Generalized Instrumental Variable Models, Methods, and Applications[*]

Andrew Chesher
University College London

Adam M. Rosen
Duke University

August 16, 2019

Abstract

This chapter sets out the extension of the scope of the classical IV model to cases in which unobserved variables are set-valued functions of observed variables. The resulting Generalized IV (GIV) models can be used when outcomes are discrete while unobserved variables are continuous, when there are rich specifications of heterogeneity as in random coefficient models, and when there are inequality restrictions constraining observed outcomes and unobserved variables. There are many other applications and classical IV models arise as a special case. The chapter provides characterizations of the identified sets delivered by GIV models. It gives details of the application of GIV analysis to models with an interval censored endogenous variable and to binary outcome models – for example probit models – with endogenous explanatory variables. It illustrates how the identified sets delivered by GIV models can be represented by moment inequality characterizations that have been the focus of recently developed methods for inference. An empirical application to a binary outcome model of female labor force participation is worked through in detail.

Contents

1 Introduction

1.1 Scope and Motivation

Since the earliest days of econometrics, instrumental variable (IV) restrictions have been a key element in structural models of economic processes.[1] There are two standout reasons for the ubiquity of IV models in econometrics. First, IV restrictions that limit the covariation of exogenous observable variables and unobservable variables can often be convincingly motivated by consideration of the economics of the process being studied. Second, IV models can be incomplete, in that they need not fully specify the determination of endogenous variables as a single-valued function of exogenous observable and unobservable variables. Incompleteness is a great advantage in modelling complex processes when there is limited understanding of some aspects of the process and a desire for robustness to misspecification of such elements.

    Until the early 2000's the scope of application of incomplete IV models in structural econometrics was limited to cases in which unobservable variables can be expressed as single-valued functions of observable variables. This ruled out the use of IV models in many applications arising in modern econometric analysis – including cases with discrete outcomes, high dimensional heterogeneity, and structural restrictions involving inequalities – unless the researcher was willing to assert a complete specification for the determination of all endogenous variables. This limitation no longer applies. The aim of this chapter is to set out the wide range of structural econometric models to which IV methods can now be applied, to show how identified sets of structures and structural features can be characterized and to discuss the implications for econometric practice.

---

[1]The first record of the IV estimator is in Wright (1928).

We start by juxtaposing the familiar linear IV model for a single continuous outcome with an IV model for the binary outcome obtained when only the sign of the continuous outcome is reported. This is a simple example of an IV model in which the unobservable is a set-valued function of observable variables. This is a case in which an incomplete IV model is set identifying, but not point identifying even if strong parametric restrictions are brought on board.

To illustrate, consider the single equation linear IV model

$$Y_1 = Y_2\beta + Z_1\gamma + U, \tag{1}$$

where scalar Y_1 and Y_2 are endogenous variables, jointly determined in the process being studied, variables Y_1, Y_2, Z_1, Z_2 are observable, where Y_2, Z_1, and Z_2 are row vectors, and U is unobservable. Z_2 are excluded instruments and the dependence of exogenous variables $Z = (Z_1, Z_2)$ and unobservable U is restricted, for example requiring $E[Z'U] = 0$. This model is incomplete because there is no restriction on the determination of the endogenous variable Y_2, so there is a set of values of Y_1 and Y_2 that may eventuate given realizations of exogenous observable variables Z and unobservable U. However, unobservable U is a unique-valued function of observable variables,

$$U = Y_1 - Y_2\beta - Z_1\gamma$$

and with an independence restriction there may be point identification of the parameter values. For example under the restriction $E[Z'U] = 0$ there is point identification if there is a unique value of (β, γ) such that

$$E[Z'(Y_1 - Y_2\beta - Z_1\gamma)] = 0.$$

A classical rank condition requiring $E[Z'(Y_2, Z_1)]$ to have full column rank suffices for point identification.

Now consider the case in which only the sign of the outcome Y_1 is observed so that Y_1 is binary, as for example in a probit model:

$$Y_1 = 1[Y_2\beta + Z_1\gamma + U > 0]. \tag{2}$$

As before Y_2 is endogenous and there are observable excluded exogenous variables Z_2. Now unobserved U is not a single-valued function of observable variables. Knowledge of the value of the observable variables only restricts U to one of two sets of possible values, namely those that lie either above or below the threshold $-Y_2\beta - Z_1\gamma$, depending on whether Y_1 takes the value one or zero. Now, even if U is restricted to be normally distributed with unit variance, and fully independent of Z – a much stronger restriction than $E[Z'U] = 0$ – the parameters are generally not point identified. The problem here is that with Y_1 discretely varying, the force of the parametric distributional restriction and the independence restriction on the distribution of U and Z is not

4

sufficiently strong to achieve point identification of the parameter values, even if the classical linear model rank condition holds. However, these restrictions can rule out certain parameter values. The IV probit model is *set* identifying.[2]

The topic of this chapter is the extension of the IV approach to this and other, more complex cases. One significant feature of these Generalized IV (GIV) models is that they are generically partially identifying with identified sets characterized by systems of moment inequalities. A case in point is the class of unordered multiple discrete choice models studied in Chesher, Rosen, and Smolinski (2013), (CRS13). The models studied are random utility models such as that of McFadden (1974), but where some of the observable variables that affect utility are allowed to be endogenous. No restriction is placed on the covariation of these endogenous utility shifters with unobservable utility shifters, which have a Type I Extreme Value distribution in the conditional logit version of the model, and no specification of a function determining the values of endogenous utility shifters is made. A vector of instrumental variables Z independent of U is assumed available, and the model is thus another example of a GIV model.

A key insight of CRS13 is the recognition that this sort of incomplete IV model shares a similar mathematical structure to previously studied models in which incompleteness arises for an entirely different reason, namely the possibility of multiple equilibria. Examples of papers that provide set identification analysis in models with multiple equilibria include Tamer (2003), Ciliberto and Tamer (2009), Beresteanu, Molchanov, and Molinari (2011), and Galichon and Henry (2011). As in these models, the incomplete unordered discrete choice IV models of CRS13 are ones in which the mapping from structural unobservables to endogenous variables is many-to-many, and random set theory can be employed for identification analysis, as it was for example in Beresteanu, Molchanov, and Molinari (2011). For the model studied in CRS13 it was however found convenient to work with random sets defined in the space of unobservable heterogeneity, as this affords a tractable way to incorporate the restriction that the structural unobservables U and instrument Z are independent.

Chesher and Rosen (2017a) extends this analysis to a much broader class of structural econometric models, in which a variety of restrictions may be imposed on the joint distribution of structural unobservables and instrumental variables. This chapter shows how to exploit IV restrictions in identification analysis of structural, possibly incomplete, econometric models in which unobservable heterogeneity may not be expressible as a single-valued function of observable variables. The purpose of this chapter is to lay out an orderly and user-friendly approach to using IV restrictions in such settings.

This extension is made operational by employing a systematic approach to the consideration of the *sets* of values of the structural model's unobservable variables that could have feasibly

---

[2]Stata (StataCorp (2017)) provides an "ivprobit" command but the command name is a misnomer. It does not provide estimates of the parameters of a probit IV model. Instead it provides estimates of the parameters of a parametric Gaussian triangular, "control function" model with a probit main equation and linear equations for the endogenous explanatory variables.

generated the realizations of the observable variables. The main idea behind this development is that a given set of primitives of the structural model could only have produced the distribution of observable data if there exists a compatible unobserved random vector that, with probability one, is an element of this set of values of the unobservable variables. The set of primitives for which there is such a random vector characterizes the identified set of model primitives, that is, *structures* in the terminology which will be defined shortly.

This is the path taken in the analysis of Chesher and Rosen (2017a) (CR17), formalized using random set theory to model stochastic variation in sets of values of unobservable variables. Our goal in this chapter is to illuminate the key concepts of that analysis and make plain its wide applicability in IV models. This approach enables consideration of a variety of IV restrictions that limit the joint dependence of unobservables and exogenous variables.

Extension of the scope of incomplete IV models is needed because the alternatives chosen in practice are unattractive in many cases and may lay researchers open to delivering results which have spurious accuracy. We find two main alternatives to incomplete IV models employed in practice.

One alternative involves constructing complete models. This is often done with the aim of achieving point identification of policy-relevant parameters. The difficulty here is that there are typically many alternative complete models amongst which to choose, each one potentially giving a different identifying correspondence, and hence estimator for a structural feature of interest. One incomplete GIV model may encompass all these complete models and deliver an identified set of values of a structural feature comprising all the values of the feature that are point or set identified by all the various possible completions of it. The analysis described in this chapter provides the tools to analyze that encompassing incomplete model.

Another alternative is to assume that unobserved variables and endogenous variables are independent to some suitable degree *conditional* on certain observed variables or on some functions of them, sometimes known as control functions. Models containing this restriction can be point identifying. Control functions may be estimated in a first stage analysis and the estimates used as conditioning variables in a second stage. Precisely how the estimated conditioning variables are used depends on what additional assumptions are made.[3] The control function approach can be motivated by a triangular model. The recursive determination of the values of endogenous variables in triangular models guarantees completeness but triangular structures are not suitable when there is simultaneous determination of outcomes.

The conditional independence restriction underlying the control function approach requires us to think of covariation of endogenous variables and unobservables as arising because they are both

---

[3]Heckman and Robb (1985) present an early example of a control function estimator for use in program evaluation, and other examples beyond the classical linear system framework include those studied by Newey, Powell, and Vella (1999), Blundell and Powell (2003), Chesher (2003), Lee (2007), Florens, Heckman, Meghir, and Vytlacil (2008), Imbens and Newey (2009), Torgovitsky (2015), and D'Haultfoeuille and Fevrier (2015), among others.

affected by variation in some observable magnitudes. This generally does not apply in nonlinear processes in which there is genuine simultaneous determination of outcomes and the formal conditions that justify a control function approach when endogenous variables are genuinely simultaneously determined are not easily satisfied outside the context of linear models with additive errors.[4]

Many times, econometric models built on a firm foundation of plausible restrictions will be partially identifying for structural features of interest. The GIV models studied here are a case in point.

