal{X}}\}$ be any collection of subsets of $\overline{\mathbb{R}}^L$ for which the following statements are true:*

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

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

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

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

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

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 $\mathcal{U}_{l,x}(h)$ 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. *$\mathcal{U}_x(h) = \mathcal{U}_{x'}(h)$ for all $x, x' \in (\overline{\mathcal{X}} \cap \mathcal{X}_0^\dagger)$.*

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

If $p \in \mathcal{P}^\star$ then there exists an $h \in \mathcal{H}^\dagger$ and function $\overline{F}(\cdot|x) : \mathcal{U}_x(h) \to [0, 1]$ for each

$x \in \overline{\mathcal{X}}$ *such that:*

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

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

$$\overline{\rho}\left(h, \left\{\overline{F}(\cdot|x) : x \in \overline{\mathcal{X}}\right\}\right) \geq 0_{d_\rho}, \tag{T1.3}$$

$$\overline{F}_l(\cdot|x) \in \overline{\mathcal{F}}^{\dagger}_{l,x} \text{ for each } l = 1, \ldots, L \text{ and } x \in \overline{\mathcal{X}}. \tag{T1.4}$$

$$\overline{F}(u|x) = \overline{F}(u|x') \text{ for every } x, x' \in (\overline{\mathcal{X}} \cap \mathcal{X}_0^{\dagger}) \text{ and all } u \in \mathcal{U}_x(h). \tag{T1.5}$$

$$\overline{F}_l(u_l|x) = \overline{F}_l(u_l|x') \text{ for every } x, x' \in (\overline{\mathcal{X}} \cap \mathcal{X}_l^{\dagger}) \text{ and all } u_l \in \mathcal{U}_{l,x}(h), \text{ where } \overline{F}_l(\cdot|x) \text{ is the } l\text{th margin of } \overline{F}(\cdot|x). \tag{T1.6}$$

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

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

The proof of Theorem 1, which is given in Appendix C, is an extensive generalization of the proofs of Propositions 2.1, 2.2 and 2.5 in Section 2. The theorem clarifies the role of the subdistribution extension lemma in providing characterizations of the identified set for the parameter π. However, several considerations need to be made for Theorem 1 to achieve the same practicality as the low-level results in Section 2.

The first consideration is the choice of the sets $\{\mathcal{U}_x(h) : x \in \overline{\mathcal{X}}\}$. Observe that conditions U1–U5 are satisfied trivially by taking $\mathcal{U}_x(h) = \overline{\mathbb{R}}^L$ for every $x \in \overline{\mathcal{X}}$ and any $h \in \mathcal{H}^{\dagger}$. However, this choice of $\{\mathcal{U}_x(h) : x \in \overline{\mathcal{X}}\}$ renders Theorem 1 essentially a tautology, and finding functions $\overline{F}(\cdot|x)$ that satisfy (T1.1)–(T1.7) remains as difficult as finding proper distributions satisfying these conditions. On the other hand, if $\mathcal{U}_x(h)$ can be taken to be finite for every $x \in \overline{\mathcal{X}}$ and $h \in \mathcal{H}^{\dagger}$, then Theorem 1 reduces the infinite-dimensional problem to a potentially finite-dimensional one. This was the case in the ordered discrete response models discussed in Section 2, and is also the case for the Roy model discussed in Section 4.

The cardinality of a given $\mathcal{U}_x(h)$ is determined by conditions U2 and U3. In particular, for $\mathcal{U}_x(h)$ 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 (h, F) only through the value of F on a finite subset of its domain. This condition was satisfied in the ordered discrete response model of Section 1, and will also be satisfied in the Roy model discussed in Section 4. However, there are modifications of that analysis for which this condition would not hold. For example, choosing $\pi(h, F) = \mathbb{E}_F(U_1)$ in the ordered discrete response model of Section 2 would not lead to a function $\overline{\pi}$ that depends on F at only a finite number

of points, since the expectation of a random variable depends on its distribution at almost all points of its support. For the same reason, incorporating mean-zero conditions into ρ will not in general lead to a function $\overline{\rho}$ that depends on F through only a finite number of values. It is possible that one could circumvent these drawbacks through an approximation argument, but I leave such a development for future work. These limitations comprise the main weaknesses of the PIES approach. Further discussion of the merits of PIES relative to other methods proposed in the literature is provided in Section 5.

Even if $\omega_{y|x}, \pi$, and ρ have these type of finite representations for a given (y, x) pair, (T1.1) requires consideration of all $y \in \overline{\mathcal{Y}}$ and $x \in \overline{\mathcal{X}}$. Hence, for $\mathcal{U}_x(h)$ to be finite also generally requires $\overline{\mathcal{Y}}$ and $\overline{\mathcal{X}}$ to be finite. This restricts Theorem 1 to only provide practical sharp characterizations for models in which both Y and X are discrete. I do not view this as a serious drawback for two reasons. First, as shown in Theorem 1, PIES still provides a tractable characterization for outer identified sets by taking $\overline{\mathcal{Y}} \subseteq \mathcal{Y}$ and $\overline{\mathcal{X}} \subseteq \mathcal{X}$ to be finite subsets of $\mathcal{Y}$ and $\mathcal{X}$. Hence, with sufficient computing power, these outer identified sets 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 $\mathcal{X}$. Any limitation in this regard is purely computational. Second, the infinite cardinality of $\mathcal{Y}$ and $\mathcal{X}$ is only an issue in the rarefied world of simulations, where it is presumed that the distribution of observables is known without error. In practice, the empirical distribution of the observed data is always discrete, even if Y and X are modeled as continuous random variables with infinite support. It is reasonable to expect that outer sets constructed using all support points of the empirical distribution will be consistent (in an appropriate sense) for the sharp identified set in the population. A formal development of such an argument requires statistical concerns that are important, but beyond the scope of the current paper.

Putting aside these dimensionality issues, another important practical consideration to Theorem 1 is whether the system of equations (T1.1)–(T1.7) can be reliably solved for a given $h \in \mathcal{H}^\dagger$. This was the case in the ordered discrete response models of Section 2, where the corresponding system of equations were linear in $\overline{F}$. Importantly, the crucial subdistribution condition, (T1.2), always places a linear restriction on $\overline{F}$, given a fixed $h \in \mathcal{H}^\dagger$—recall Definition D. The conditional independence restrictions (T1.5) and (T1.6) are also always linear for a fixed $h \in \mathcal{H}^\dagger$. As a result, linearity of the entire system in Theorem 1 requires that for a fixed $h \in \mathcal{H}^\dagger$, each of $\overline{\omega}_{y|x}$, $\overline{\rho}$ and $\overline{\pi}$ are linear in the values of $\overline{F}$. The same must also be true of whatever low-level condition defines T1.4. These properties turns out to be satisfied for all of the models discussed in this paper, although it is possible to think of examples where it is not satisfied. For

instance, imposing the positive quadrant dependence condition mentioned in Remark 3.2 at a finite collection of (u_1, u_2) points would require ρ to be nonlinear. Hence, while the linearity of these functions still allows for many interesting examples, including those discussed in this paper, it need not always be satisfied. Theorem 1 is still valid for cases in which this linearity fails, but it may be prohibitively difficult to implement due to the difficulties involved in reliably solving non-convex programs.

An additional consideration that was highlighted in Proposition 2.4 is whether characterizing the identified set for π can be rephrased as a series of optimization problems. As stated, Theorem 1 would suggest that one needs to determine the existence of non-existence of a solution to (T1.1) for every $h \in \mathcal{H}^\dagger$ and every $p \in \mathcal{P}^\star$. However, if π is scalar-valued and $\overline{\omega}_{y|x}, \overline{\rho}$ and $\overline{\pi}$ are all linear functions of $\overline{F}$ (for a fixed $h \in \mathcal{H}^\dagger$), then one can, as in Proposition 2.4, determine $\mathcal{P}^\star$ by minimizing and maximizing $\overline{\pi}$ subject to (T1.1)–(T1.6) once for each $h \in \mathcal{H}^\dagger$, and then taking the union of all min/max intervals. More generally, this is true if $\overline{\pi}$ is a continuous function of $\overline{F}$ (for a fixed $h \in \mathcal{H}^\dagger$) and if the constraint set represented by (T1.1)–(T1.6) is a closed and convex set, see e.g. Theorem 4.22 of Rudin (1976). The proof of these statements follows the same argument as in Proposition 2.4 and so is omitted.

Searching over every $h \in \mathcal{H}^\dagger$ could be computationally daunting in more complicated models. Proposition 2.3 provided one way of ameliorating this issue by partitioning the space of admissible structural functions into sets for which the domains of $\overline{F}$ were order-invariant, in the sense discussed there. Using this scheme, Proposition 2.3 established that one only needs to consider a single representative function from each order-invariant class. This property of PIES carries over to the more general construction in Theorem 1, but a general statement appears to be inherently cumbersome. Moreover, implementing such a computational scheme is necessarily application-dependent, since it depends on the choice of $\mathcal{H}^\dagger$ and the general composition of the model at hand. Hence, I only remark here that constructions similar to that discussed in Proposition 2.3 can be used to greatly reduce the computational burden of the characterization in Theorem 1, but I leave a more detailed discussion of the appropriate algorithm for specific applications in future work.

3.3 Statistical Inference

This paper is about identification and the analysis therefore presumes that the population distribution of the observable variables is known without error. In practice, this population distribution must be estimated from a finite sample, which raises issues of statistical inference. In this section, I observe that the characterization of the identified

set provided by the PIES methodology can be rephrased in terms of moment equalities. This enables the application of results from the recent literature on inference in conditional and unconditional moment (in)equality models, such as Andrews and Soares (2010), Bugni (2010), Canay (2010), Andrews and Barwick (2012), Andrews and Shi (2013) and Romano et al. (2014).

For any $h \in \mathcal{H}^\dagger$, let $\{\mathcal{U}_x(h) : x \in \overline{\mathcal{X}}\}$ denote a collection of sets satisfying conditions U1–U5 in Theorem 1. Let $\theta \equiv (h, \{\overline{F}(u|x) : u \in \mathcal{U}_x(h), x \in \overline{\mathcal{X}}\})$ denote the parameters relevant to the characterization in Theorem 1. The identified set corresponding to θ is given by

$$\Theta^\star \equiv \left\{\theta : h \in \mathcal{H}^\star \text{ and } \overline{F}(u|x) \in \mathcal{F}^\star(u|x) \text{ for all } u \in \mathcal{U}_x(h), x \in \overline{\mathcal{X}}\right\},$$

where $\mathcal{F}^\star(u|x) \equiv \{F(u|x) : F \in \mathcal{F}^\star\}$ for any $u \in \overline{\mathbb{R}}^L$ and $x \in \mathcal{X}$.[37] Define the admissible set of θ by

$$\Theta^\dagger \equiv \left\{\theta : h \in \mathcal{H}^\dagger \text{ and } \{\overline{F}(u|x) : u \in \mathcal{U}_x(h), x \in \overline{\mathcal{X}}\} \text{ satisfies (T1.2)–(T1.6)}\right\}.$$

Also, define the conditional moment functions

$$m_{y|x}(\theta) \equiv \mathbb{E}\left[\mathbb{1}[Y \leq y] - \overline{\omega}_{y|x}(h, \overline{F}(\cdot|x))|X = x\right]$$

for all $y \in \overline{\mathcal{Y}}$ and $x \in \overline{\mathcal{X}}$. Then an implication of Theorem 1 is that

$$\Theta^\star = \left\{\theta \in \Theta^\dagger : m_{y|x}(\theta) = 0 \text{ a.e. } y \in \overline{\mathcal{Y}}, x \in \overline{\mathcal{X}}\right\}. \tag{22}$$

Andrews and Shi (2013) discuss and analyze test statistics for the null hypothesis $H_0 : \theta \in \Theta^\star$ when $\Theta^\star$ has a representation like (22). Their results are directly applicable here. Andrews and Shi (2013) recommend constructing confidence regions for $\theta^\star$, i.e. the value of θ that generated the data, by inverting this test. However, such an inversion procedure raises practical difficulties if the dimension of θ is large, as it often is in the PIES characterization. More tailored approaches for conducting inference on functionals of θ, such as the feature π introduced in the previous section, have been proposed by Bugni et al. (2014) and Kaido et al. (2015). These methods currently only apply to models with unconditional moments, so in practice one would need to transform the conditional moments into a set of unconditional moments, potentially leading to a loss of information. Given the trajectory of current research on inference

[37] Note that $\Theta^\star$ may be an outer set (vs. a sharp identified set) if $\overline{\mathcal{X}}$ and/or $\overline{\mathcal{Y}}$ are proper subsets of $\mathcal{X}$ and $\mathcal{Y}$, but that this is immaterial for characterization (22) ahead.

in moment (in)equality models, it seems likely that researchers in this area will soon develop projection inference approaches that exploit full conditional moment conditions. Torgovitsky (2015) shows in a related inference problem that for certain choices of criterion functions, the projection inference approach of Bugni et al. (2014) involves solving a convex program when the identified set is defined by a linear program. This is important for practice, since convex programs can be solved reliably. The same finding applies here. One useful feature of (22) is that it only involves moment equalities, not inequalities, which simplifies many of the statistical considerations discussed in the aforementioned papers.