Partially identifying models deliver a realistic view of the information content of economic theory and economic data. They offer a route to honest reporting of information about economic magnitudes. Their use gives the ability to compare and contrast the identifying power of different models employing more or less severe restrictions, without requiring each such model to invoke a sufficiently rich set of restrictions to guarantee point identification. They make a contribution to achieving the goal of taking the "con" out of econometrics argued for in Leamer (1983). Point identifying, restriction heavy alternatives are prey to Manski's (2003) law of decreasing credibility.[5]

The developments reported on in this chapter provide a framework for identification analysis in the wide class of cases in which instrumental variable restrictions are imposed. The GIV models studied in this chapter deliver identified sets of structures that have a straightforward characterization as systems of moment inequalities. Their production can become routine. The general characterization is, as we show, easy to specialize to particular cases and it always delivers sharp identified sets, dispensing with the need for case-by-case proofs of sharpness. Recent developments in econometric inference in partially identifying models open the door to the possibility of routine implementation.

At this point in the history of econometrics working with partially identifying models is not routine, and there are many case-by-case treatments of different problems. There is an extensive literature on the topic of partial identification, and there are other approaches to partial identification analysis, including the complementary general treatments for incomplete models (Galichon and Henry (2011)) and models with convex moment predictions (Beresteanu, Molchanov, and Molinari (2011)). The chapter Molinari (2019) in this volume covers the broader literature on the use of partial identification in econometrics, and we refer to that chapter and the book Molchanov and Molinari (2018) for a more extensive overview of the literature on partial identification and the use of random sets in econometrics. For a survey of applications of partially identifying models with an emphasis on common themes across different areas see Ho and Rosen (2017).

Our focus in this chapter is on GIV models, which are instrumental variable models that may be incomplete and for which there may not exist a unique inverse of the mapping from observed

---

[4]See Blundell and Matzkin (2014) for the required *control function separability* conditions.

[5]The credibility of inference decreases with the strength of assumptions maintained.

endogenous and exogenous variables to unobservable variables. These are structural models that impose restrictions on a mapping from unobserved heterogeneity to observable outcomes. An alternative framework for application of IV methods is the program evaluation approach, which takes randomized experiments with a potential outcome representation as its starting point. In this chapter we focus on the structural approach, but the program evaluation approach is a valuable complementary approach in the modern econometrician's toolkit. Our approach can be used to conduct partial identification analysis in program evaluation models and in extensions of those models in which structural econometric restrictions appear, as in the models studied in Chesher and Rosen (2015). For further discussion and application of instrumental variables using the program evaluation approach, there are several excellent sources available, including Angrist, Imbens, and Rubin (1996), Angrist and Pischke (2009) and Imbens (2014), as well as Heckman (2010) for ways in which the structural and program evaluation approach may be combined.

There are other distinct areas of research on IV models which this chapter does not cover, but which are extensively covered elsewhere. One such area is the impact of weak instruments on the performance of inference procedures, for which Stock, Yogo, and Wright (2002) and Andrews and Stock (2007) provide instructive surveys. Another related topic is measurement error and econometric methods to address it, which sometimes involve the use of instruments. For an overview of this literature, see for example Chen, Hong, and Nekipelov (2011) and Schennach (2016), as well as the chapter Schennach (2019) in this volume. Horowitz (2011) provides an accessible overview of nonparametric instrumental variables estimation, focusing on the main ideas and motivation for their use and providing a thorough set of references to the broader literature on that topic.

1.2   Plan of the Chapter

This chapter is laid out as follows. In Section 2 the scope of analysis is cast in the light of the foundational framework of structural models studied by the Cowles Commission during its time at the University of Chicago from 1939 to 1955. As described in this section, the class of structural models studied here is a natural extension of their early framework, suitably expanded to accommodate incomplete models, set-valued unobservables, and partial identification.

In Section 3 a wide class of Generalized Instrumental Variable (GIV) models is defined. We give further examples which include some classical IV models and complete as well as incomplete models. Some of the models that fall in the GIV class do not involve IVs at all. A structure denoted $(h, \mathcal{G}_{U|Z})$ is defined to have two coupled elements: a structural function $h(Y, Z, U)$ which defines the combinations of endogenous Y, observed Z, and unobserved U admitted by a model *via* the restriction $\mathbb{P}[h(Y, Z, U) = 0] = 1$ and a collection of conditional distributions of U given Z. For instance, in the linear IV model of equation (1), $Y_1 = Y_2\beta + Z_1\gamma + U$, and the structural function h may be defined $h(Y, Z, U) = Y_1 - Y_2\beta - Z_1\gamma - U$, since then $h(Y, Z, U) = 0$ is identical to (1). Moreover, the restriction $E[Z'U] = 0$ considered earlier defines a set of admissible collections

8

of conditional distributions of U given Z, namely any collection such that $E[Z'U] = 0$ holds. If the model instead imposes $E[U|Z] = 0$, this defines an alternative (and smaller) set of admissible collections of conditional distributions of U given Z.

The identified set of structures delivered by a model and a probability distribution of observables comprises the structures admitted by the model which are observationally equivalent to each other and which can deliver the probability distribution of observables under consideration. Identified sets of structural features are obtained by projection.

Section 4 is concerned with observational equivalence of structures. Standard definitions of observational equivalence of structures apply when models are complete. To carry this definition through to incomplete models requires extension of the definition of observational equivalence, as set out in Chesher and Rosen (2017a) and expounded here. The definition is naturally expressed in terms of conditional distributions of sets of *observable* outcomes $\mathcal{Y}(U, Z; h)$ where

$$\mathcal{Y}(U, Z; h) \equiv \{y : h(y, Z, U) = 0\}$$

is a level set of the structural function $h(\cdot, \cdot, \cdot)$. We show how it can be equivalently defined using the conditional distributions of sets of *unobservable* variables $\mathcal{U}(Y, Z; h)$ where

$$\mathcal{U}(Y, Z; h) \equiv \{u : h(Y, Z, u) = 0\}$$

is another level set of the structural function $h(\cdot, \cdot, \cdot)$ that is dual to the level set $\mathcal{Y}(U, Z; h)$. Expression of observational equivalence in these terms is most convenient when considering the impact of IV restrictions on the conditional distributions of U given Z. It is shown in Section 4 how these alternative expressions of observational equivalence lead to alternative characterizations of identified sets of structures based on properties of the dual level sets $\mathcal{Y}(U, Z; h)$ and $\mathcal{U}(Y, Z; h)$.

Section 5 sets out observable implications of these characterizations and discusses their use in practice. One of the characterizations involves consideration of probability inequalities, each one associated with a set of values of the unobservable variables. The important concept of core-determining sets is explained. Using properties of core determining sets a particular collection of observable implications can be guaranteed to produce the full characterization of the identified set, as given in the previous section. This can lead to substantial reduction in the number of inequalities required to characterize an identified set.

Section 6 provides characterizations of identified sets of GIV models under a variety of different restrictions on the conditional distributions of unobservables common in the econometrics literature. These include independence restrictions as well as conditional mean and conditional quantile independence restrictions.

Sections 7 and 8 demonstrate application of the chapter's analysis to IV models with censored endogenous variables, and binary outcomes. The models studied in Section 7 offer an IV gen-

eralization of models studied in Manski and Tamer (2002), in which censored variables may be endogenous explanatory variables. A range of restrictions on the joint dependence of the structural unobservable variable and the exogenous variables are considered – namely independence restrictions, conditional quantile restrictions, and conditional mean restrictions. The moment inequality characterizations of identified sets that result from application of earlier analysis in the chapter are demonstrated and compared.

Section 8 provides an application to IV models with a binary outcome, including one employing the probit specification (2) discussed above. Identification analysis is applied to characterize identified sets in such models and analog set estimates are reported in an application to female labor force participation using data from Angrist and Evans (1998). In the context of this application, it is demonstrated how inference can be carried out, with confidence sets for parameters and certain projections of them computed using methods developed in Chernozhukov, Lee, and Rosen (2013).

Section 9 discusses developing strands of research in the literature and concludes.

2   Structural IV Models

The framework employed in this chapter builds on the seminal work done at the Cowles Commission during its tenure at the University of Chicago from 1939 to 1955. In this period a formal apparatus for identification analysis was laid out in a collection of papers appearing in the Cowles Commission for Research in Economics Monograph 10, Koopmans (1950). The monograph, referred to here as Cowles 10, was based on presentations given at a Cowles Commission conference in 1945. This work was notably influenced by ideas set out in Haavelmo (1943, 1944) and the framework employed for studying identification is elegantly summarized by Koopmans (1949) building on ideas set out by Leonid Hurwicz in his 1945 conference paper, Hurwicz (1950).

Marschak (1950) referred to the system of equations characterizing economic relationships in the models studied as the "stochastic model", by which was meant a model whose defining relationships feature random variables that are unobservable to the econometrician. In today's terminology, such a model comprising a system of equations including unobservable variables would be described as a "structural model". This is the framework in which the ideas set out in this chapter are exposited.

2.1   The Cowles 10 Stochastic Model

On page 19 of the introductory chapter of Cowles 10, Marschak set out the stochastic model in which the contributions of the monograph are cast. In this section only we use Marschak's notation. Unobservable random disturbances, denoted $w = (w_1, ..., w_J)$, are explicitly incorporated. The model constitutes *a priori* restrictions on a system of G equations

$$\varphi_g\left(x, w; \alpha_{(g)}\right) = 0, \qquad g = 1, ..., G, \tag{3}$$

and on the joint density function of w

$$f(w; \varepsilon), \qquad (4)$$

where ε and $\alpha = \left(\alpha_{(1)}, ..., \alpha_{(G)}\right)$ denote model parameters. As explained in Koopmans (1949), a model may be represented by a collection of admissible structures. A *structure* is defined as all properties of the system of equations (3) and the distribution (4), including those not known *a priori*. Observable variables x are grouped into a vector of endogenous components y and exogenous components z. Independence of w and z is presumed so that the marginal distribution $f(w; \varepsilon)$ coincides with the conditional distribution of w given z.

As Marschak points out, the monograph focuses almost exclusively on the special case described as a simple shock model, in which $w = (u_1, ..., u_G)$, with (3) replaced by

$$\varphi_g \left(x; \alpha_{(g)}\right) = u_g, \qquad g = 1, ..., G. \qquad (5)$$

These are models in which for any value of x knowledge of the functions $\varphi = (\varphi_1, ..., \varphi_G)$ and parameters α delivers a unique value of the unobservable variables, w.

Notably, the papers in the monograph focus on models that are complete, with the sole exception of Wald (1950). Wald's six page chapter raises issues concerning estimation and inference in incomplete models that came to be studied in depth only much later. Wald framed his discussion around an incomplete model comprising a system of equations as in (5), in which the distribution of unobservable variables was insufficient to determine that of the endogenous variables. Wald writes on page 306, "It will appear from what follows that the estimation problems in incomplete systems are essentially different from those in complete systems discussed in other contributions in this volume." Wald later continued on page 306, running onto page 307, "...we cannot be sure that a consistent estimate exists. This difficulty, however, is not as serious as it would appear at first sight. In fact, instead of point estimates, we are usually more interested in constructing a confidence region for the unknown parameters corresponding to a given confidence coefficient." He thus sidestepped the issue of identification, and proceeded to examine the question of inference irrespective of point identification in a particular example. It would be more that 50 years until Imbens and Manski (2004) took up the problem of constructing confidence intervals for a set identified parameter, the first of many studies of inference on set identified parameters in the recent partial identification literature.[6]

---

[6]It should be noted that there were earlier (and later) studies in which confidence regions were proposed for *identified sets* of parameters rather than the parameters themselves, notable predecessors being the Bonferroni and bootstrap procedures proposed by Horowitz and Manski (1998, 2000). Section 4.3 of the chapter Molinari (2019) provides a discussion of different coverage notions considered in the literature.

2.2 The Framework Employed Here

The IV models and methods set out in this chapter are presented in a framework that expands upon the Cowles 10 stochastic model by permitting structures to be *incomplete* and to have unobservable variables that may *not* be single-valued functions of observable variables. As discussed in the introduction, this widens the scope of application of IV restrictions to a very broad class of models. The notation used from here onwards reflects common current practice. Greek letters are used to denote finite dimensional parameters. When considering stochastic quantities upper and lower case Roman letters are used to denote random vectors and their realizations, respectively. We reserve calligraphic font ($\mathcal{R}$) to denote sets and set-valued random elements and sans serif font (R) to denote collections of sets.

The processes considered here deliver values of observed endogenous outcomes Y given values of observed Z and unobserved U, with Y, Z, and U all finite dimensional random vectors with support on a subset of a Euclidean space.

$$
\begin{aligned}
Y &\equiv (Y_1, \ldots, Y_N), \\
Z &\equiv (Z_1, \ldots, Z_K), \\
U &\equiv (U_1, \ldots, U_R).
\end{aligned}
$$

There may be discrete components in any of these vectors.

A *structure* comprises two elements, analogous to (3) and (4). One element is a collection of conditional distributions of U given $Z = z$, one distribution for every value z in the support, $\mathcal{R}_Z$, of Z . This is given by a collection of conditional probability measures denoted $\mathcal{G}_{U|Z}$ defined formally in Restriction A3 in Section 3 below. The other element comprises a specification of the combinations of Y, Z and U that can occur, given by the specification of a structural function $h : \mathcal{R}_{YZU} \to \mathbb{R}$ such that

$$h(Y, Z, U) = 0, \tag{6}$$

with probability one.[7] This relation plays the role of the system of equations expressed in (3). That there is only one equation rather than G equations is unimportant, because, reverting momentarily to Marschak's notation, letting $U = w$ and $(Y, Z) = x$, any system (3) can be expressed as

$$h(Y, Z, U) = \max_{g \in \{1, \ldots, G\}} \left| \varphi_g \left(x, w; \alpha_{(g)} \right) \right| = 0.$$

There are many other ways in which this can be done, for example using the sum of squares of the functions $\varphi_g \left(x, w; \alpha_{(g)} \right)$.

As in the treatment given in Cowles 10, models place restrictions on the structures m that

---

[7]Throughout this Chapter $\mathbb{R}$ denotes the real line.

generate outcomes, with $m = \left(h, \mathcal{G}_{U|Z}\right)$ in the notation of this Chapter. A model $\mathcal{M}$ is defined as the set of structures that obey a particular collection of restrictions. For example, considering an IV model $\mathcal{M}$ could require the structural function to be linear with coefficient zero on certain excluded exogenous variables and with unobserved variables and exogenous variables independently distributed.

Structures $m \in \mathcal{M}$ obeying the model's restrictions are referred to as *admissible*. The goal of identification analysis is to ascertain which admissible structures are capable of producing the conditional distributions of observable variables Y given exogenous variables Z that a process delivers.

For any set $\mathcal{Y} \subseteq \mathcal{R}_Y$, the notation $F_{Y|Z}(\mathcal{Y}|z)$ denotes the probability of the event $Y \in \mathcal{Y}$ conditional on $Z = z$. The sampling process is assumed to reveal the conditional probability measure $F_{Y|Z}(\cdot|z)$ for each $z \in \mathcal{R}_Z$ with

$$\mathcal{F}_{Y|Z} \equiv \{F_{Y|Z}(\cdot|z) : z \in \mathcal{R}_Z\}$$

denoting this collection of conditional measures.

The identified set of structures delivered by a model $\mathcal{M}$ and a collection of probability measures $\mathcal{F}_{Y|Z}$ is denoted $\mathcal{I}\left(\mathcal{M}, \mathcal{F}_{Y|Z}, \mathcal{R}_Z\right)$. This notation makes explicit the dependence of the identified set on (i) restrictions embedded in specification of $\mathcal{M}$, (ii) the conditional distributions of Y given Z, and (iii) the support of exogenous variables Z, over which conditional distributions $F_{Y|Z}$ are point identified.

In contrast to the models studied in Cowles 10, the models considered here will admit structures with structural functions h for which equation (6) may have multiple solutions for Y for some realizations of (Z, U), and may have multiple solutions for U for some realizations of (Y, Z). Structures $\left(h, \mathcal{G}_{U|Z}\right)$ with h allowing multiple solutions for Y for some realizations of (Z, U) are incomplete. Wald's six-page chapter Wald (1950) makes clear that researchers were aware of the possibility of incompleteness, and that it would require an alternative treatment. The issue of incompleteness and the difficulties it can cause for identification have been more recently studied in the context of models that allow for multiple equilibria, with some key contributions including Heckman (1978), Jovanovic (1989), Tamer (2003), Galichon and Henry (2011), and Beresteanu, Molchanov, and Molinari (2011). Incompleteness also arises in single equation IV models that do not include a full specification of the determination of endogenous variables and in models that require observed and unobserved variables to satisfy inequality restrictions.

The focus of Cowles 10 was on complete models with systems of simultaneous equations in which the values of endogenous variables are fully determined as part of the system. Moreover, Cowles 10 primarily addressed linear systems, in which the conditions required for point identification of the parameters of each equation do not require a complete specification of the determination of all

endogenous variables. As a simple case in point, in the classical two equation model

$$Y_1 = Y_2\beta_1 + Z_1\gamma_1 + U_1, \tag{7}$$

$$Y_2 = Y_1\beta_2 + Z_2\gamma_2 + U_2, \tag{8}$$

with $E\left[U_1|Z\right] = E\left[U_2|Z\right] = 0$, the rank condition for point identification of (β_1, γ_1) is that $E\left[Z'\left(Y_2, Z_1\right)\right]$ has full column rank. This is the same condition required for point identification of (β_1, γ_1) in an incomplete single equation IV model that imposes only (7), with (8) absent.

In nonlinear models this simple equivalence does not apply. One can augment the equations of an incomplete model with additional equations such that the augmented system uniquely determines the values of all endogenous variables. However, if any restrictions are placed on the additional equations, then the conditions required for point identification of the parameters of the complete model may not be the same as the conditions required in the incomplete model. When the model considered is partially identifying, the identified set for the parameters of the incomplete model may vary as alternative augmentations that produce complete models are employed. This point is elaborated in Section 3.3.

2.3 Outcomes and Unobservables

In a classical IV model the specification of feasible combinations of variables takes the form of one or more structural equations setting out relationships between Y, Z, and U, which are here encoded in the structural function h. In the models studied in this chapter systems of inequalities can also arise. Using a structural function h as defined here accommodates a wide range of possibilities. There are many examples in Section 3.2.

As set out in the Introduction, let $\mathcal{Y}(u, z; h)$ denote the set of values of endogenous variables Y that solve $h\left(Y, Z, U\right) = 0$ with structural function h when $Z = z$ and $U = u$. This is the set of values of Y that may occur when $Z = z$ and $U = u$. By an incomplete model we mean a model that admits structures with structural functions h such that this set can be non-singleton. Let $\mathcal{U}\left(y, z; h\right)$ denote the set of values of unobservable variables U that solve $h\left(Y, Z, U\right) = 0$ with structural function h when $Z = z$ and $U = u$.

The sets $\mathcal{U}\left(y, z; h\right)$ and $\mathcal{Y}(u, z; h)$ are zero level sets of h with respect to U and Y defined for each $(y, z) \in \mathcal{R}_{YZ}$ and $(z, u) \in \mathcal{R}_{ZU}$, respectively:

$$\mathcal{U}\left(y, z; h\right) \equiv \{u : h\left(y, z, u\right) = 0\}, \quad \mathcal{Y}\left(u, z; h\right) \equiv \{y : h\left(y, z, u\right) = 0\}. \tag{9}$$

These level sets are dual to each other in the sense that for all z and h, a value u^* lies in $\mathcal{U}\left(y^*, z; h\right)$ if and only if y^* lies in $\mathcal{Y}\left(u^*, z; h\right)$. This is so because for all z and h

$$u^* \in \mathcal{U}\left(y^*, z; h\right) \Longleftrightarrow h\left(y^*, z, u^*\right) = 0 \Longleftrightarrow y^* \in \mathcal{Y}\left(u^*, z; h\right).$$

An IV model places restrictions on (i) the way in which certain observed exogenous variables affect the determination of outcomes Y and (ii) the nature of the dependence amongst these exogenous variables and unobserved U. Restrictions of the first type are *exclusion* restrictions. Restrictions of the second type are *independence* restrictions. For instance, the classical linear IV model set out in (1) excludes Z_2 from playing any role in determining Y_1, and uses one of several possible independence restrictions, such as $E[Z'U] = 0$, $E[U|Z] = 0$, or $U \perp\!\!\!\perp Z$. A suitable structural function h in this case is

$$h(y, z, u) = y_1 - y_2\beta - z_1\gamma - u,$$

but the square or fourth power of this function would also serve.