4 PIES for the Two-Sector Roy Model

Consider the binary potential outcomes model

$$Y = DY_1 + (1 - D)Y_0, \tag{23}$$

where $D \in \{0, 1\}$ is a binary treatment and (Y_0, Y_1) are latent potential outcomes corresponding to different states of this treatment. The researcher observes (Y, D, X) where X is a vector of covariates with respect to which certain exclusion and/or independence conditions might be maintained. Analysis of this problem frequently maintains a weakly separable selection equation

$$D = \mathbb{1}[U_D \leq g_D(X)], \tag{24}$$

where U_D is a latent variable and g_D is an unknown function (Vytlacil, 2002; Heckman and Vytlacil, 2005). The two-sector Roy model refers to a specific case of this model in which the effect of X on Y_0 and Y_1 is made explicit through a latent variable formulation, say

$$Y_d = g_d(X, U_d) \quad \text{for } d = 0, 1, \tag{25}$$

where U_d, $d = 0, 1$ are latent random variables and $g_d, d = 0, 1$ are unknown functions. The functions g_d can be parameterized, or a completely agnostic approach can be taken by setting $g_d(X, U_d) = U_d$, in which case U_d is simply a relabelling of the potential outcome Y_d.

In the terminology of the previous section, (23)–(25) comprise a two-equation model

$$\begin{aligned} Y &= Dg_1(X, U_1) + (1 - D)g_0(X, U_0) \\ D &= \mathbb{1}[U_D \leq g_D(X)] \end{aligned}$$

with two internal random variables (Y, D), one external random vector X, and an $L =$ 3-dimensional vector of unobservables (U_0, U_1, U_D). The model structure $S = (h, F)$ is composed of the structural function $h = (g_0, g_1, g_D)$ and the conditional distribution function $F : \overline{\mathbb{R}}^3 \times \mathcal{X} \to [0, 1]$ for (U_0, U_1, U_D). Assuming for concreteness that g_0 and g_1 are invertible in their latent components,[38] a given structure S generates a distribution of (Y, D) (conditional on X) through the relationships

$$\begin{aligned} \mathbb{P}^S[Y \leq y, D = 0|X = x] &= \mathbb{P}^S\left[U_0 \leq g_0^{-1}(x, y), U_D > g_D(x)|X = x\right] \\ &= F(g_0^{-1}(x, y), \infty, \infty|x) - F(g_0^{-1}(x, y), \infty, g_D(x)|x) \qquad (26) \end{aligned}$$

and

$$\begin{aligned} \mathbb{P}^S[Y \leq y, D = 1|X = x] &= \mathbb{P}^S\left[U_1 \leq g_1^{-1}(x, y), U_D \leq g_D(x)|X = x\right] \\ &= F(\infty, g_1^{-1}(x, y), g_D(x)|x). \qquad (27) \end{aligned}$$

The right-hand sides of (26) and (27) provide expressions for what was called $\omega_{y|x}(h, F)$ in the previous section, noting the slight notation collision that now both Y and D are parts of "Y" from that section. Notice that, for a given h and (y, x) pair, both (26) and (27) are linear in the values of F evaluated at a finite number of points. The definition of the $\overline{\omega}$ function in Theorem 1 follows immediately for any choice of $\mathcal{U}_x(h)$ that contains the evaluation points in (26) and (27) for all (y, x) of interest.

If Y is continuously distributed, a typical restriction on h would be to assume that g_0 and g_1 are additively separable in their respective error distributions, in which case the invertibility maintained for (26) and (27) would be satisfied. If Y is a binary or more generally ordered discrete outcome, one might adopt a specification for g_0, g_1 that is similar to those in Section 2. Invertibility does not generally hold for these types of specifications, so (26) and (27) would be replaced by forms similar to (P2.1) from Section 2, but otherwise there is no conceptual difference. Mourifié et al. (2015) derived analytic expressions for sharp identified sets of certain parameters in nonparametric two-sector Roy models. The PIES approach provides a general method to compute these sets that is also applicable to the type of semiparametric Roy models commonly used in empirical work. However, the PIES procedure is computational, and does not

[38] As shown in Sections 2 and 3, this is not necessary for the analysis.

provide analytic expressions for identified sets.

A commonly maintained identifying assumption in the two-sector Roy model is that X is independent of the latent terms (U_0, U_1, U_D), see e.g. Assumption 1 in Eisenhauer et al. (2015). This can be imposed via the formalization in Assumption A3. Using Assumption A4, one could refine this strategy by imposing conditional independence restrictions only on certain components of (U_0, U_1, U_D). For both of these strategies, these assumptions could be imposed in such a way that all or only part of the components of X are used in these conditional independence statements. Much additional flexibility is possible. When considering the binary selection equation (24), it is common to maintain a nonparametric view of g_D, in which case the marginal distribution of U_D can be normalized, with a typical choice being the uniform distribution on the $[0, 1]$ interval. This type of normalization can be accommodated in PIES through Assumption A2; see Example 3.1.

Implementing PIES for the two-sector Roy model is now a matter of deciding which assumptions to impose and then appealing to Theorem 1. As a concrete example, write X as $X = (X_0, X_1, X_D)$, where each of these subvectors has a constant term, and all may have some components in common. Then assume that $g_0(X, U_0) = X_0'\beta_0 + U_0$, $g_1(X, U_1) = X_1'\beta_1 + U_1$, and $g_D(X) = X_D'\beta_D$ for some parameter vectors $\beta = (\beta_0, \beta_1, \beta_D)$. Suppose in addition that each of these errors has median 0, conditional on X. Let $\overline{\mathcal{Y}}$ and $\overline{\mathcal{X}}$ denote subsets of the supports of Y and X. For a fixed value of β, define

$$\begin{aligned} \mathcal{U}_{0,x}(\beta) &\equiv \{y - x_0'\beta_0 : y \in \overline{\mathcal{Y}}\} \cup \{0, \pm\infty\} \\ \mathcal{U}_{1,x}(\beta) &\equiv \{y - x_1'\beta_1 : y \in \overline{\mathcal{Y}}\} \cup \{0, \pm\infty\} \\ \text{and } \mathcal{U}_{D,x}(\beta) &\equiv \{x_D'\beta_D, 0, \pm\infty\}, \end{aligned}$$

and define $\mathcal{U}_x(\beta) \equiv \mathcal{U}_{0,x}(\beta) \times \mathcal{U}_{1,x}(\beta) \times \mathcal{U}_{D,x}(\beta)$. Theorem 1 shows that if β is in the identified set, then there exist functions $\overline{F}(\cdot|x) : \mathcal{U}_x(\beta) \to [0, 1]$ that satisfy the following linear system of equations:

$$\begin{aligned} &\overline{F}(y - x_0'\beta_0, \infty, \infty|x) - \overline{F}(y - x_0'\beta_0, \infty, x_D'\beta_D|x) \\ &\qquad = \mathbb{P}[Y \le y, D = 0|X = x] \qquad (28) \\ \text{and } &\overline{F}(\infty, y - x_1'\beta_1, x_D'\beta_D|x) = \mathbb{P}[Y \le y, D = 1|X = x] \text{ for all } y \in \overline{\mathcal{Y}},\ x \in \overline{\mathcal{X}}, \qquad (29) \\ &\overline{F}(\cdot|x) \text{ is a 3–dimensional subdistribution for each } x \in \overline{\mathcal{X}}, \qquad (30) \\ &\overline{F}(0, \infty, \infty|x) = \overline{F}(\infty, 0, \infty|x) = \overline{F}(\infty, \infty, 0|x) = \frac{1}{2} \text{ for all } x \in \overline{\mathcal{X}}. \qquad (31) \end{aligned}$$

If $\overline{\mathcal{Y}} = \mathcal{Y}$ and $\overline{\mathcal{X}} = \mathcal{X}$ then the existence of a solution to this linear system provides a sharp characterization, i.e. the only β for which it is satisfied are those that are elements of the identified set. If $\overline{\mathcal{Y}}$ and/or $\overline{\mathcal{X}}$ are infinite, then the characterization can be made arbitrarily close to sharp by making $\overline{\mathcal{Y}}$ and/or $\overline{\mathcal{X}}$ arbitrarily large subsets of $\mathcal{Y}$ and/or $\mathcal{X}$. The system of equations in (28)–(31) is just one example that corresponds to a specific choice of assumptions. The generality of the PIES approach in Theorem 1 shows how this example can be tailored in many directions for specific applications.

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 grounded in copula theory, and in particular uses a generalization of Sklar's (1959; 1996) multilinear interpolation lemma for subcopulas 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 from standard probability theory. I have shown that applying PIES involves solving linear programming problems in at least two important semiparametric nonlinear models: (i) An ordered discrete response model with possibly endogenous explanatory variables, and (ii) the two-sector Roy model. In work in preparation, I show how to apply PIES to random coefficient binary response and multinomial discrete choice models. There appear to be many other interesting applications as well.

It is methodologically important to have general computational procedures for computing sharp identified sets. Sharp analytic characterizations of identified sets take considerable effort to derive in even fairly simple models. While these characterizations can provide theoretical insights, practitioners cannot be expected to re-work such results to suit the specific needs of their application. Moreover, the simulation results presented in Section 2 suggest that even in a relatively simple semiparametric bivariate binary response model, the sharp identified sets for the index parameters can have wild and irregular shapes. (Indeed, it would be challenging enough to find analytic expressions for some of the sets in Figures 2 and 3 for even a single distribution of the observables, let alone in general.) The development of suitable computational approaches for characterizing sharp identified sets is a necessary ingredient for broadening the empirical impact of partial identification analysis.[39]

[39]There is an argument to be made in favor of the simplicity of utilizing only outer identified sets. For

Other authors have proposed different general methodologies for constructing sharp identified sets. The frameworks that I am aware of are those using random set theory (Beresteanu et al. (2011), Galichon and Henry (2011), Chesher and Rosen (2014a,b), Chesher et al. (2013), Aradillas-Lopez and Rosen (2014)) and information theory (Schennach, 2014).[40] Ultimately, the utility of the PIES methodology relative to the approaches advocated by these papers can only be judged retrospectively by the quantity and quality of their subsequent applications. However, I will provide some brief and admittedly speculative comments on their relative strengths as I understand them. Schennach (2014) provides a discussion of her ELVIS approach compared to the approaches of Beresteanu et al. (2011) and Galichon and Henry (2011).