This model is incomplete for Y because all that can be known of the value y of Y when $Z = z$ and $U = u$ is that it lies on the manifold

$$\mathcal{Y}(u, z; h) = \{(y_1, y_2) : y_1 - y_2\beta = z_1\gamma + u\},$$

and structural functions h are characterized by values of parameters (β, γ). Here incompleteness arises because the process determines values of multiple outcomes but the model specifies only one structural equation. Even in a model specifying as many equations as there are outcomes there would be incompleteness if there could be multiple solutions to those equations.[8] A model that specifies inequality relationships amongst Y, Z, and U will usually be incomplete.

Complete models have the feature that specifying the distribution of unobservable variables U conditional on the realization of exogenous variables $Z = z$ leads to a specification of the conditional distribution of endogenous variables Y given $Z = z$. In parametrically specified models, this means that $F_{Y|Z}(\cdot|z)$ may be expressed as a known function of model parameters, enabling the use of maximum likelihood methods. On the other hand, in the context of an incomplete model, specifying a conditional distribution of U given $Z = z$ does not lead to a specification of the conditional distribution of Y given $Z = z$ and so alternative approaches are needed as Wald (1950) pointed out.

A common feature of the IV models employed in practice is that terms capturing unobserved heterogeneity are expressible as *single-valued functions* of observed variables. It is a convenient shorthand to talk of such models as having *point-valued residuals*. A model embodying the point-valued residual restrictions admits only structures such that $\mathcal{U}(y, z; h)$ is singleton with probability one.

The point-valued residual restriction clearly holds in the linear model (1) where $U = Y_1 - Y_2\beta - Z_1\gamma$. It also holds in the nonparametric instrumental variables (NPIV) model studied in Newey

---

[8]In Koopmans, Rubin, and Leipnik (1950), this possibility is ruled out in a linear simultaneous system by requiring the matrix of coefficients multiplying endogenous variables have rank equal to the number of such variables, a condition there termed "completeness".

and Powell (2003) which has $Y_1 = f(Y_2, Z_1) + U$, and in the non-additive NPIV model studied in Chernozhukov and Hansen (2005) which has $Y_1 = f(Y_2, Z_1, U)$ with U scalar and the function f *strictly monotone* in its third argument. It also holds in the simultaneous equations models studied by Brown (1983), Roehrig (1988), Matzkin (2008) and Berry and Haile (2018).

To see why the point-valued residual condition is convenient, suppose a model requires that condition to hold, that is the model requires $U = a(Y, Z)$ for some single-valued function $a(\cdot, \cdot)$. In the linear model example above $a(Y, Z) = Y_1 - Y_2\beta - Z\gamma$. If the model requires the distribution of U given Z to satisfy some condition, here denoted Γ (e.g. stochastic independence, mean independence, or zero covariance), then determining the identifying power of the model boils down to finding the set $\mathcal{A}$ of functions $a(\cdot, \cdot)$ such that, for the distribution of Y and Z under consideration, the joint distribution of $a(Y, Z)$ and Z satisfies the condition Γ, that is:

$$\mathcal{A} = \{a(\cdot, \cdot) : \text{distribution of } a(Y, Z) \text{ and } Z \text{ satisfy condition } \Gamma\}.$$

For example, Γ may be the condition that $E[Z'U] = 0$, $E[U|Z] = 0$, or some other restriction limiting the joint distribution of U and Z. The distribution of U given Z associated with a function a in this set is identified as the distribution of $a(Y, Z)$ given Z.

A major focus in the econometrics literature is the determination of rank and completeness conditions under which the set of functions $\mathcal{A}$ is a singleton. These are conditions on the distribution of observed Y and Z over and above the restrictions on structures embodied in the model under consideration. In the linear model just considered, the condition on the distribution of (Y, Z) that $E[Z'(Z_1, Y_1)]$ has full column rank along with the restriction of the model that $E[Z'U] = 0$ suffices to ensure that $\mathcal{A}$ is a singleton set, that is that β and γ are point identified. In this chapter such rank and completeness conditions are distinguished from the *restrictions* of *models* because they are conditions on the distribution of observable variables, which is treated as known in an identification analysis.

When unobservable U is a set-valued function of observed variables it is not obvious how the force of restrictions on the joint distribution of U and Z can be determined. Consider again the threshold crossing model for binary outcome Y_1 with an endogenous explanatory variable Y_2, as in (2), for which

$$\mathcal{U}(y, z; h) = \begin{cases} (-\infty, -y_2\beta - z_1\gamma] & , \quad y_1 = 0 \\ (-y_2\beta - z_1\gamma, \infty) & , \quad y_1 = 1 \end{cases} \tag{10}$$

where Z_2 denotes excluded instruments and $Z \equiv (Z_1, Z_2)$. Knowledge of a value of Y and Z does not identify a unique value of U for any value of the parameters and it turns out that, barring a large support condition on the distribution of Z, many values of β and γ could be consistent with the independence restriction $U \perp\!\!\!\perp Z$ for a particular distribution of Y and Z even if the distribution of U is fully specified, for example as standard normal or logistic.

However the probability distribution of Y and Z can still convey some information about the value of β in the context of this model. To see this, consider a case in which the distribution of Y and Z is such that $\mathbb{P}[Y_1 = 1 | Z_1 = z_1, Z_2 = z_2]$ varies with z_2. In this case β cannot be zero in any structure that generates this distribution of Y and is admitted by the IV model. This is so because if β were zero there would be no route for Z_2 to affect the conditional distribution of Y_1 given Z with Z_1 fixed, since Z_2 is excluded from the structural relationship and U and Z are independent. Indeed the restrictions of the model are falsifiable because if Y_2 and Z were independent while Y_1 and Z were not then one could be sure that the binary outcome IV model is misspecified.[9]

The methods we set out in this Chapter reveal for any distribution of Y and Z precisely the identified set of values of β and γ delivered by the binary outcome IV model. We return to this single equation IV probit model in Section 8 as an example in order to illustrate many aspects of the analysis of GIV models.

2.4   Duality

In identification analysis one asks: what can be learned of the values of parameters of interest from observation of an economic process using observable implications of economic models? To this end one can employ two distinct, complementary approaches, focusing either on properties of the conditional distributions of observable Y given Z or properties of the conditional distributions of unobservable U given Z.

Taking the first approach we ask what the properties of a structure $m = \left(h, \mathcal{G}_{U|Z} \right)$ admitted by a model $\mathcal{M}$ imply about the conditional distributions of Y given Z delivered by a process of interest, and in particular whether these conditional distributions are compatible with these implications. Let $\mathcal{F}_{Y|Z}(\mathcal{Z})$ denote the distributions of Y conditional on $Z = z$ for each $z \in \mathcal{Z} \subseteq \mathcal{R}_Z$ delivered by a process of interest, and let $\mathcal{P}_{Y|Z}(m, \mathcal{Z})$ denote the conditional probability measures of Y given Z for each $z \in \mathcal{Z}$ compatible with structure m. If the conditional measures $\mathcal{F}_{Y|Z}(\mathcal{Z})$ delivered by the process are compatible with the implications of m for a set $\mathcal{Z}$ satisfying $\mathbb{P}[Z \in \mathcal{Z}] = 1$, then structure m is in the identified set of structures obtained with the model $\mathcal{M}$ and the distributions $\mathcal{F}_{Y|Z}(\mathcal{Z})$. Put succinctly, structure m belongs to the identified set of structures obtained with the model $\mathcal{M}$ and the distributions $\mathcal{F}_{Y|Z}(\mathcal{Z})$ if m is admitted by $\mathcal{M}$ and $\mathcal{F}_{Y|Z}(\mathcal{Z})$ and $\mathcal{P}_{Y|Z}(m, \mathcal{Z})$ agree for a set $\mathcal{Z}$ satisfying $P[Z \in \mathcal{Z}] = 1$. If there is only one structure m for which this is so, then the structure is point identified.

This is the logical framework underpinning maximum likelihood estimation, where the model specifies the distributions of Y given Z as a function of model parameters, say θ, each value of which corresponds to a unique structure m. Point identification of these parameters is achieved if there is a unique θ_0 compatible with observable conditional distributions of Y given Z for almost

---

[9]The first results on the identifying power of the binary outcome IV model are in Chesher (2010).

every Z, in which case the information inequality guarantees that the expected log-likelihood is uniquely maximized at θ_0.[10]

Taking the second approach to identification analysis we ask what the conditional distributions of Y given Z delivered by a process, combined with structure m admitted by a model $\mathcal{M}$, imply regarding the conditional distributions of U given Z. If these implications are compatible with restrictions placed on the conditional distributions of unobservable heterogeneity by the model, for example conditional mean independence restrictions, then structure m could have produced the conditional distributions of Y given Z.

This is the idea behind moment based estimation, such as generalized method of moments (GMM) as in Hansen (1982). Restrictions placed on unobservable variables can sometimes be expressed as a requirement that certain functions of observable variables have zero mean. For example, in a non-linear model in which

$$Y_1 = g\left(Y_2, Z_1, \theta\right) + U$$

with $Z = (Z_1, Z_2)$ and the restriction $E\left[U|Z = z\right] = 0$, the moment conditions

$$E\left[\left(Y_1 - g\left(Y_2, Z_1, \theta\right)\right) w\left(Z\right)\right] = 0,$$

can be used as a basis for estimation and inference on θ for suitable collections of functions $w(\cdot)$. This is feasible here, in contrast to the first approach, because in a model such as this, knowledge of the conditional distributions of U given Z is insufficient to uniquely determine the conditional distributions of Y given Z.

The dual roles played by distributions of observable variables and unobservable variables conditional on exogenous variables Z is exploited in the analysis of GIV models in this chapter. The framework used here works with conditional distributions of unobservable variables, and operationalizes a way to verify that a given structure m combined with conditional distributions of Y given Z is compatible with restrictions placed on the distribution of unobservable variables. The framework accommodates models that are incomplete and models that admit set-valued unobservables. The formalization lies in the characterization of observational equivalence provided in Section 4.

The benefit of this approach is that it is straightforward to use in models that place IV restrictions on the conditional distributions of unobservable variables U. This is because it does not require determining the implications for conditional distributions of Y given Z of restrictions on the conditional distribution of U given Z. That can be a complex task when U is multivariate or when U does not enter the structural function in a simple additive fashion. The approach taken here also delivers characterizations of identified sets in some cases in which instrumental variables

---

[10]See for example Newey and McFadden (1994) Lemma 2.2 for a complete proof.

play no role at all. The analysis in CR17 and Chesher and Rosen (2017b) of the auction model of Haile and Tamer (2003) is an example.

Alternative approaches allowing for incompleteness and set-valued unobservables include those of Beresteanu, Molchanov, and Molinari (2011) and Galichon and Henry (2011). Those approaches employ characterizations of conditional distributions of Y given Z compatible with a given structure m, Beresteanu, Molchanov, and Molinari (2011) using a set-valued conditional expectation of outcome variables, Galichon and Henry (2011) using a generalization of a likelihood approach, allowing for the possibility that the distributions of Y given Z compatible with a structure m need not be unique. These sorts of approaches are convenient in models with restrictions that are placed directly on the conditional distributions of Y given Z, perhaps because unobservable variables do not explicitly appear in the models' specifications, or in models in which restrictions on U given Z straightforwardly translate into restrictions on sets of feasible Y given Z.[11] Notable examples of such models include the treatment effect models in Beresteanu, Molchanov, and Molinari (2012), the Roy model studied in Mourifie, Henry, and Meango (2017), and the insurance choice models in Barseghyan, Coughlin, Molinari, and Teitelbaum (2017).

3 GIV Structures and Models

This section gives a formal presentation of the generalized instrumental variables (GIV) framework on which this chapter focuses.

3.1 Restrictions

As discussed in the introduction, GIV models admit structural functions h whose level sets $\mathcal{U}(y, z; h)$ of values of unobserved U may have cardinality *exceeding* one. Both IV and GIV models admit structural functions h whose level sets $\mathcal{Y}(u, z; h)$ of values of endogenous Y may have cardinality greater than one, in which case the model is incomplete. This occurs for example when a model does not specify the way in which some endogenous explanatory variables Y_2 are determined, even if the other endogenous variables Y_1 are uniquely determined by (Y_2, Z, U). Models imposing inequality restrictions on observable and unobservable variables are typically incomplete. Some leading examples are given in Section 3.2 below.

At various points in the exposition we use restrictions chosen from the following collection.

Restriction A1: (Y, Z, U) are random vectors defined on a probability space $(\Omega, \mathsf{L}, \mathbb{P})$, endowed with the Borel sets on Ω. The support of (Y, Z, U), denoted $\mathcal{R}_{YZU}$, is a subset of a finite-dimensional Euclidean space. $\square$

---

[11]Models in which unobservable variables do not explicitly appear include those with potential outcomes and in which individuals possess response functions, both of which provide a different but no more restrictive approach to incorporating unobservable heterogeneity.

Restriction A2: The collection of conditional probability measures

$$\mathcal{F}_{Y|Z} \equiv \left\{ F_{Y|Z}\left(\cdot|z\right) : z \in \mathcal{R}_Z \right\},$$

is identified by the sampling process, where for all measurable $\mathcal{Y} \subseteq \mathcal{R}_{Y|z}$, $F_{Y|Z}\left(\mathcal{Y}|z\right) \equiv \mathbb{P}\left[Y \in \mathcal{Y}|z\right]$.
$\square$

Restriction A3: There is an L-measurable function $h\left(\cdot,\cdot,\cdot\right) : \mathcal{R}_{YZU} \rightarrow \mathbb{R}$ such that $h\left(Y,Z,U\right) = 0$ with probability one and there is a collection of conditional probability measures

$$\mathcal{G}_{U|Z} \equiv \left\{ G_{U|Z}\left(\cdot|z\right) : z \in \mathcal{R}_Z \right\},$$

where for all measurable $\mathcal{S} \subseteq \mathcal{R}_{U|z}$, $G_{U|Z}\left(\mathcal{S}|z\right) \equiv \mathbb{P}\left[U \in \mathcal{S}|z\right]$. $\square$

Restriction A4: The pair $\left(h, \mathcal{G}_{U|Z}\right)$ belongs to a known set of admissible structures $\mathcal{M}$. $\square$

Restriction A5: $\mathcal{U}\left(Y,Z;h\right)$ is closed almost surely $\mathbb{P}\left[\cdot|z\right]$, each $z \in \mathcal{R}_Z$. $\square$

Restriction A6: $\mathcal{Y}\left(U,Z;h\right)$ is closed almost surely $\mathbb{P}\left[\cdot|z\right]$, each $z \in \mathcal{R}_Z$. $\square$

Restriction A1 defines the underlying probability space and restricts the support of (Y, Z, U) to a finite dimensional Euclidean space, as is typically the case in applications. Restriction A2 requires that the conditional distribution of Y given $Z = z$ is identified for each $z \in \mathcal{R}_Z$. Simple random sampling of (Y, Z) is sufficient for this but not required. For instance, choice based sampling and endogenous stratification may deliver identification of $\mathcal{F}_{Y|Z}$. Restriction A3 formalizes the existence of structural relation h and lays out notation for the collection of conditional measures $\mathcal{G}_{U|Z}$ of U given Z. Restrictions A1-A3 are maintained throughout.

Restriction A4 imposes model $\mathcal{M}$, the collection of admissible structures $\left(h, \mathcal{G}_{U|Z}\right)$. These encompass the researcher's *a priori* restrictions on the process generating observable (Y, Z). For instance, the researcher may require that the structural function h and conditional distributions of unobservable variables $\mathcal{G}_{U|Z}$ are restricted to parametric classes, so that $\mathcal{M}$ comprises $\left(h, \mathcal{G}_{U|Z}\right)$ such that $h = h_\theta$ and $G_{U|Z}\left(\cdot|z\right) = G_{U|Z}\left(\cdot|z, \theta\right)$ for some $\theta \in \Theta \subseteq \mathbb{R}^{d_\theta}$. Semiparametric or nonparametric restrictions on $\mathcal{M}$ are also allowed. In contrast to Restrictions A1-A3, Restriction A4 may be refutable based on knowledge of $\mathcal{F}_{Y|Z}$.[12] Characterizations of identified sets given admissible structures $\mathcal{M}$ entail those structures $\left(h, \mathcal{G}_{U|Z}\right) \in \mathcal{M}$ that, under Restrictions A1-A3, could deliver the identified conditional distributions $\mathcal{F}_{Y|Z}$. It is possible that there is no $\left(h, \mathcal{G}_{U|Z}\right)$ belonging to $\mathcal{M}$ such that $\mathbb{P}\left[h\left(Y,Z,U\right) = 0\right] = 1$ for some random variable U with conditional distributions belonging to $\mathcal{G}_{U|Z}$. This possibility is allowed, and in such cases the identified set of structures is empty, indicating that the model is misspecified. The development of specification tests for models allowing for set identification is an ongoing area of research, with a notable contribution provided by Bugni, Canay, and Shi (2015), who provide specification tests based on moment inequalities.

---

[12]For a formal statement of refutability, see Breusch (1986).

Restrictions A5 and A6 restrict the level sets $\mathcal{U}(Y, Z; h)$ and $\mathcal{Y}(U, Z; h)$ to be random closed sets, respectively. This enables application of results from random set theory to establish that certain collections of conditional moment inequalities characterizing bounds on structures are sharp. This restriction is usually easy to accommodate. Restriction A5 is for example guaranteed whenever the structural function $h(y, z, u)$ is continuous in u in a neighborhood of those u such that $h(y, z, u) = 0$. In models in which the usual formulations do not automatically satisfy these restrictions, the structural function can often be redefined in an equivalent way to ensure that one of these restrictions does hold.

An example of such a redefinition arises in the binary outcome model with outcome Y_1 determined by (2), which is the commonly used way of specifying binary outcome threshold-crossing models, such as the binary probit model. A corresponding structural function h such that $h(Y, Z, U)$ is equal to zero almost surely when (2) holds is given by

$$h(y, z, u) = (1 - y_1) \cdot 1\left[y_2\beta + z_1\gamma + u > 0\right] + y_1 \cdot 1\left[y_2\beta + z_1\gamma + u \leq 0\right],$$

which produces the U-level sets $\mathcal{U}(y, z; h)$ of the form given in (10). When $y_1 = 1$, this set is the open interval $(-y_2\beta - z_1\gamma, \infty)$, and so is not closed in the Euclidean topology. This is because the outcome variable is assigned the value 1 only if the latent index $y_2\beta + z_1\gamma + u$ *strictly* exceeds zero. This is however simply a matter of convention. In such threshold crossing models the unobservable U is restricted to be continuously distributed, for example standard normal in the probit model. Then the conditional probability that the index $Y_2\beta + Z_1\gamma + U$ is equal to zero is itself zero, and whether Y_1 takes the value 1 or 0 when this zero probability event is realized is of no substantive consequence.

One may therefore redefine the model without loss of generality to allow Y_1 to take either value 1 or 0 when $Y_2\beta + Z_1\gamma + U = 0$. A structural function that permits either value of Y_1 in this event is given by

$$h(y, z, u) = y_1\left|y_2\beta + z_1\gamma + u\right|_- + (1 - y_1)\left|y_2\beta + z_1\gamma + u\right|_+ \tag{11}$$

where $|a|_+ = \max(0, a)$ and $|a|_- = \max(0, -a)$ denote respectively the positive and negative parts of a. The set $\mathcal{U}(y, z; h)$ defined as those u such that $h(y, z, u) = 0$ is the closure of the U-level set implied by (10).[13]

To see that Restriction A6 holds in this model note that

$$\mathcal{Y}(u, z; h) = \mathcal{Y}_0(u, z; h) \cup \mathcal{Y}_1(u, z; h), \tag{12}$$

---

[13] Closed sets are sets that contain their limit points, so the intervals $(-\infty, -\alpha y_2 - \beta]$ and $[-\alpha y_2 - \beta, \infty)$ are closed in $\mathbb{R}$ despite being "open" at $\pm\infty$.

where

$$
\begin{aligned}
\mathcal{Y}_0\left(u, z; h\right) &\equiv \left\{y \in \mathcal{R}_Y : y_1 = 0 \wedge y_2\beta + z_1\gamma + u \leq 0\right\}, \\
\mathcal{Y}_1\left(u, z; h\right) &\equiv \left\{y \in \mathcal{R}_Y : y_1 = 1 \wedge y_2\beta + z_1\gamma + u \geq 0\right\},
\end{aligned}
$$

so $\mathcal{Y}\left(u, z; h\right)$ is a union of two closed sets and is therefore closed.