Unlike this paper, Beresteanu et al. (2011) and Schennach (2014) consider models defined by finite sets of moment conditions. A benefit of this focus is that they can directly consider mean and correlation restrictions, which appear to be less straightforward to impose using PIES. A drawback is that these approaches cannot easily handle the types of independence restrictions commonly imposed in nonlinear models, since such restrictions require an infinite number of moment conditions.[41] For the same reason, these approaches are also unsuited to imposing shape restrictions (such as symmetry) on the distribution of unobservables. The use of distributional shape restrictions to shrink identified sets in semiparametric nonlinear models seems like a particularly promising direction for future research.

The analysis of Galichon and Henry (2011) and Chesher and Rosen (2014b), who consider characterizations based on the random set theory concept of a capacity functional, seems to be more closely comparable to PIES. Both approaches consider the entire distribution of the latent variables, rather than pre-specified moments. Galichon and Henry (2011) only analyze models with unobservables that are finitely parameterized, which is different than the focus on semiparametric models emphasized here. On the other hand, those authors (along with Beresteanu et al. (2011)) appear to be

example, Ciliberto and Tamer (2009) and Pakes (2010) are able to make substantive conclusions using outer sets derived from revealed preference arguments. However, the lack of sharpness leaves open the tantalizing possibility that even stronger conclusions could be reached without maintaining stronger assumptions by finding a low-level characterization of the sharp identified set.

[40]Note that there are many other papers in the literature that employ computational approaches for *specific* partially identified models. Many of these approaches reduce—like PIES—to linear programming problems. Examples include Honoré and Lleras-Muney (2006), Honoré and Tamer (2006), Manski (2007, 2014), Molinari (2008), Kitamura and Stoye (2013), Freyberger and Horowitz (2014), and Lafférs (2015). The focus of this discussion is on *general* approaches for constructing partially identified sets that are applicable to large classes of models.

[41]Schennach (2014, Section 4.3) provides an extension of ELVIS to the case of an infinite number of moment conditions by using an asymptotic approximation argument. In contrast, the application of PIES to these types of restrictions often yields exact characterizations.

primarily motivated by game theoretic models with multiple equilibria, which seems to be an area in which random set theory approaches may have substantial theoretical and practical advantages over PIES. Chesher and Rosen (2014b) consider models that may have unobservables with distributions that are not finitely parameterized, and so their aim is more directly comparable to the goals considered in this paper.

The practical differences between the method of Chesher and Rosen (2014b) and the PIES approach in this paper appear to primarily turn on the relative difficulty of the computational implementation. For Chesher and Rosen (2014b), the difficulty is finding a way to approximate the uncountably infinite number of restrictions that comprise their characterization of the identified set. For the simulation exercises considered by Chesher and Rosen (2014b, Section 4), the restrictions that need to be satisfied correspond to the set of all closed intervals contained in $[0, 1]$. The authors address this by using a relatively small and arbitrary grid of such intervals, which nevertheless appears to provide a reasonably close outer approximation of the sharp identified set, although this of course cannot be a theorem. The PIES approach faces a similar problem in situations where the observed variables have an uncountably infinite support, since a potentially infinite number of restrictions play a role in characterizing the identified set for these cases. However, in contrast to Chesher and Rosen (2014b), these restrictions are indexed by all possible realizations of the observed variables. This difference is important because in practice the empirical distribution associated with any given sample of data provides only a finite number of support points. It is reasonable to expect that under relatively weak conditions, sharp identified sets based on this discrete empirical distribution will converge as the sample size increases to sharp identified sets based on the population distribution. I leave a precise formulation of such a result to future work.

One positive and very important quality that appears to be unique to PIES is the ability to easily consider causal parameters that depend both on the structural function (h in the preceding discussion) and the distribution of unobservables (F). The authors of the aforementioned papers tend to view F as a nuisance parameter to be eliminated in the service of characterizing the identified set of h.[42] However, causal parameters in nonlinear models typically depend on the distribution of unobservables. PIES makes it straightforward for a researcher to consider sharp identified sets for these parameters, since it reduces the distribution of unobservables to a subdistribution, rather than trying to eliminate it completely. The same does not appear to be true for the methods proposed by the other papers. A related benefit that appears to be unique to PIES

[42]For example, Schennach (2014, pg. 364) explicitly describes the distribution of unobservables as an infinite-dimensional nuisance parameter.

is that, by more carefully considering the relationship between h and F, it provides a method of reducing the computational problem of searching over all h, see Proposition 2.3. This is practically important, since simple grid searches become computationally infeasible with even relatively small numbers of parameters. None of the other general approaches appear to provide an obvious solution to this problem.

The point of this discussion is not to claim that PIES or any of the other cited methods strictly dominates any one of the others. It would be naive to believe this to be true. The utility of a given method must be judged in the context of the model and empirical problem at hand. In this paper, I have provided two examples of widely used empirical models to which the PIES method seems better suited than the frameworks proposed in the aforementioned papers. Those papers contain examples for which the opposite is true. The purpose of providing another general approach to characterizing identified sets is to increase the size of the toolbox available to empirical researchers seeking to apply partial identification methods. The applications of PIES in this paper to the widely used ordered discrete response model, and to the two-sector Roy model, should solidify its claim to be a useful part of this toolbox, without detracting from any of the other tools already there.

Appendix A Copula Theory

This section records some definitions and results from copula theory that are used in the proof of Lemma 3.2. See Schweizer and Sklar (1983) and Nelsen (2006) for further discussion.

An L-dimensional copula is similar to a distribution function, but defined instead on $[0,1]^L$, and with 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 most important result in copula theory is a two-part theorem due to Sklar (1959) (see e.g. 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 3.2. 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) recent monograph on copula theory (see Lemma 2.3.5). Sklar (1996) provides 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 Proof of Lemma 3.2 and Corollary 3.1

The proof of Lemma 3.2 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 3.2 (case** $L = 1$**). Suppose that $\overline{F}$ is a one-dimensional subdistribution with domain $\mathcal{U}$ that is closed and contains $\{-\infty, +\infty\}$. 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 $\mathcal{U} \setminus \{-\infty, +\infty\}$ 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 $\overline{\mathcal{U}}$, i.e. $\overline{\mathcal{U}}^c \equiv \overline{\mathbb{R}} \setminus \overline{\mathcal{U}}$ into three sets

$$\overline{\mathcal{U}}^c = \overline{\mathcal{U}}^c_- \cup \overline{\mathcal{U}}^c_0 \cup \overline{\mathcal{U}}^c_+, \tag{32}$$

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

$$F(u) = \begin{cases} 0 & \text{if } u = -\infty \\ \overline{F}(\underline{u})e^{u-\underline{u}} & \text{if } u \in \overline{\mathcal{U}}^c_- \\ \ell(u; [a_u, b_u]) & \text{if } u \in \overline{\mathcal{U}}^c_0 \\ \overline{F}(u) & \text{if } u \in \overline{\mathcal{U}} \\ \overline{F}(\overline{u}) + (1 - \overline{F}(\overline{u}))(1 - e^{\overline{u}-u}) & \text{if } u \in \overline{\mathcal{U}}^c_+ \\ 1 & \text{if } u = +\infty, \end{cases} \tag{33}$$

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, i.e.

$$\ell(u;[a_u,b_u]) \equiv \overline{F}(a_u) + \left(\frac{\overline{F}(b_u)-\overline{F}(a_u)}{b_u-a_u}\right)(u-a_u), \tag{34}$$

noting that $a_u < b_u$ for all $u \notin \overline{\mathcal{U}}$. 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 10. It is straightforward to verify that F is a proper, non-defective, 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 continuous on its entire domain. *Q.E.D.*

Proof of Lemma 3.2 (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 $\overline{F}$, i.e.

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

Lemma 3.1 shows that each $\overline{F}_l$ is itself a one-dimensional subdistribution function.

Next, define the set

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

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

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

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 $\{-\infty, +\infty\} \in \mathcal{U}_l$ with $\overline{F}_l(-\infty) = 0$ and $\overline{F}_l(+\infty) = 1$. Next, notice that

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

and similarly if any other collection of the arguments of $\overline{C}$ are 0. This shows that C2

is satisfied. Condition C3 is satisfied because, by construction,

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

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

To see that $\overline{C}$ satisfies 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 $t'' = (\overline{F}_1(u''_1), \ldots, \overline{F}_L(u''_L))$.[43] 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,

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

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 $t \in \mathrm{Vrt}(t', t'')$. Hence,

$$\begin{aligned}\mathrm{Vol}_{\overline{C}}(t', t'') &\equiv \sum_{t \in \mathrm{Vrt}(t',t'')} \mathrm{sgn}_{(t',t'')}(t)\overline{C}(t) \\ &= \sum_{t \in \mathrm{Vrt}(t',t'')} \mathrm{sgn}_{(u',u'')}(\xi(t))\overline{F}(\xi(t)) \\ &= \sum_{u \in \mathrm{Vrt}(u',u'')} \mathrm{sgn}_{(u',u'')}(u)\overline{F}(u) \equiv \mathrm{Vol}_{\overline{F}}(u', u''), \end{aligned} \tag{35}$$

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 $u' \leq u''$. Hence, by (35), $\mathrm{Vol}_{\overline{C}}(t', t'') \geq 0$ as well. Since t', t'' were arbitrary elements of $\mathcal{T}$ with $t' \leq t''$, this shows that $\overline{C}$ satisfies C1.

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, non-

[43] 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.

defective, 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

$$F : \overline{\mathbb{R}}^L \to [0,1] : F(u) = C(F_1(u_1), \ldots, F_L(u_L)). \tag{36}$$

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 $u \in \mathcal{U}$,

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

Finally, notice that because copulas are continuous (Theorem 2.2.4 of Nelsen (2006)) the non-defectivity of each F_l implies that F is non-defective as well. In addition, 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. This establishes the claims of the lemma. *Q.E.D.*

Proof of Corollary 3.1. The case where $L = 1$ follows tautologically from the assumption and definition of extendibility, so assume that $L > 1$. Follow the proof of Lemma 3.2 up to (36). 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 (36) with these margins F_l for $l = 1, \ldots, L$ and appealing to Sklar's Theorem shows that F is a proper, non-defective, L-dimensional distribution function with margins $F_l \in \mathcal{F}_l$ for each $l = 1, \ldots, L$. *Q.E.D.*

Appendix C Proof of Theorem 1

Suppose that $p \in \mathcal{P}^\star$. By definition, there exists an $h \in \mathcal{H}^\star$ and an $F \in \mathcal{F}^\star(h)$ such that $\pi(h, F) = p$. For every $x \in \overline{\mathcal{X}}$ and $u \in \mathcal{U}_x(h)$, 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 $\mathcal{U}_x(h)$ is a subdistribution on that subset.
2. Condition U2, $F \in \mathcal{F}^\star(h)$, and A1 imply that (T1.1), (T1.3) and (T1.7) are satisfied.
3. Condition U3 and A2 imply that (T1.4) is satisfied.

4. Condition U4 and A3 imply that (T1.5) is satisfied.

5. Condition U5 and A4 imply that (T1.6) is satisfied.

Hence, there exists an $h \in \mathcal{H}^\dagger$ and functions $\overline{F}(\cdot|x) : \mathcal{U}_x(h) \to [0,1]$ 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 exists an $h \in \mathcal{H}^\dagger$ and functions $\overline{F}(\cdot|x) : \mathcal{U}_x(h) \to [0,1]$ for $x \in \mathcal{X}$ such that (T1.1)–(T1.7) are satisfied. As with the propositions in Section 2, the proof that $p \in \mathcal{P}^\star$ will use an extension argument based on Lemma 3.2 and Corollary 3.1. Given the general nature of the theorem, the extension argument will be conducted separately for different $x \in \mathcal{X}$, so I will first partition $\mathcal{X}$ in a particular way.