Other models allowing discrete outcomes can be similarly modified to ensure Restriction A5 holds. More generally, in some models where Restriction A5 does not hold automatically, it may be possible to employ topologies other than the Euclidean topology to satisfy the closedness requirement.

Restrictions A1-A6 are very general in the sense that they can be applied in a wide variety of contexts. Consequently, identification analysis built on these restrictions is widely applicable. The level of generality allows for the possibility that identified sets of structures are large or small, for example the entire admissible space at one extreme, or a singleton point or possibly even the empty set at the other.

The characterizations can simplify under the restrictions imposed by a particular model specification. The general characterizations laid out in Section 4 apply *whatever* restrictions are embodied in a model, but the simplification afforded by particular restrictions can be enormous, and sometimes essential in obtaining characterizations amenable to use in practice. This is the focus of Section 6, where the identifying power of restrictions on the conditional distributions of U given Z are considered. The widely applicable characterizations in Section 4 provide an important step in obtaining these. This is because with the general characterization in hand, to show that a given set of observable implications comprise the identified set, the task is simply to establish that, with the additional restrictions in hand, their implications in fact imply all of the implications of the general characterization.

In addition to restrictions on structural functions and conditional distributions of unobservables that are imposed through the model specification, the identified set of structures depends crucially upon the joint distribution of the observed variables (Y, Z). This distribution, and consequently the collection of conditional distributions $\mathcal{F}_{Y|Z}$, are point identified under Restriction A2 which relies on the suitability of the sampling scheme under which the process of interest is observed. It is unnecessary to place restrictions on $\mathcal{F}_{Y|Z}$, when, as here, developing characterizations of identified sets of structures. This may appear at odds with existing identification results involving rank or completeness conditions which place restrictions on $\mathcal{F}_{Y|Z}$, but it is not. Such conditions are invoked as conditions under which there is *point* identification of parameter values or functions. Such conditions do not arise in characterizations of identified sets of structures but they are of great interest when considering what classes of distributions $\mathcal{F}_{Y|Z}$ can deliver identified sets that are singleton, and what characteristics of the distributions $\mathcal{F}_{Y|Z}$ affect the size of the identified set.

Here are some examples of models that fall within the GIV framework studied here. In each example there can be alternative restrictions on the joint distribution of (U, Z), for example full independence, $U \perp\!\!\!\perp Z$, mean independence, $E[U|Z] = 0$, or quantile independence, $q_{U|Z}(\tau|Z) = 0$, and/or parametric restrictions on the distribution of U.

3.2 Examples

Example 1. *The classical linear IV model set out in the Introduction has structural function*

$$h(y, z, u) = y_1 - y_2\beta - z_1\gamma - u$$

with level set $\mathcal{Y}(u, z; h) = \{((y_2\beta + z_1\gamma + u), y_2) : y_2 \in \mathcal{R}_{Y_2}\}$ which is not a singleton set. The level set $\mathcal{U}(y, z; h)$ is the singleton set $\{(y_1 - y_2\beta - z_1\gamma)\}$. ∎

Example 2. *A binary threshold crossing model with $Y_1 = 1[g(Y_2, Z_1) < U]$ and U normalized uniformly distributed on $[0, 1]$ as studied in Chesher and Rosen (2013) has structural function*

$$h(y, z, u) = y_1 |u - g(y_2, z_1)|_- + (1 - y_1)|u - g(y_2, z_1)|_+$$

where $y_1 \in \{0, 1\}$.[14] *The corresponding level sets are values of pairs (y_1, y_2)*

$$\mathcal{Y}(u, z; h) = \{(y_1, y_2) \in \mathcal{R}_{Y_1 Y_2} : \{y_1 = 1 \wedge u \geq g(y_2, z_1)\} \text{ or } \{y_1 = 0 \wedge u \leq g(y_2, z_1)\}\}$$

and intervals

$$U(y, z; h) = \begin{array}{l} [0, g(y_2, z_1)] \text{ if } y_1 = 0, \\ [g(y_2, z_1), 1] \text{ if } y_1 = 1, \end{array}$$

both of which are closed sets in $\mathbb{R}$. ∎

Example 2*. *In the binary outcome model of Example 2, suppose the form of the structural function is further restricted such that*

$$g(y_2, z_1) \equiv \Lambda(-y_2\beta - z_1\gamma),$$

where $\Lambda(\cdot)$ denotes a strictly increasing CDF on $\mathbb{R}$, for instance the standard normal CDF, in which case $\Lambda(\cdot) = \Phi(\cdot)$. Then $Y_1 = 1[\Phi^{-1}(U) + Y_2\beta + Z_1\gamma > 0]$, and with a change of variable replacing the unobservable $\Phi^{-1}(U)$ with a standard normally distributed variate $V = \Phi^{-1}(U)$ we arrive at the IV probit specification (2) of the Introduction. Once again ignoring the zero probability

---

[14] Following the same argument provided previously for the more restrictive binary outcome model, we have that if U is continuously distributed conditional on realizations of (Y_2, Z_1) it is straightforward to show that since $g(Y_2, Z_1) = U$ occurs with zero probability, in the ensuing identification analysis it is of no consequence whether Y_1 takes value 1 or 0 when this occurs. To simplify the exposition we define the structural function h such that either value of Y_1 is permitted when $g(Y_2, Z_1) = U$.

event that the unobservable is exactly equal to the threshold determining Y_1 gives the U-level sets represented in (10) and the Y-level sets expressed in (12). ∎

Example 3. *Multiple discrete choice with endogenous explanatory variables as studied in Chesher, Rosen, and Smolinski (2013). This is the random utility model of McFadden (1974) extended to allow some explanatory variables to be endogenous. The structural function is*

$$h(y, z, u) = \left| \min_{k \in \{1,\dots,J\}} \left(\pi_{y_1} (y_2, z_1, u_j) - \pi_k (y_2, z_1, u_k) \right) \right|_-,$$

where $\pi_j (y_2, z_1, u_j)$ is the utility associated with choice $j \in \mathcal{J} \equiv \{1, \dots, J\}$ and $u = (u_1, \dots, u_J)$ is a vector of unobserved preference heterogeneity. Y_1 is the outcome variable indicating the choice made and Y_2 contains endogenous explanatory variables. The exogenous variables Z_1 are allowed to enter the utility functions $\pi_1, \dots, \pi_J$, while Z_2 are excluded exogenous variables. The Y and U level sets are respectively

$$\mathcal{Y}(u, z; h) = \left\{ \left(\arg \max_{j \in \mathcal{J}} \pi_j (y_2, z_1, u_j), y_2 \right) : y_2 \in \mathcal{R}_{Y_2} \right\}$$

and

$$\mathcal{U}(y, z; h) = \left\{ u \in \mathcal{R}_U : y_1 = \arg \max_{j \in \mathcal{J}} \pi_j (y_2, z_1, u_j) \right\}$$

∎

Example 4. *A continuous-outcome random coefficients model with an endogenous explanatory variable has structural function*

$$h(y, z, u) = y_1 - z_1 \gamma - (\beta_2 + u_2) y_2 - (\beta_1 + u_1).$$

The random coefficients are $(\beta_1 + U_1)$ and $(\beta_2 + U_2)$, with means β_1 and β_2 respectively. The coefficient γ multiplying exogenous variables in h could also be random. The Y and U level sets are

$$\mathcal{Y}(u, z; h) = \{ (z_1 \gamma + (\beta_2 + u_2) y_2 + (\beta_1 + u_1), y_2) : y_2 \in \mathcal{R}_{Y_2} \},$$

and

$$U(y, z; h) = \{ u \in \mathcal{R}_U : u_1 = y_1 - z_1 \gamma - \beta_1 - \beta_2 y_2 - u_2 y_2 \},$$

respectively. ∎

Example 5. *Interval censored endogenous explanatory variables. Let $g(\cdot, \cdot, \cdot) : \mathbb{R} \times \mathbb{R}^k \times \mathbb{R} \to \mathbb{R}$ be monotone in its first argument and strictly monotone nondecreasing in its third argument such that*

$$Y_1 = g(Y_2^*, Z_1, U),$$

where endogenous variable $Y_2^* \in \mathbb{R}$ is interval censored with

$$\mathbb{P}\left[Y_{2l} \leq Y_2^* \leq Y_{2u}\right] = 1,$$

for observed variables Y_{2l}, Y_{2u}. No further restriction is placed on the process determining the realizations of Y_{2l}, Y_{2u}. The structural function is

$$h(y, z, u) = \left|y_1 - g\left(y_{2l}, z_1, u\right)\right|_- + \left|y_1 - g\left(y_{2u}, z_1, u\right)\right|_+,$$

with $y \equiv (y_1, y_{2l}, y_{2u})$, and $y_{2l} \leq y_{2u}$. The resulting level sets are

$$\mathcal{Y}(u, z; h) = \left\{y \in \mathcal{R}_Y : g\left(y_{2l}, z_1, u\right) \leq y_1 \leq g\left(y_{2u}, z_1, u\right) \wedge y_{2l} \leq y_{2u}\right\},$$

and

$$\mathcal{U}(y, z; h) = \left[g^{-1}\left(y_{2u}, z_1, y_1\right), g^{-1}\left(y_{2l}, z_1, y_1\right)\right],$$

where the function $g^{-1}(\cdot, \cdot, \cdot)$ is the inverse of $g(\cdot, \cdot, \cdot)$ with respect to its third argument, so that for all y_2, z_1, and u, $g^{-1}(y_2, z_1, g(y_2, z_1, u)) = u$. This example is a generalization of a model studied by Manski and Tamer (2002), in which the interval censored variable was restricted to be exogenous. ∎