To this end, observe first that, necessarily $\mathcal{X}_0^\dagger \subseteq \mathcal{X}_l^\dagger$ for all $l = 1, \ldots, L$, given their definitions in A3 and A4. Let $\mathcal{L} \subseteq \{1, \ldots, L\}$ denote a subset of the integers between 1 and L, and let $\mathfrak{L}$ denote the collection of all such subsets $\mathcal{L}$. In particular, note that $\emptyset \in \mathfrak{L}$. For any $\mathcal{L} \in \mathfrak{L}$, define $\mathcal{X}_\mathcal{L}^\dagger \equiv \{x \in \mathcal{X} \setminus \mathcal{X}_0^\dagger : x \in \mathcal{X}_l^\dagger \text{ for all } l \in \mathcal{L}\}$. Then $\mathcal{X}$ can be partitioned into $\{\mathcal{X}_0^\dagger\} \cup \{\mathcal{X}_\mathcal{L}^\dagger : \mathcal{L} \in \mathfrak{L}\}$. That is, every $x \in \mathcal{X}$ satisfies exactly one of the following conditions: (i) $x \in \mathcal{X}_0^\dagger$, (ii) $x \notin \mathcal{X}_0^\dagger$ but $x \in \mathcal{X}_l^\dagger$ for some non-empty subset $\mathcal{L}$ of $l \in \{1, \ldots, L\}$, or (iii) $x \notin \mathcal{X}_0^\dagger$ and $x \notin \mathcal{X}_l^\dagger$ for every l, i.e. $x \in \mathcal{X}_\mathcal{L}^\dagger$ with $\mathcal{L} = \emptyset$.

Now, fix an arbitrary $\overline{x} \in \mathcal{X}_0^\dagger$. Then $\overline{F}(\cdot|\overline{x})$ is a subdistribution with domain $\mathcal{U}_{\overline{x}}(h)$ by (T1.2) and $\overline{F}_l(\cdot|\overline{x}) \in \overline{\mathcal{F}}_{l,\overline{x}}^\dagger$ for each $l = 1, \ldots, L$ by (T1.4). Given U3, Corollary 3.1 implies that there exists a proper L-dimensional distribution function $\widetilde{F}_0$ such that $\widetilde{F}_0(u) = \overline{F}(u|\overline{x})$ for all $u \in \mathcal{U}_{\overline{x}}(h)$, and such that the lth margin of $\widetilde{F}_0$ is in $\mathcal{F}_{l,\overline{x}}^\dagger$ for every $l = 1, \ldots, L$. Define $F_0 : \overline{\mathbb{R}}^L \times \mathcal{X}_0^\dagger \to [0,1] : F_0(u|x) = \widetilde{F}_0(u)$. Observe that A2 and A3 together require that $\mathcal{F}_{l,x}^\dagger = \mathcal{F}_{l,x'}^\dagger$ for all $x, x' \in \mathcal{X}_0^\dagger$. Hence, the lth margin of $F_0(\cdot|x)$ is an element of $\mathcal{F}_{l,\overline{x}}^\dagger = \mathcal{F}_{l,x}^\dagger$ for all $x \in \mathcal{X}_0^\dagger$. In addition, $F_0(u|x) = \widetilde{F}_0(u) = \overline{F}(u|\overline{x}) = \overline{F}(u|x)$ for all $x \in \mathcal{X}_0^\dagger$ and $u \in \mathcal{U}_x(h)$, by U4 and (T1.5).

Next, let $\mathcal{L}$ be an element of $\mathfrak{L}$ (possibly the emptyset) and fix an arbitrary $\overline{x} \in \mathcal{X}_\mathcal{L}^\dagger$ (if any exist). For each $l \in \mathcal{L}$ (if any), let $\overline{F}_l(\cdot|\overline{x})$ be the lth margin of $\overline{F}(\cdot|\overline{x})$, which is a one-dimensional subdistribution function, given U1 and (T1.2). By U3 and (T1.4), for each $l \in \mathcal{L}$ there exists a function $\widetilde{F}_l : \overline{\mathbb{R}} \to [0,1]$ such that $\widetilde{F}_l(u) = \overline{F}_l(u|\overline{x})$ for every $u \in \mathcal{U}_{l,\overline{x}}(h)$ and such that $\widetilde{F}_l \in \mathcal{F}_{l,\overline{x}}^\dagger$. Observe that A2 and A4 together require that $\mathcal{F}_{l,x}^\dagger = \mathcal{F}_{l,x'}^\dagger$ for all $x, x' \in \mathcal{X}_l^\dagger$ and every $l = 1, \ldots, L$. Hence, $\widetilde{F}_l(u)$ is an element of $\mathcal{F}_{l,\overline{x}}^\dagger = \mathcal{F}_{l,x}^\dagger$ for all $l \in \mathcal{L}$ and $x \in \mathcal{X}_\mathcal{L}^\dagger \subseteq \mathcal{X}_l^\dagger$. In addition, $\widetilde{F}_l(u) = \overline{F}_l(u|\overline{x}) = \overline{F}_l(u|x)$ for all $x \in \mathcal{X}_\mathcal{L}^\dagger$, all $u \in \mathcal{U}_{l,x}(h)$, and each $l \in \mathcal{L}$. Hence, for each $x \in \mathcal{X}_\mathcal{L}^\dagger$, the singleton set $\{\overline{F}_l(\cdot|x)\}$ is extendible to the singleton set $\{\widetilde{F}_l\}$ for every $l \in \mathcal{L}$, while for every $l \notin \mathcal{L}$, (T1.4) and U3 imply that the singleton set $\{\overline{F}_l(\cdot|x)\}$ is extendible to $\mathcal{F}_{l,x}^\dagger$. Given (T1.2), Corollary 3.1 then implies that for each $x \in \mathcal{X}_\mathcal{L}^\dagger$, there exists a proper

L-dimensional distribution function $F_{\mathcal{L}}(\cdot|x)$ such that $F_{\mathcal{L}}(u|x) = \overline{F}(u|x)$ for all $u \in \overline{\mathbb{R}}^L$, and such that the lth margins of $F_{\mathcal{L}}(\cdot|x)$—call them $F_{l,\mathcal{L}}(\cdot|x)$—are equal to $\widetilde{F}_l$ for every $l \in \mathcal{L}$, and elements of $\mathcal{F}^{\dagger}_{l,x}$ for every $l \notin \mathcal{L}$. It follows that $F_{l,\mathcal{L}}(\cdot|x) \in \mathcal{F}^{\dagger}_{l,x}$ for every $l = 1, \ldots, L$ and all $x \in \mathcal{X}^{\dagger}_l$, and that $F_{l,\mathcal{L}}(u|x) = \widetilde{F}_l(u) = F_{l,\mathcal{L}}(u|x')$ for every $l \in \mathcal{L}$ and all $x, x' \in \mathcal{X}^{\dagger}_l$.

Finally, combine these constructions together into the function

$$F : \overline{\mathbb{R}}^L \times \mathcal{X} \to [0,1] : F(u|x) = \begin{cases} F_0(u|x) & \text{if } x \in \mathcal{X}^{\dagger}_0 \\ F_{\mathcal{L}}(u|x) & \text{if } x \in \mathcal{X}^{\dagger}_{\mathcal{L}}. \end{cases} \tag{37}$$

Observe that F is well-defined through (37), since $\mathcal{X}^{\dagger}_0 \cup \{\mathcal{X}^{\dagger}_{\mathcal{L}} : \mathcal{L} \in \mathfrak{L}\}$ forms a partition of $\mathcal{X}$. Also observe that, given the previous discussion, it is known that $F(u|x) = \overline{F}(u|x)$ for all $u \in \mathcal{U}_x(h)$ and all $x \in \mathcal{X}$. In addition, the previous discussion already showed that F satisfies A2–A4. Given (T1.3), U2, and the established fact that each $F(\cdot|x)$ extends each $\overline{F}(\cdot|x)$, it is also known that $\rho(h, F) \geq \vec{0}$, so that A1 is satisfied. Hence, $F \in \mathcal{F}^{\dagger}$. Moreover, U2, (T1.1), and the assumption that $\overline{\mathcal{Y}} = \mathcal{Y}, \overline{\mathcal{X}} = \mathcal{X}$ imply that $\omega_{y|x}(h, F) = \mathbb{P}[Y \leq y | X = x]$ for all $y \in \mathcal{Y}$ and $x \in \mathcal{X}$, so that $F \in \mathcal{F}^{\star}(h)$. Lastly, U2 and (T1.7) imply that $\pi(h, F) = p$ which, since $F \in \mathcal{F}^{\star}(h)$ and $h \in \mathcal{H}^{\dagger}$, implies that $p \in \mathcal{P}^{\star}$.

Q.E.D.

	Exogenous X_1			Endogenous X_1 (Y_2)					
Assumption	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]
$\text{med}(U_1 \mid X_1, X_2) = 0$	✓	✓	✓						
$U_1 \mid X_1, X_2$ symmetric around 0		✓	✓						
$U_1 \perp\!\!\!\perp (X_1, X_2)$			✓						
$\text{med}(U_1 \mid X_2, X_3) = 0$				✓	✓	✓	✓	✓	✓
$U_1 \mid X_2, X_3$ symmetric around 0						✓			✓
$U_1 \perp\!\!\!\perp (X_2, X_3)$					✓	✓	✓	✓	✓
$Y_2 = \mathbb{1}[g_{21}(X_2, X_3) > U_2]$							✓		
$(U_1, U_2) \perp\!\!\!\perp (X_2, X_3)$							✓	✓	✓
$Y_2 = \mathbb{1}[\pi_0 + \pi_2 X_2 + \pi_3 X_3 > U_2]$								✓	✓
$\text{med}(U_2 \mid X_2, X_3) = 0$								✓	✓
$U_2 \mid X_2, X_3$ symmetric around 0									✓

Table 1: Assumptions used in the binary response model simulations. Specifications [1]–[3] model only the outcome equation and assume that X_1 and X_2 are exogenous, using increasing concepts of exogeneity. Specifications [4]–[6] allow for X_1 (now called Y_2) to be endogenous and place restrictions on the relationship between (X_2, X_3) and U_1. These specifications can be seen as a single equation semiparametric instrumental variables model with excluded instrument X_3. Specification [7] adds a nonparametric first stage equation and imposes independence between (X_2, X_3) and (U_1, U_2). Specification [8] imposes a linear-in-parameters parametric form for this first stage equation. Specification [9] additionally imposes conditional symmetry on both U_1 and U_2.

(K_2, K_3)	(3, 3)	(5, 3)	(3, 5)	(5, 5)
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, 0.757]	[-0.254, 0.735]	[-0.219, 0.754]	[-0.238, 0.733]
[5]	[-0.173, -0.027] ∪ [0.027, 0.685]	[-0.189, -0.030] ∪ [0.030, 0.660]	[-0.109, -0.053] ∪ [0.053, 0.638]	[-0.119, -0.060] ∪ [0.060, 0.615]
[6]	[-0.173, -0.027] ∪ [0.027, 0.666]	[-0.189, -0.030] ∪ [0.030, 0.629]	[-0.109, -0.053] ∪ [0.053, 0.616]	[-0.119, -0.060] ∪ [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]

Table 2: Sharp identified sets for the specifications listed in Table 1, for several values of (K_2, K_3). The supports for different values of (K_2, K_3) are given in (8). Note that the second header row of the table lists the true value of the average treatment effect which varies with K_2. 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.