Example 6. *English ascending auction. This is similar to the model studied in Haile and Tamer (2003), with reserve price set to zero. There are J symmetric bidders making non-negative final bids $Y = (Y_1, \ldots, Y_J)$, there are continuously distributed non-negative valuations $U = (U_1, \ldots, U_J)$ and U has probability distribution $G_{U|Z}(\cdot|z)$ conditional on auction characteristics $Z = z$. Let $a_{[j]}$ denote the jth smallest element of vector $a \equiv (a_1, \cdots, a_J)'$, so that for instance $a_{[1]} = \min\{a_1, \cdots, a_J\}$ and $a_{[J]} = \max\{a_1, \cdots, a_J\}$. The structural function and resulting level sets are as follows.*

$$h(y, z, u) = \sum_{j=1}^{J} \left|y_j - u_j\right|_+ + \left|y_{[J]} - u_{[J-1]}\right|_-$$

$$\mathcal{Y}(u, z; h) = \left\{y \in \mathcal{R}_Y : \bigwedge_{j=1}^{j}\left(y_j \leq u_j\right) \wedge \left(y_{[J]} \geq u_{[J-1]}\right)\right\}$$

$$\mathcal{U}(y, z; h) = \left\{u \in \mathcal{R}_U : \bigwedge_{j=1}^{J}\left(y_j \leq u_j\right) \wedge \left(y_{[J]} \geq u_{[J-1]}\right)\right\}$$

The structural function embodies the restrictions imposed in Haile and Tamer (2003) that no person bids in excess of their valuation and no person allows another to win at a price below their valuation. The structural function h is known and does not depend on z. The unknown structural feature is

the distribution of valuations, $\mathcal{G}_{U|Z} = \{G_{U|Z}(\cdot|z) : z \in \mathcal{R}_Z\}$. In Haile and Tamer (2003) there is the restriction that conditional on any value of Z the elements of U are identically and independently distributed. This example is studied in detail in CR17, in which a characterization of the identified set of valuation distributions is derived. ■

These examples are just a selection from the wide array of structural models to which the GIV framework can be applied. Examples 2-5 will be used later to demonstrate application of several subsequent developments of this chapter. Example 6 is studied in detail as the lead example of CR17, in which it is shown that the identified set for the common valuation distribution in the Independent Private Value (IPV) framework refines the bounds previously available. The analysis is extended to non-IPV settings in the working paper Chesher and Rosen (2017b).

There are in addition several quite recent applications of the GIV framework that conduct set identification analysis through the use of sets of values of unobservables that are compatible with values of observed variables. Kim (forthcoming) applies GIV analysis to characterize identified sets for single equation IV count data models, and features an application investigating the effect of supplemental health insurance on the frequency of doctor visits made by U.S. individuals over the age of 65. Berry and Compiani (2019) study dynamic models in which unobservable variables are allowed to be serially correlated, which renders observed states endogenous. To deal with the endogeneity problem, they use GIV methods to set identify dynamic policy functions in both single agent and oligopoly models. They use inference methods from Chernozhukov, Chetverikov, and Kato (forthcoming) in an application the ready-mix concrete industry using data from Collard-Wexler (2013). Tebaldi, Torgovitsky, and Yang (2019) apply an unordered discrete choice model to study insurance plan choice under the Affordable Care Act in California. They conduct partial identification analysis through careful use of the properties of sets of values of unobservable variables as in the GIV framework and in CRS13 for multiple discrete choice models. They show how, using their model, sharp bounds on several economically meaningful projections of the identified set of structures can be characterized by solutions to linear programming problems, lending computational tractability to estimation.

3.3   Completing Incomplete Models?

Consider again the comparison of complete and incomplete models started in Section 2.2. An incomplete model $\mathcal{M}$ allows that amongst the collection of structures $m = \left(h, \mathcal{G}_{U|Z}\right)$ that it admits there can be level sets $\mathcal{Y}(U, Z; h)$ that are not singleton sets. Any such model may be completed by augmenting the model with a class of permissible selection functions q that map from sets $\mathcal{Y} \subseteq \mathcal{R}_Y$ to unique values of $y \in \mathcal{Y}$. The selection function may in general be a function of any other components of the model, so that

$$q : 2^{\mathcal{R}_Y} \times \mathcal{R}_U \times \mathcal{R}_Z \times \mathcal{H} \to \mathcal{R}_Y, \qquad q\left(\mathcal{Y}, u, z, h\right) \in \mathcal{Y} \tag{13}$$

where $\mathcal{H}$ is the set of structural functions admitted by the incomplete model. An augmented complete model may then be obtained by taking all possible compositions of such q and h to obtain complete structural functions h^c for which $\mathcal{Y}(U, Z; h^c)$ is a singleton:

$$\mathcal{Y}(U, Z; h^c) \equiv q\left(\mathcal{Y}(U, Z; h), U, Z, h\right),$$

with $\mathcal{M}^c$ denoting the collection of all compositions $m^c = \left(h^c, \mathcal{G}_{U|Z}\right) = \left(\left(q\left(\mathcal{Y}(\cdot, \cdot; h), \cdot, \cdot, h\right)\right), \mathcal{G}_{U|Z}\right)$ taken over admissible combinations of q and h.

Hypothesizing the existence of such a completion q and defining a complete model $\mathcal{M}^c$ in this way is always possible, but is not without cost. Completing an incomplete model and restricting allowable selection functions q will in general alter the model, so that the projection of the identified set for m^c onto $\left(h, \mathcal{G}_{U|Z}\right)$ will in general be a subset of the identified set for $\left(h, \mathcal{G}_{U|Z}\right)$ based on the original incomplete model. In general, adding such a completion to the model imposes additional restrictions on the process generating outcomes, which serve to shrink the identified set for $\left(h, \mathcal{G}_{U|Z}\right)$. The only way to avoid this possibility is to allow *any completion* (13), without restriction, which will guarantee no identifying content is imposed. This is not without cost.

Completing an incomplete model without imposing further substantive restrictions requires keeping track of and characterizing a larger set of model primitives, specifically whichever ones are required for the specification of q. If one is interested in learning about features of the incomplete model structures $\left(h, \mathcal{G}_{U|Z}\right)$, then these additional components of q are nuisance parameters that have to be projected away. Their presence increases the number of primitives in the model, which may complicate characterization of the relevant features of $\left(h, \mathcal{G}_{U|Z}\right)$. As a case-in-point, in the context of econometric models of games with multiple equilibria, Beresteanu, Molchanov, and Molinari (2011) showed that an intractable complete specification featuring an infinite dimensional equilibrium selection mechanism (a type of completion function) could be replaced with a more tractable incomplete model specification.

If, on the other hand, there are credible and substantive restrictions that restrict $\mathcal{Y}(U, Z; h)$ to be a singleton set then these should be incorporated in the specification of h and, with that done, the GIV framework remains applicable as discussed in Section 5.2.1.

4   Observational Equivalence

The notion of observational equivalence of structures has been a useful basis for identification analysis since it was introduced in the early days of structural econometrics at the Cowles Commission, in particular in Koopmans (1949) and the Cowles 10 chapters Koopmans and Reiersøl (1950) and Hurwicz (1950). The identified set of structures $\mathcal{I}\left(\mathcal{M}, \mathcal{F}_{Y|Z}, \mathcal{R}_Z\right)$ delivered by a complete model $\mathcal{M}$ and conditional distributions $\mathcal{F}_{Y|Z}$ are those structures admitted by $\mathcal{M}$ which are observationally equivalent to one another, and which generate the conditional distributions $\mathcal{F}_{Y|Z}$ for almost

every $z \in \mathcal{R}_Z$. The explicit use of observational equivalence is not the only path to identification analysis, but its use provides an orderly formalization within which to work. It has been used to good purpose in key papers on parametric identification such as Rothenberg (1971) and Bowden (1973), and applied in nonparametric identification analysis as set out in Matzkin (2007, 2008).

Section 4.1 reviews the classical definition of observational equivalence that applies in complete models in which admissible structures are required to produce a unique distribution of outcome variables Y conditional on exogenous variables Z for each possible realization $z \in \mathcal{R}_Z$. This formulation of observational equivalence does not apply to incomplete models.

Section 4.2 provides a generalization of the classical definition of observational equivalence suitable for incomplete models. This was first developed in CR17. Relative to that analysis, some subtleties are discussed that expand on the difference from the classical definition, and which are illustrated in the context of the IV probit model of Example 2*.

Section 4.3 provides an equivalent formulation of the new definition of observational equivalence for incomplete models in terms of sets of unobservable variables implied by realizations of observable variables. This formulation is then applied to provide a convenient formulation of identified sets for structural models for which restrictions on unobservable variables are easily incorporated.

Section 6 uses this formulation to provide an orderly characterization of identified sets for structures in models employing a variety of particular restrictions on the joint distribution of unobservable variables and instruments, such as independence, conditional mean, and conditional quantile restrictions, as are commonly used in instrumental variable models.

4.1   Observational Equivalence in Complete Models

The classical definition of observational equivalence applies in contexts in which each structure, m, delivers a single conditional probability measure $P_{Y|Z}(\cdot|z;m)$ for each $z \in \mathcal{R}_Z$, the collection of which is denoted

$$\mathcal{P}_{Y|Z}(m) \equiv \{P_{Y|Z}(\cdot|z;m) : z \in \mathcal{R}_Z\}.$$

Two structures m and m' are said to be observationally equivalent if $\mathcal{P}_{Y|Z}(m) = \mathcal{P}_{Y|Z}(m')$ almost surely. This holds if there exists a set $\mathcal{Z}$ such that (i) $Z \in \mathcal{Z}$ with probability one and (ii) $P_{Y|Z}(\cdot|z;m) = P_{Y|Z}(\cdot|z;m')$ for all $z \in \mathcal{Z}$.

As an example consider the binary probit model with exogenous explanatory variables:

$$Y = 1\left[Z\beta + U > 0\right], \quad U \sim N(0,1), \quad U \per\!\!\!\perp Z.$$

The distribution of unobserved heterogeneity is known, and the structural function is parametrically specified. The structure m is determined by the value of β and the notation $P_{Y|Z}(\cdot|z;\beta)$ may be used in place of $P_{Y|Z}(\cdot|z;m)$. In this probit model $P_{Y|Z}(\cdot|z;\beta)$ places probability mass $\Phi(z\beta)$ on 1, and mass $1 - \Phi(z\beta)$ on 0. Two parameter vectors β and b are observationally equivalent if and

only if $\Phi(Z\beta) = \Phi(Zb)$ with probability one $\mathbb{P}_Z$.

Note that under this classical definition of observational equivalence, the question of whether two structures m and m' are observationally equivalent is addressed by considering only the conditional distributions of Y given Z, $\mathcal{P}_{Y|Z}(m)$ and $\mathcal{P}_{Y|Z}(m')$, produced by these two structures. Whether or not two structures admitted by complete models are observationally equivalent is answered without reference to the conditional distributions $\mathcal{F}_{Y|Z}$ delivered by the process being studied.