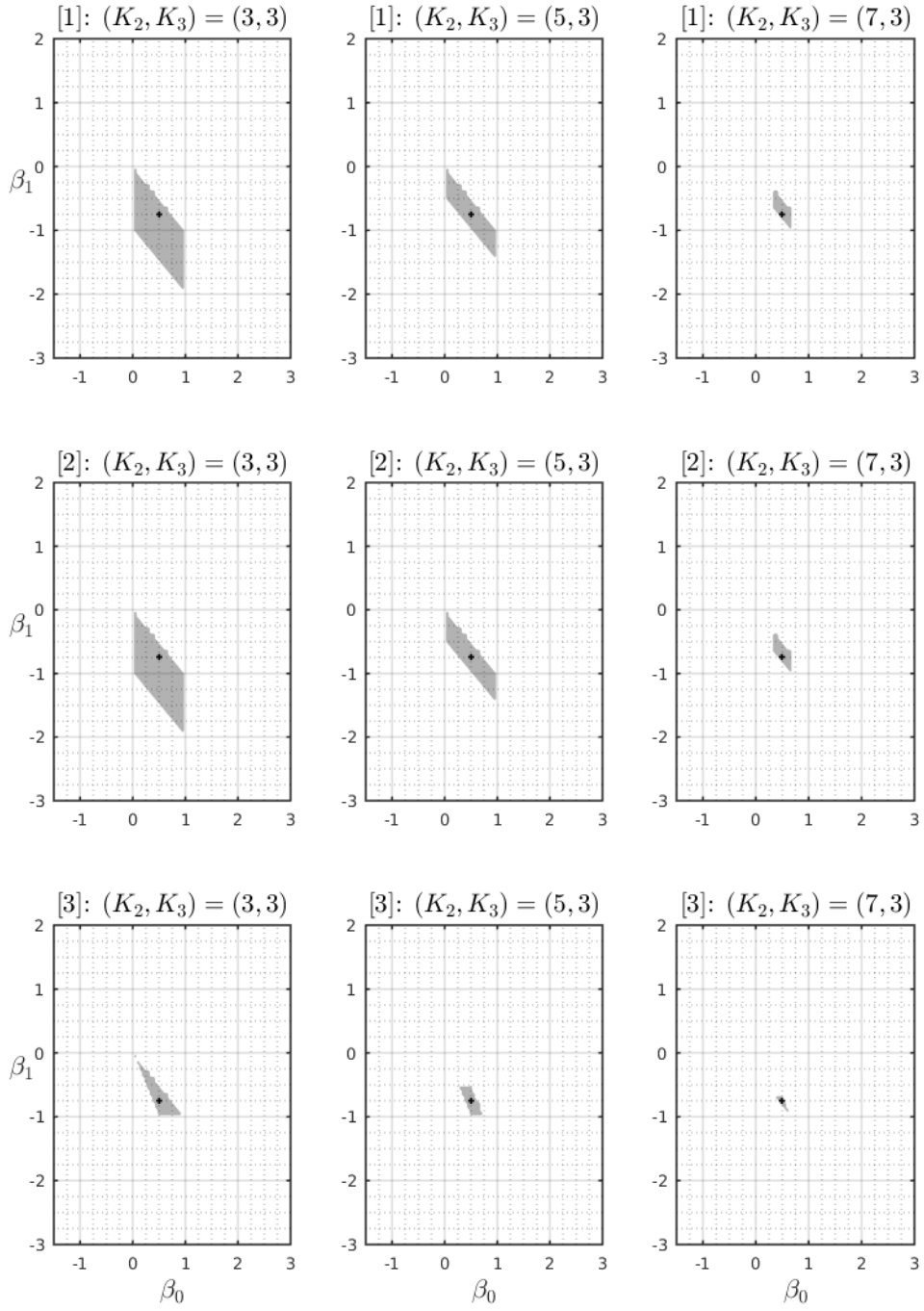

Figure 1: Sharp identified sets for (β_0, β_1) under specifications [1]–[3] in Table 1 with different choices of K_2 and K_3. The shaded area indicates the identified set. The black hash mark is placed at $(\beta_0^\star, \beta_1^\star) = (.5, -.75)$.

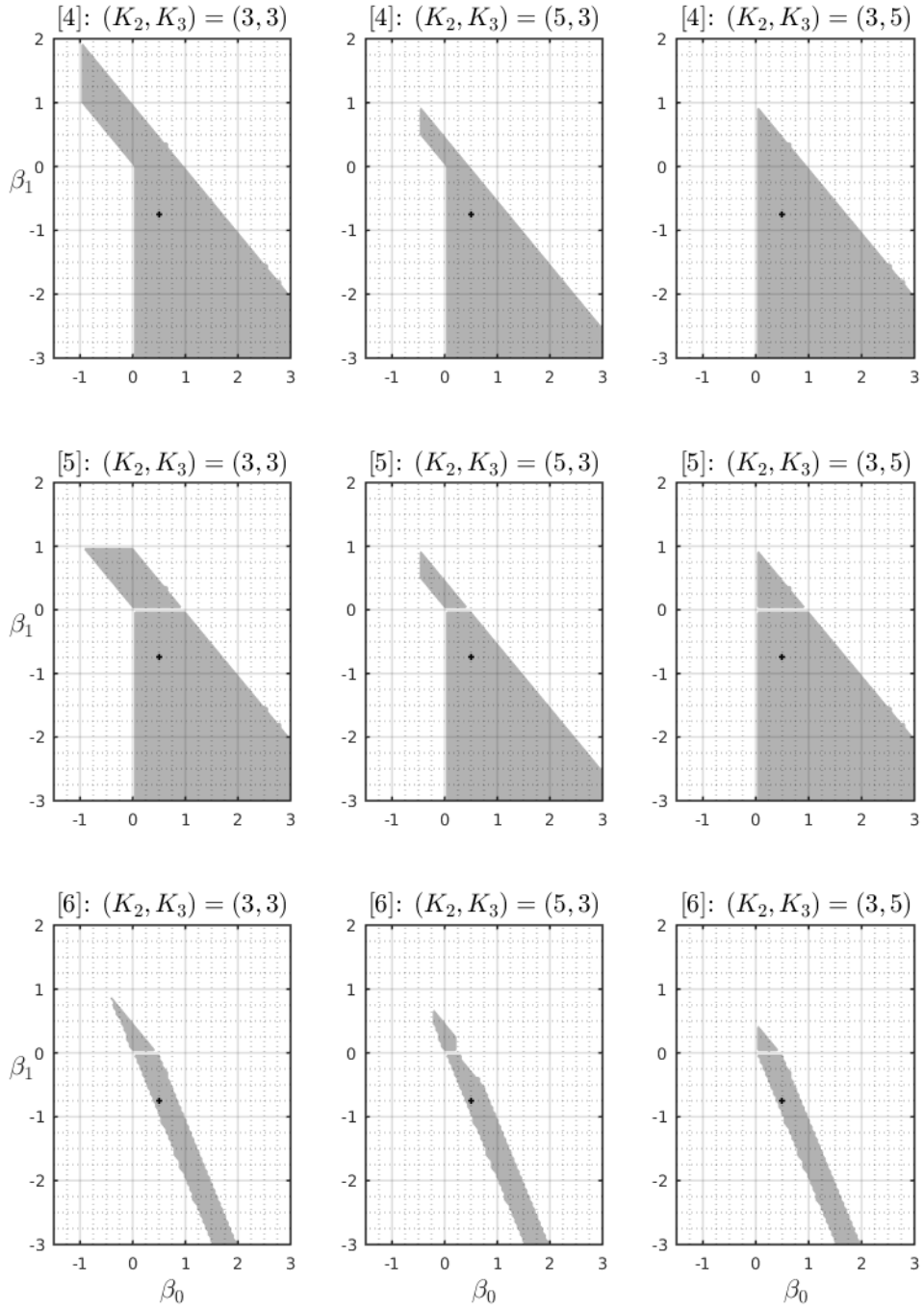

Figure 2: Sharp identified sets for (β_0, β_1) under specifications [4]–[6] in Table 1.

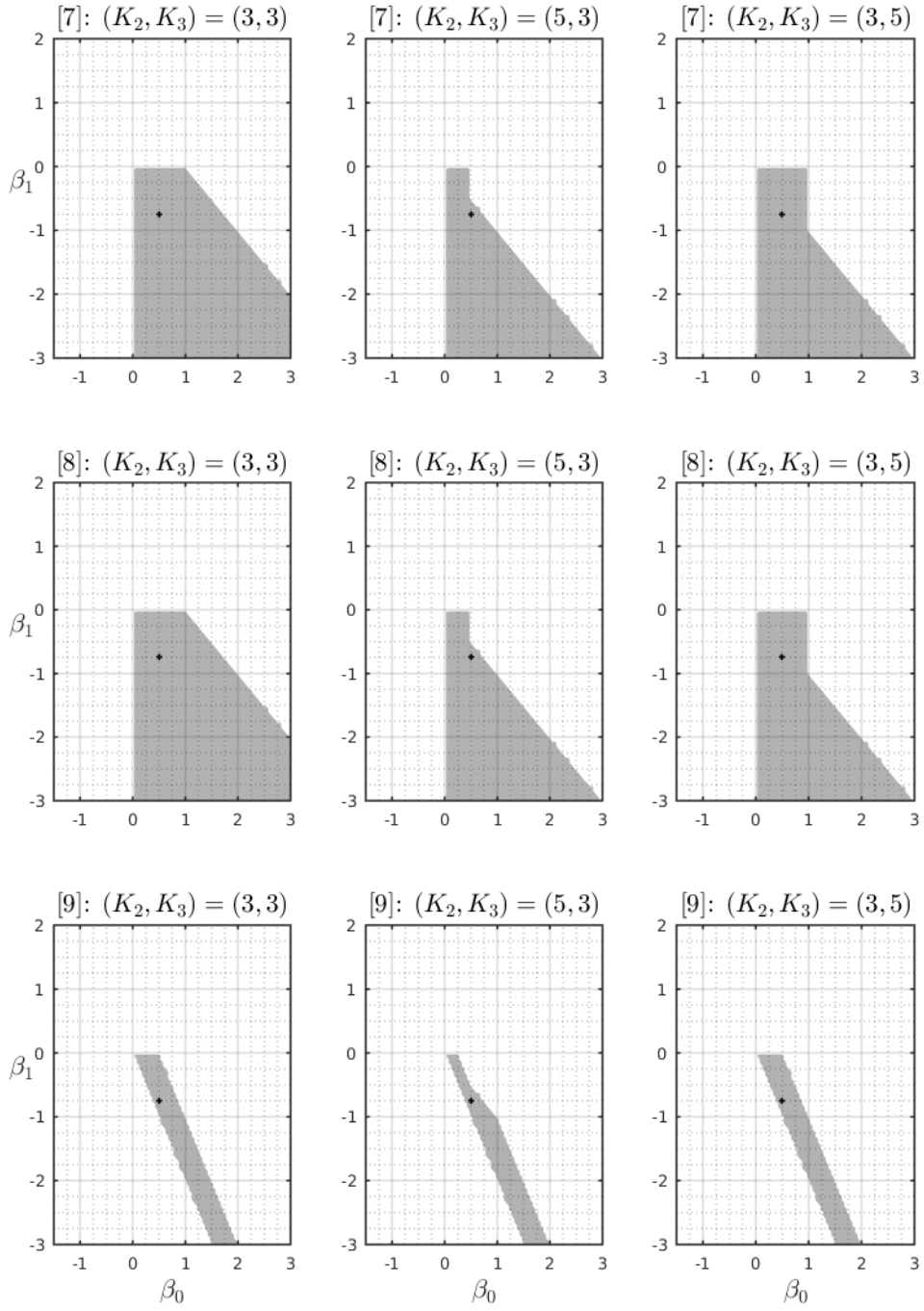

Figure 3: Sharp identified sets for (β_0, β_1) under specifications [7]–[9] in Table 1.

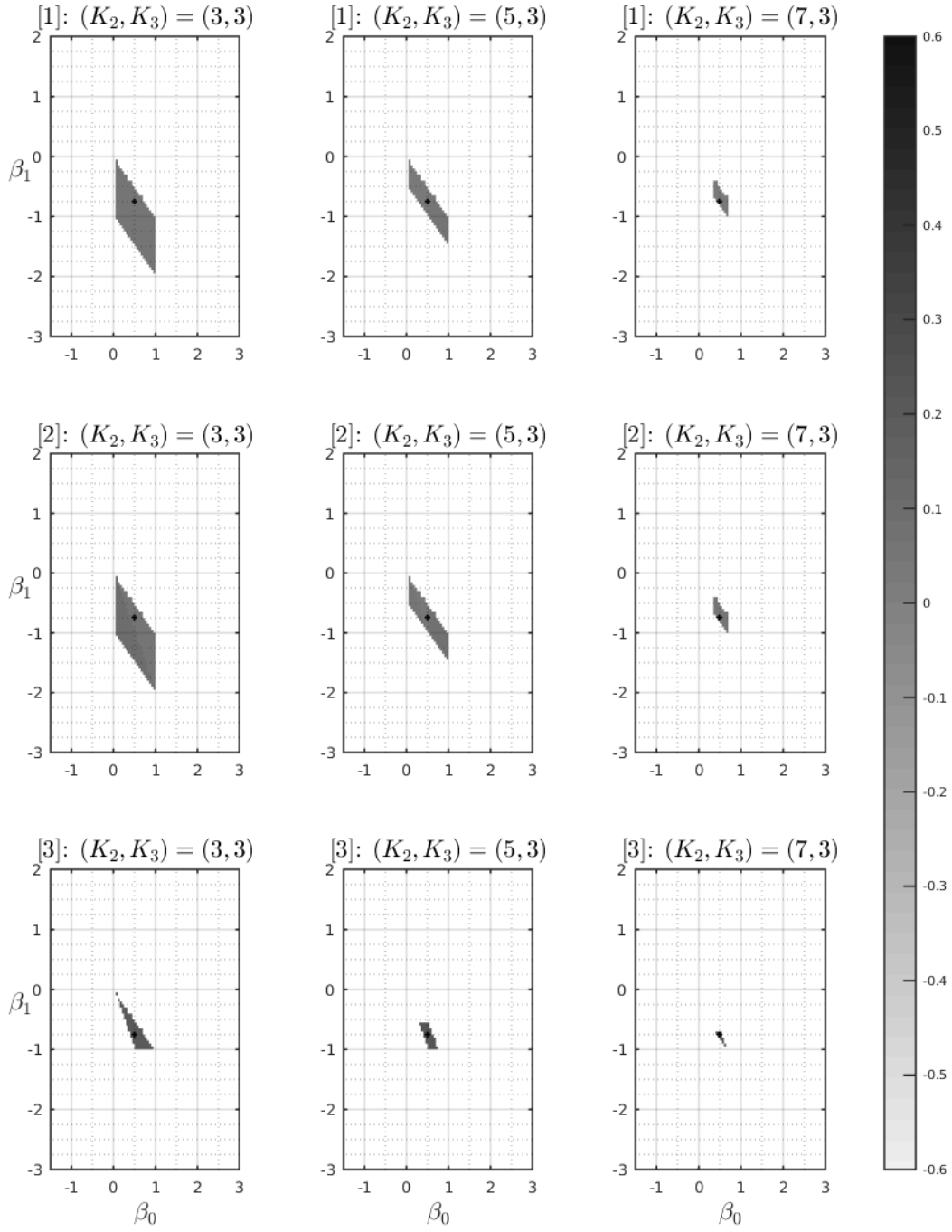

Figure 4: Sharp lower bounds for the ATE at a given fixed value of (β_0, β_1) corresponding to the sharp identified sets of (β_0, β_1) in Figure 1. Darker regions represent larger lower bounds.

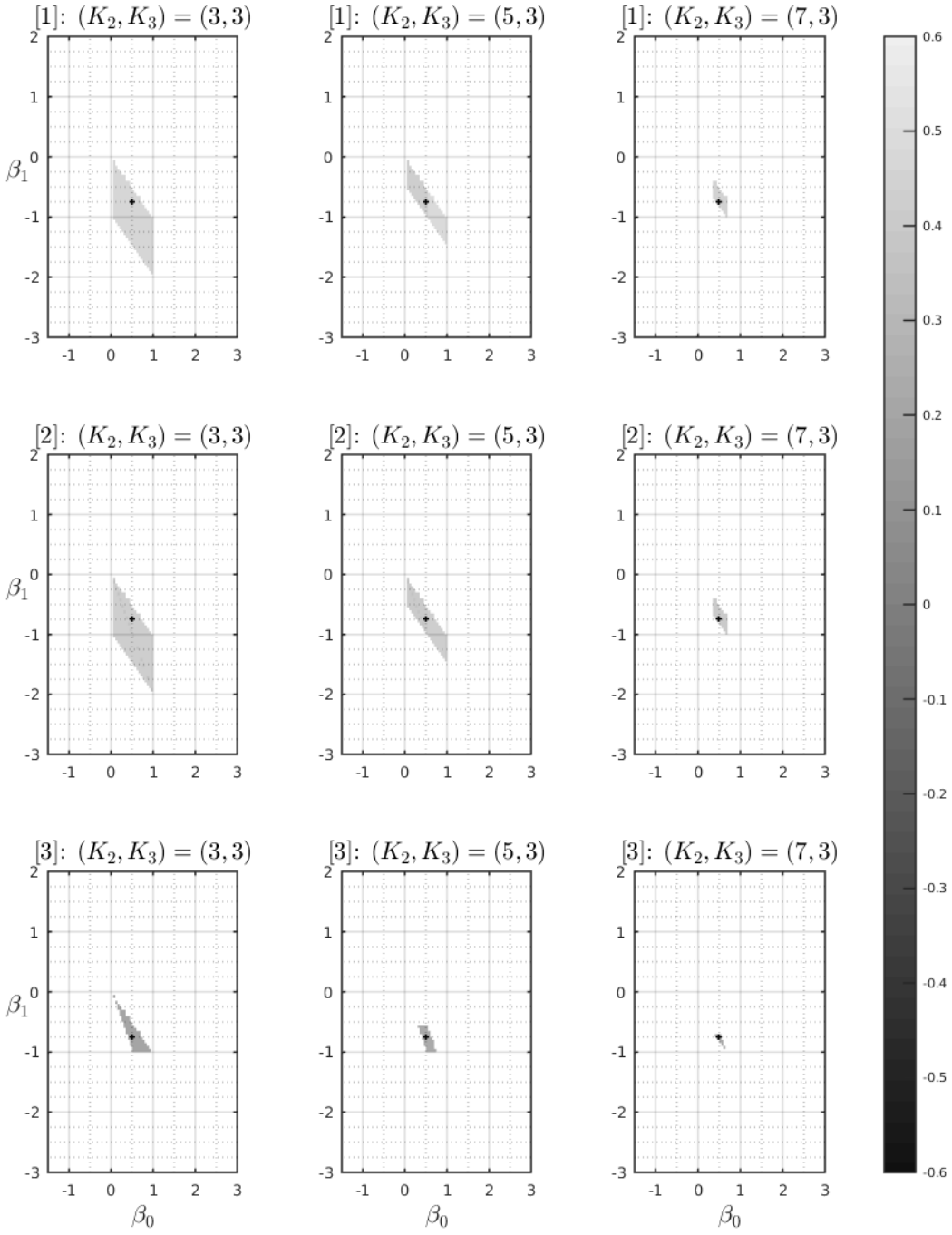

Figure 5: Sharp upper bounds for the ATE at a given fixed value of (β_0, β_1) corresponding to the sharp identified sets of (β_0, β_1) in Figure 1. Darker regions represent smaller upper bounds.

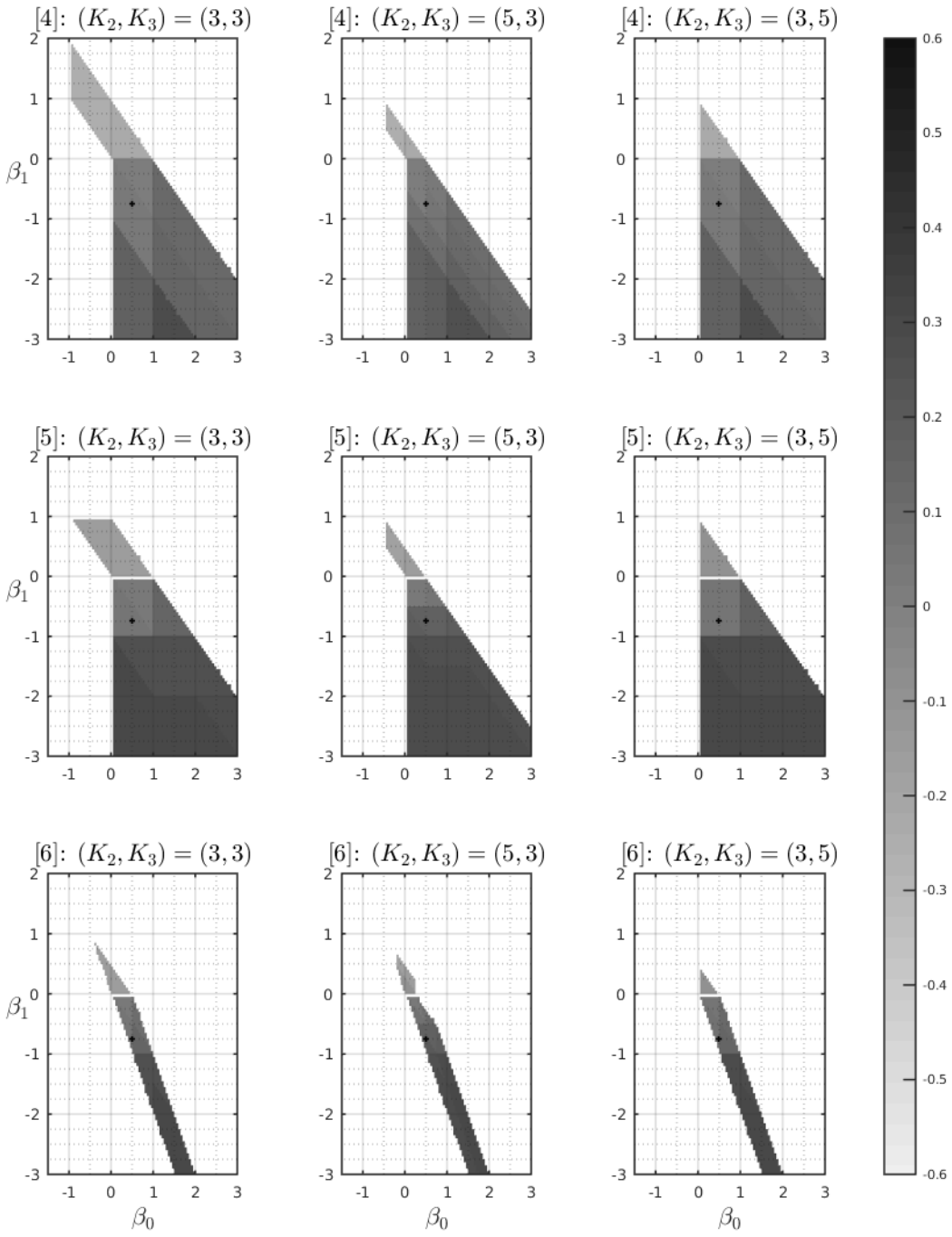

Figure 6: Sharp lower bounds for the ATE at a given fixed value of (β_0, β_1) corresponding to the sharp identified sets of (β_0, β_1) in Figure 2. Darker regions represent larger lower bounds.

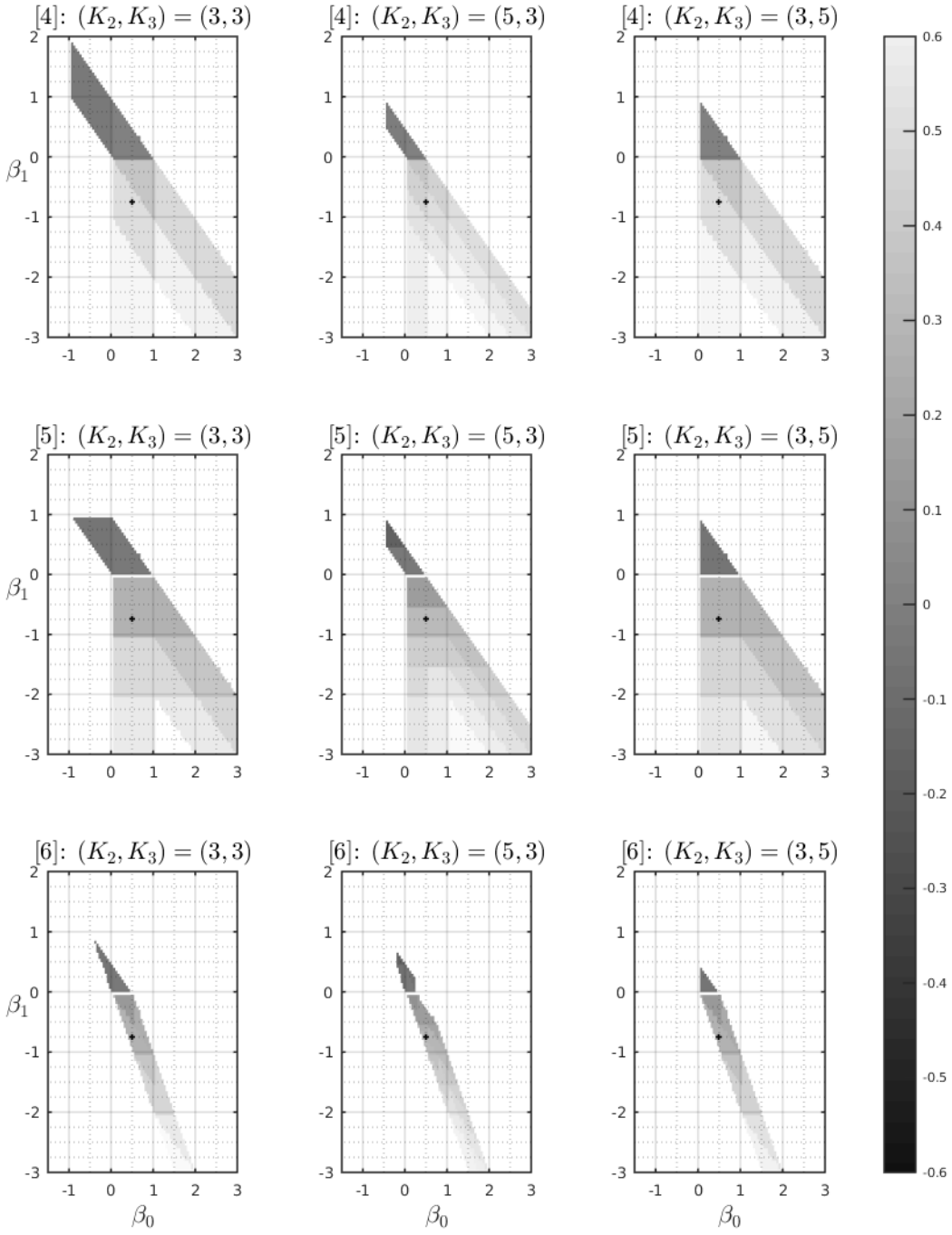

Figure 7: Sharp upper bounds for the ATE at a given fixed value of (β_0, β_1) corresponding to the sharp identified sets of (β_0, β_1) in Figure 2. Darker regions represent smaller upper bounds.

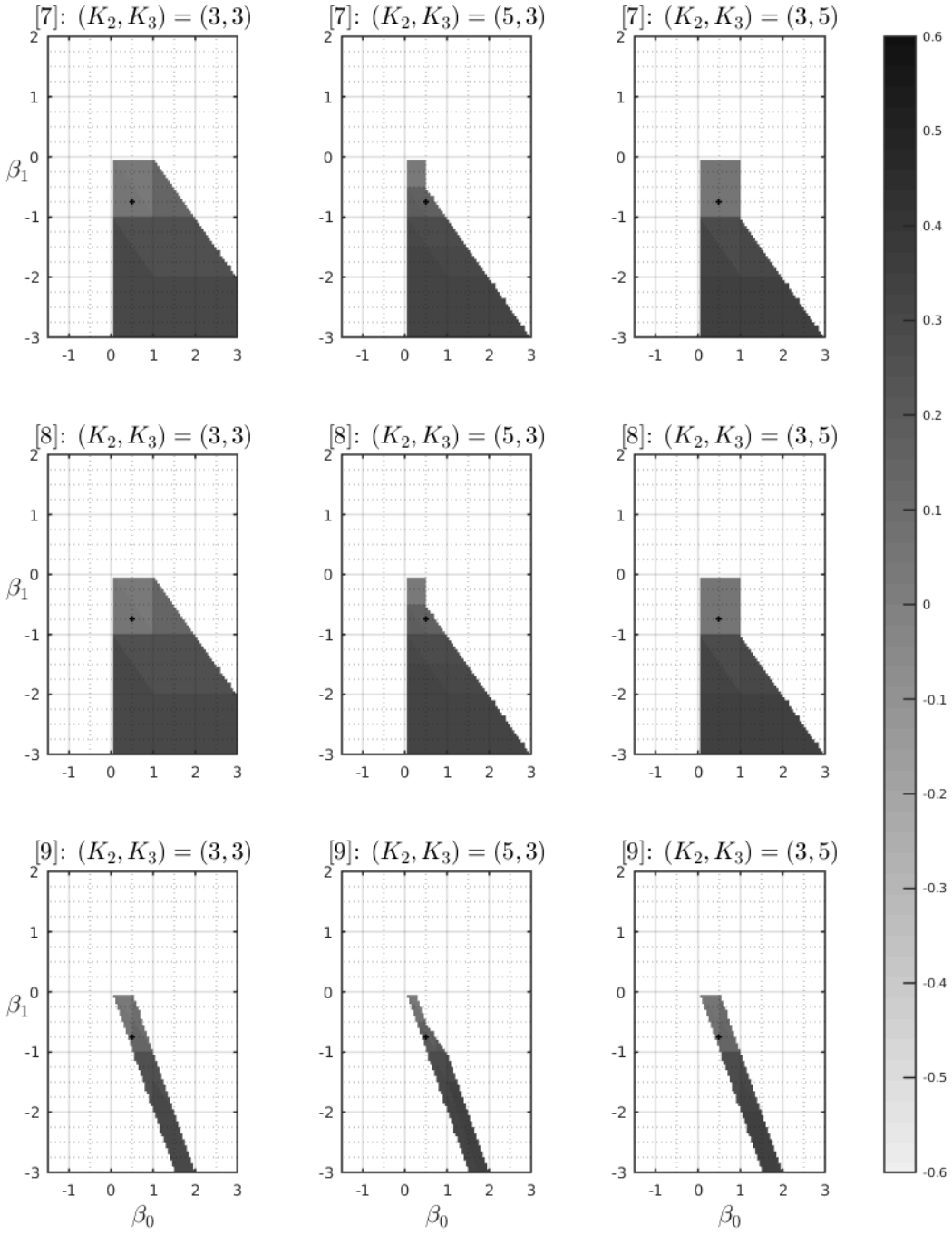

Figure 8: Sharp lower bounds for the ATE at a given fixed value of (β_0, β_1) corresponding to the sharp identified sets of (β_0, β_1) in Figure 3. Darker regions represent larger lower bounds.

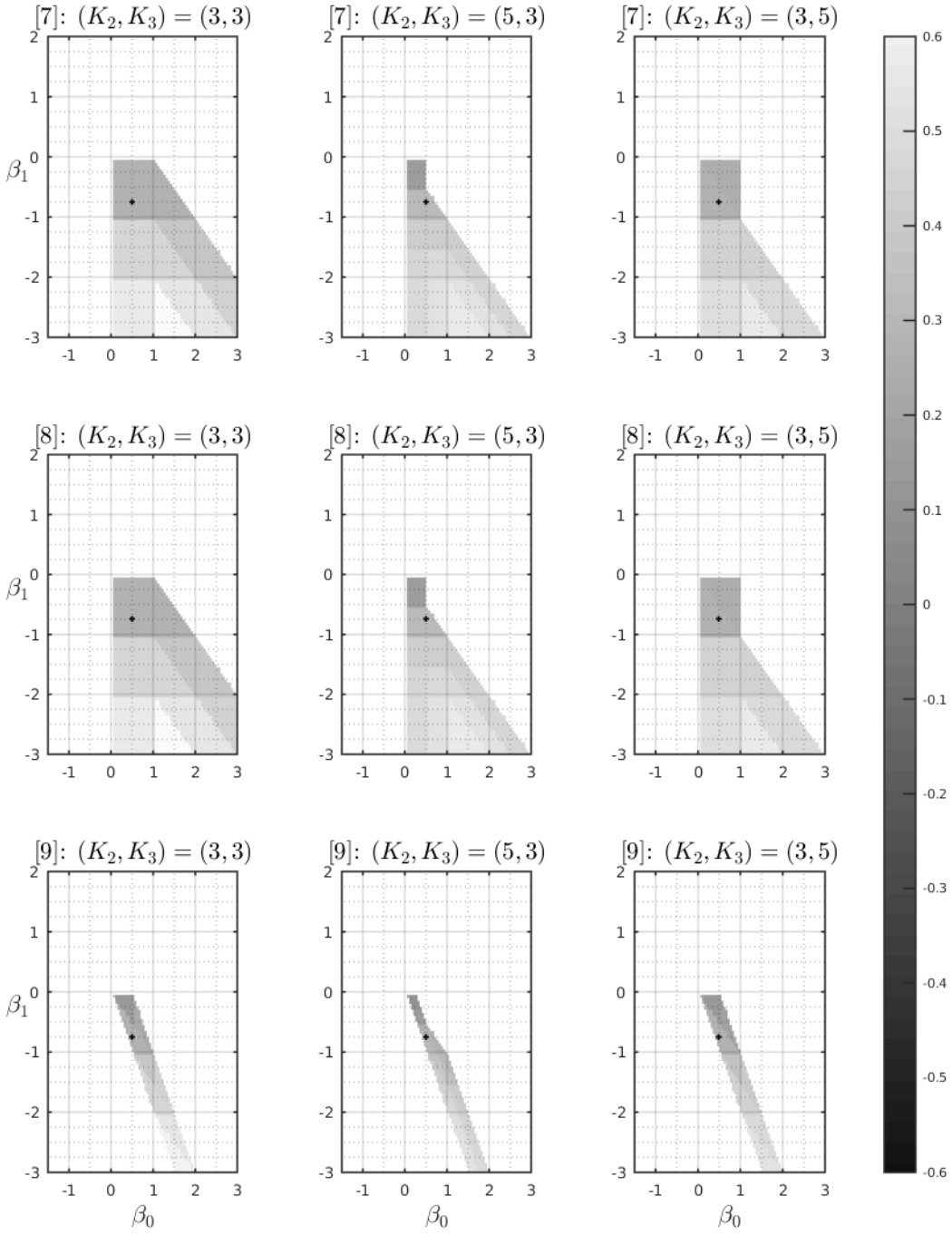

Figure 9: Sharp upper bounds for the ATE at a given fixed value of (β_0, β_1) corresponding to the sharp identified sets of (β_0, β_1) in Figure 3. Darker regions represent smaller upper bounds.

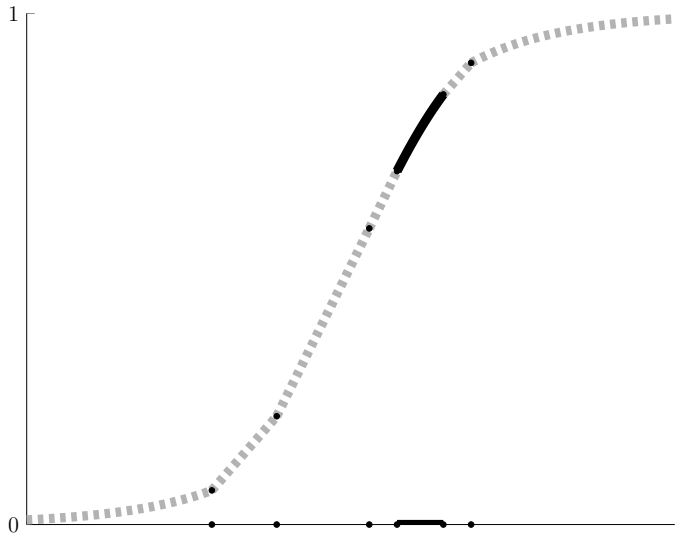

Figure 10: An example of the construction used in Lemma 3.2. The set $\overline{\mathcal{U}}$ is shown in bold on the horizontal axis and the subdistribution $\overline{F}$ is plotted in bold on its domain $\overline{\mathcal{U}}$. On $\overline{\mathcal{U}}$, F must be chosen to match $\overline{F}$. 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.

References

Andrews, D. W. K. and P. J. Barwick (2012): "Inference for Parameters Defined by Moment Inequalities: A Recommended Moment Selection Procedure," *Econometrica*, 80, 2805–2826. 37

Andrews, D. W. K. and X. Shi (2013): "Inference Based on Conditional Moment Inequalities," *Econometrica*, 81, 609–666. 5, 37

Andrews, D. W. K. and G. Soares (2010): "Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection," *Econometrica*, 78, 119–157. 37

Angrist, J. D. (2001): "Estimation of Limited Dependent Variable Models With Dummy Endogenous Regressors," *Journal of Business & Economic Statistics*, 19, 2–28. 4

Angrist, J. D. and J.-S. Pischke (2009): *Mostly Harmless Econometrics: An Empiricist's Companion*, Princeton University Press. 4, 16

Aradillas-Lopez, A. and A. Rosen (2014): "Inference in Ordered Response Games with Complete Information," *cemmap working paper CWP36/14*. 5, 42

Beresteanu, A., I. Molchanov, and F. Molinari (2011): "Sharp Identification Regions in Models With Convex Moment Predictions," *Econometrica*, 79, 1785–1821. 5, 42

Blevins, J. R. (2015): "Non-Standard Rates of Convergence of Criterion-Function-Based Set Estimators for Binary Response Models," *Econometrics Journal*, Forthcoming. 12

Blundell, R. and J. L. Powell (2003): "Endogeneity in Nonparametric and Semiparametric Regression Models," in *Advances in Economics and Econometrics*, ed. by M. Dewatripont, L. Hansen, and S. Turnovsky, Cambridge: Cambridge University Press, vol. II, 312–357. 14

Bugni, F., I. Canay, and X. Shi (2014): "Inference for Functions of Partially Identified Parameters in Moment Inequality Models," *cemmap working paper 22/14*. 37, 38

Bugni, F. A. (2010): "Bootstrap Inference in Partially Identified Models Defined by Moment Inequalities: Coverage of the Identified Set," *Econometrica*, 78, 735–753. 37

Canay, I. A. (2010): "EL inference for partially identified models: Large deviations optimality and bootstrap validity," *Journal of Econometrics*, 156, 408–425. 37

Chernozhukov, V. and C. Hansen (2005): "An IV Model of Quantile Treatment Effects," *Econometrica*, 73, 245–261. 27

Chesher, A. (2003): "Identification in Nonseparable Models," *Econometrica*, 71, 1405–1441. 32

——— (2005): "Nonparametric Identification under Discrete Variation," *Econometrica*, 73, 1525–1550. 23, 32

——— (2007): "Instrumental values," *Journal of Econometrics*, 139, 15–34. 32

——— (2010): "Instrumental Variable Models for Discrete Outcomes," *Econometrica*, 78, 575–601. 13, 22, 23, 27

——— (2013): "Semiparametric Structural Models of Binary Response: Shape Restrictions and Partial Identification," *Econometric Theory*, 29, 231–266. 13, 22, 23

Chesher, A. and A. Rosen (2012): "Simultaneous Equations Models for Discrete Outcomes: Coherence, Completeness and Identification," *CeMMAP working paper CWP21/12*. 32

——— (2014a): "An Instrumental Variable Random Coefficients Model for Binary Outcomes," *Econometrics Journal*. 5, 42

——— (2014b): "Generalized instrumental variable models," *cemmap working paper 43/13*. 5, 42, 43

Chesher, A., A. M. Rosen, and K. Smolinski (2013): "An instrumental variable model of multiple discrete choice," *Quantitative Economics*, 4, 157–196. 5, 42

Chesher, A. and K. Smolinski (2012): "IV models of ordered choice," *Journal of Econometrics*, 166, 33–48. 13, 22, 23

Chiburis, R. C. (2010): "Semiparametric bounds on treatment effects," *Journal of Econometrics*, 159, 267–275. 13, 16, 23, 26

Ciliberto, F. and E. Tamer (2009): "Market Structure and Multiple Equilibria in Airline Markets," *Econometrica*, 77, 1791–1828. 42

Eisenhauer, P., J. J. Heckman, and E. Vytlacil (2015): "The Generalized Roy Model and the Cost-Benefit Analysis of Social Programs," *Journal of Political Economy*, 123, 413–443. 4, 40

Fourer, R., D. M. Gay, and B. W. Kernighan (2002): *AMPL: A Modeling Language for Mathematical Programming*, Cengage Learning. 12

Freyberger, J. and J. Horowitz (2014): "Identification and Shape Restrictions in Nonparametric Instrumental Variables Estimation," *Working paper*. 42

Galichon, A. and M. Henry (2011): "Set Identification in Models with Multiple Equilibria," *The Review of Economic Studies*, 78, 1264–1298. 5, 42

Han, S. and E. J. Vytlacil (2015): "Identification in a Generalization of Bivariate Probit Models with Endogenous Regressors," *Working paper*. 21

Heckman, J. J. (1978): "Dummy Endogenous Variables in a Simultaneous Equation System," *Econometrica*, 46, 931–959. 21

Heckman, J. J. and B. E. Honoré (1990): "The Empirical Content of the Roy Model," *Econometrica*, 58, 1121–1149. 4

Heckman, J. J. and E. Vytlacil (2005): "Structural Equations, Treatment Effects, and Econometric Policy Evaluation," *Econometrica*, 73, 669–738. 4, 38

Hong, H. and E. Tamer (2003): "Endogenous binary choice model with median restrictions," *Economics Letters*, 80, 219–225. 22

Honoré, B. E. and A. Lleras-Muney (2006): "Bounds in Competing Risks Models and the War on Cancer," *Econometrica*, 74, 1675–1698. 42

Honoré, B. E. and E. Tamer (2006): "Bounds on Parameters in Panel Dynamic Discrete Choice Models," *Econometrica*, 74, 611–629. 42

Horowitz, J. L. (2009): *Semiparametric and Nonparametric Methods in Econometrics*, Springer. 12

IBM (2010): *IBM ILOG AMPL Version 12.2*, International Business Machines Corporation. 12

Jun, S. J., J. Pinkse, and H. Xu (2011): "Tighter bounds in triangular systems," *Journal of Econometrics*, 161, 122–128. 23

——— (2012): "Discrete endogenous variables in weakly separable models," *The Econometrics Journal*, 15, 288–303. 23

Kaido, H., F. Molinari, and J. Stoye (2015): "Inference on Projections of Identified Sets," *Working paper*. 37

Khan, S. (2013): "Distribution free estimation of heteroskedastic binary response models using Probit/Logit criterion functions," *Journal of Econometrics*, 172, 168–182. 16

Kitamura, Y. and J. Stoye (2013): "Nonparametric Analysis of Random Utility Models: Testing," *cemmap working paper CWP36/13*. 42

Komarova, T. (2013): "Binary choice models with discrete regressors: Identification and misspecification," *Journal of Econometrics*, 177, 14–33. 12

Lafférs, L. (2015): "Bounding Average Treatment Effects Using Linear Programming," *Working paper*. 42

Lehmann, E. L. (1966): "Some Concepts of Dependence," *The Annals of Mathematical Statistics*, 37, 1137–1153. 32

Lewbel, A. (2000): "Semiparametric qualitative response model estimation with unknown heteroscedasticity or instrumental variables," *Journal of Econometrics*, 97, 145–177. 28

Magnac, T. and E. Maurin (2007): "Identification and information in monotone binary models," *Journal of Econometrics*, 139, 76–104. 28

Manski, C. F. (1975): "Maximum score estimation of the stochastic utility model of choice," *Journal of Econometrics*, 3, 205–228. 1, 4, 12

——— (1985): "Semiparametric analysis of discrete response: Asymptotic properties of the maximum score estimator," *Journal of Econometrics*, 27, 313–333. 1, 4, 12

——— (1988): "Identification of Binary Response Models," *Journal of the American Statistical Association*, 83, 729–738. 1, 4, 12, 17, 22, 28

——— (2003): *Partial identification of probability distributions*, Springer. 2

——— (2007): "Partial Indentification of Counterfactual Choice Probabilities," *International Economic Review*, 48, 1393–1410. 42

——— (2014): "Identification of incomeleisure preferences and evaluation of income tax policy," *Quantitative Economics*, 5, 145–174. 42

Matzkin, R. L. (1994): "Chapter 42 Restrictions of economic theory in nonparametric methods," in *Handbook of Econometrics*, ed. by R. F. Engle and D. L. McFadden, Elsevier, vol. Volume 4, 2523–2558. 22

——— (2003): "Nonparametric Estimation of Nonadditive Random Functions," *Econometrica*, 71, 1339–1375. 22, 27

——— (2007): "Chapter 73 Nonparametric identification," in *Handbook of Econometrics*, ed. by J. J. Heckman and E. E. Leamer, Elsevier, vol. Volume 6, Part 2, 5307–5368. 22

Molinari, F. (2008): "Partial identification of probability distributions with misclassified data," *Journal of Econometrics*, 144, 81–117. 42

Mourifié, I. (2015): "Sharp bounds on treatment effects in a binary triangular system," *Journal of Econometrics*, 187, 74–81. 23, 26

Mourifié, I., M. Henry, and R. Méango (2015): "Sharp Bounds For The Roy Model," *SSRN Working Paper 2043117*. 5, 39

Nelsen, R. (2006): *An introduction to copulas*, New York: Springer. 24, 45, 46, 47, 49

Pakes, A. (2010): "Alternative Models for Moment Inequalities," *Econometrica*, 78, 1783–1822. 42

Romano, J. P., A. M. Shaikh, and M. Wolf (2014): "A Practical Two-Step Method for Testing Moment Inequalities," *Econometrica*, 82, 1979–2002. 37

Rudin, W. (1976): *Principles of mathematical analysis*, New York: McGraw-Hill. 36

Schennach, S. M. (2014): "Entropic Latent Variable Integration via Simulation," *Econometrica*, 82, 345–385. 5, 42, 43

Schweizer, B. and A. Sklar (1983): *Probabilistic Metric Spaces*, North-Holland, New York. 24, 25, 26, 45, 46, 47

Shaikh, A. M. and E. J. Vytlacil (2011): "Partial Identification in Triangular Systems of Equations With Binary Dependent Variables," *Econometrica*, 79, 949–955. 23

Sklar, A. (1959): "Fonctions de répartition à n dimensions et leurs marges," *Publ. Inst. Statist. Univ. Paris*, 8, 1. 2, 8, 41, 45

——— (1996): "Random Variables, Distribution Functions, and Copulas: A Personal Look Backward and Forward," *Institute of Mathematical Statistics Lecture Notes-Monograph Series*, 28, 1–14. 2, 41, 46

Torgovitsky, A. (2015): "Partial Identification of State Dependence," *SSRN Working Paper 2564305*. 38

Vytlacil, E. (2002): "Independence, Monotonicity, and Latent Index Models: An Equivalence Result," *Econometrica*, 70, 331–341. 38

Vytlacil, E. and N. Yildiz (2007): "Dummy Endogenous Variables in Weakly Separable Models," *Econometrica*, 75, 757–779. 23

Wooldridge, J. M. (2010): *Econometric analysis of cross section and panel data*, MIT press. 4, 17

Yildiz, N. (2013): "Estimation Of Binary Choice Models With Linear Index And Dummy Endogenous Variables," *Econometric Theory*, 29, 354–392. 23