4.2 Observational Equivalence in Incomplete Models

The definition of observational equivalence of structures given for complete models cannot be used in identification analysis of incomplete models. This is because incomplete models admit structures $m \in \mathcal{M}$ that are incomplete, such that a particular realization of exogenous observable and unobservable variables may be associated with *multiple* alternative realizations of endogenous Y.

Considering such a structure with components h and $\mathcal{G}_{U|Z}$, the set $\mathcal{Y}(u, z; h)$ defined in equation (9) is the set of values of Y that can be obtained using the structural function h when $U = u$ and $Z = z$. When $U \sim \mathcal{G}_{U|Z}(\cdot|z)$ the structure delivers the *random* set $\mathcal{Y}(U, z; h)$. This random set can be characterized by the *selections* of the random set.[15] These are the point-valued random variables that lie in the random set with probability 1. An incomplete structure delivers a collection of conditional distributions of Y given $Z = z$, comprising the set of probability distributions that are distributions of the selections of the random set.

So, in the analysis of incomplete models there can be, for each z, not a *single* distribution $P_{Y|Z}(\cdot|z; m)$, but rather a *set* of distributions to consider. Let $\mathcal{P}_{Y|Z}(\cdot|z; m)$ denote such a set. This is so for each z, and so enumerating the set across values of z there is, associated with each structure, a collection of sets of distributions taken across $z \in \mathcal{R}_Z$, here denoted by

$$\mathsf{P}_{Y|Z}(m) \equiv \{\mathcal{P}_{Y|Z}(\cdot|z; m) : z \in \mathcal{R}_Z\}.$$

To be clear $\mathcal{P}_{Y|Z}(\cdot|z; m)$ denotes the *set* of conditional distributions for Y given $Z = z$ compatible with structure m, and $\mathsf{P}_{Y|Z}(m)$ denotes the collection of such sets taken over $z \in \mathcal{R}_Z$.

To illustrate, consider the probit model of Example 2*, which has an endogenous right hand side variable Y_2 and an IV restriction:

$$Y_1 = 1[Y_2\beta + Z_1\gamma + V > 0], \tag{14}$$

with $V \equiv \Phi^{-1}(U) \sim N(0, 1)$ and $U \perp\!\!\!\perp Z \equiv (Z_1, Z_2)$. For any realization of (Z, V), this model is silent regarding the determination of Y_2; any conditional distribution of Y_2 given (Z, V) is admissible. The conditional distribution of the unobservable variables given Z is fully specified, so a structure

---

[15]We give formal definitions shortly.

is characterized by the values of the parameters β and γ that determine the structural function.

Given parameters (β, γ), the conditional distribution of $Y \equiv (Y_1, Y_2)$ given $Z = z$, denoted $P_{Y|Z}(\cdot|z; \beta, \gamma)$, is restricted to be one of those obtained as the conditional distribution of

$$(1[Y_2\beta + Z_1\gamma + V > 0], Y_2)$$

given $Z = z$ for some random variables (Y_2, V) such that $V \perp\!\!\!\perp Z$ and $V \sim N(0, 1)$. Because any conditional distribution for Y_2 given $Z = z$ and $V = v$ is admitted by the model, there are many candidate conditional distributions $P_{Y|Z}(\cdot|z; \beta, \gamma)$, which taken together comprise the collection of conditional distributions of Y given $Z = z$, $\mathcal{P}_{Y|Z}(\cdot|z; \beta, \gamma)$. By contrast a complete model, for example specifying an equation determining Y_2 as a function of Z and unobservable variables, say W, would deliver a unique conditional distribution for Y given $Z = z$ at each choice of conditional distribution of (W, V) given $Z = z$.

Consider now the question of observational equivalence of two distinct structures m and m' that are admitted by an incomplete model. Associated with each of these structures is an accompanying collection of conditional distributions, $\mathsf{P}_{Y|Z}(m)$ and $\mathsf{P}_{Y|Z}(m')$, generated by these structures. For each $z \in \mathcal{R}_Z$ there are corresponding sets of conditional distributions $\mathcal{P}_{Y|Z}(\cdot|z; m) \in \mathsf{P}_{Y|Z}(m)$ and $\mathcal{P}_{Y|Z}(\cdot|z; m') \in \mathsf{P}_{Y|Z}(m')$. If $\mathcal{P}_{Y|Z}(\cdot|z; m)$ and $\mathcal{P}_{Y|Z}(\cdot|z; m')$ intersect, that is if there is at least one conditional distribution $P_{Y|Z}(\cdot|z)$ that is a member of both sets $\mathcal{P}_{Y|Z}(\cdot|z; m)$ and $\mathcal{P}_{Y|Z}(\cdot|z; m')$, then it is *possible* that both structures m and m' generate the same conditional distribution of Y given $Z = z$. It is also possible that $\mathcal{P}_{Y|Z}(\cdot|z; m)$ and $\mathcal{P}_{Y|Z}(\cdot|z; m')$ intersect, but are not identical sets, so that there are conditional distributions that belong to $\mathcal{P}_{Y|Z}(\cdot|z; m)$ but not $\mathcal{P}_{Y|Z}(\cdot|z; m')$, and *vice versa*.

The conditional distribution of Y given $Z = z$ delivered by the process being studied, denoted $F_{Y|Z}(\cdot|z)$, is point identified for each $z \in \mathcal{R}_Z$. The structure m imparts which conditional distributions $P_{Y|Z}(\cdot|z; m)$ are *possible*, while the researcher has data that yields identification of the *actual* conditional distributions $F_{Y|Z}(\cdot|z)$ delivered by the process being studied, the collection of which is denoted $\mathcal{F}_{Y|Z}$ as set out in Restriction A2. In general, and in contrast to the case when analyzing complete models, whether or not two structures m and m' are observationally equivalent in the consideration of incomplete models can depend on the actual distributions $F_{Y|Z}(\cdot|z)$ at hand. This is so precisely because of the possibility described above. It may be that there is for each z some $P_{Y|Z}(\cdot|z)$ that belongs to both $\mathcal{P}_{Y|Z}(\cdot|z; m)$ and $\mathcal{P}_{Y|Z}(\cdot|z; m')$, while there may also be distributions $P'_{Y|Z}(\cdot|z)$ which for some positive measure set of z belong to $\mathcal{P}_{Y|Z}(\cdot|z; m)$ but not to $\mathcal{P}_{Y|Z}(\cdot|z; m')$ or vice versa. If the collection of distributions $\mathcal{F}_{Y|Z} \equiv \{F_{Y|Z}(\cdot|z) : z \in \mathcal{R}_Z\}$ contains $F_{Y|Z}(\cdot|z)$ of the former type, belonging to $\mathcal{P}_{Y|Z}(\cdot|z; m)$ and $\mathcal{P}_{Y|Z}(\cdot|z; m')$ for almost every z, then m and m' are observationally equivalent for this $\mathcal{F}_{Y|Z}$. But if $\mathcal{F}_{Y|Z}$ is of the second type, so that there is a positive measure set of values z for which $F_{Y|Z}(\cdot|z)$ does not belong to one or the other

of $\mathcal{P}_{Y|Z}\left(\cdot|z;m\right)$ or $\mathcal{P}_{Y|Z}\left(\cdot|z;m'\right)$ then m and m' are not observationally equivalent.

Consequently, in the following development, observational equivalence is defined with respect to the (identified) collection of distributions $\mathcal{F}_{Y|Z}$. Potential observational equivalence is also defined. Structures m and m' are potentially observationally equivalent when the collections $\mathsf{P}_{Y|Z}\left(m\right)$ and $\mathsf{P}_{Y|Z}\left(m'\right)$ have sets of distributions in common a.e. $z \in \mathcal{R}_Z$. Potential observational equivalence is a property that two structures can possess irrespective of the collection of conditional distributions $\mathcal{F}_{Y|Z}$ under consideration in identification analysis. When $\mathsf{P}_{Y|Z}\left(m\right)$ and $\mathsf{P}_{Y|Z}\left(m'\right)$ have no sets of distributions in common m and m' are observationally distinct for all $\mathcal{F}_{Y|Z}$.

Formal definitions are required. We start with definitions of a *random closed set*, as well as that of a *measurable selection* from a random set and *selectionability*. These are available as Molchanov (2005, Definition 1.1, p.1, Definition 2.2, p. 26 and Definition 2.19, p. 34), see also Molchanov and Molinari (2018), and are included here for completeness.

Definition 1 *Let $\mathbb{E}$ be a locally compact Hausdorff second countable topolgical space, and let $\mathsf{F}\left(\mathbb{E}\right)$ denote the family of closed sets on $\mathbb{E}$. A map $\mathcal{W} : \Omega \to \mathsf{F}\left(\mathbb{E}\right)$ is a **random closed set** if for every compact set $\mathcal{K}$ in $\mathbb{E}$, the set $\{\omega \in \Omega : \mathcal{W}\left(\omega\right) \cap \mathcal{K} \neq \emptyset\} \in \mathsf{L}$.*

Leading examples of locally compact second countable Hausdorff topolgical spaces, sufficient for our purposes here, are finite dimensional Euclidean spaces. It is also possible to consider families of random sets satisfying other properties, for example random open sets and random convex sets, see for example Molchanov (2005, Sections 4.6 and 4.7, pp. 63-65.) We work mostly with random closed sets here, but some statements, such as the definition of a measurable selection that follows, only require that the random set in question be measurable, and so the "closed" qualifier may be dropped.

Definition 2 *Let W and $\mathcal{W}$ denote a random vector and random set defined on the same probability space. W is a **measurable selection** of $\mathcal{W}$, denoted $W \in \mathrm{Sel}\left(\mathcal{W}\right)$, if $W \in \mathcal{W}$ with probability one. The distribution F_W of random vector W is **selectionable** with respect to the distribution of random set $\mathcal{W}$, which we abbreviate $F_W \preccurlyeq \mathcal{W}$, if there exists a random variable $\tilde{W}$ distributed F_W and a random set $\tilde{\mathcal{W}}$ with the same distribution as $\mathcal{W}$ such that $\tilde{W} \in \mathrm{Sel}\left(\tilde{\mathcal{W}}\right)$.*

A given structure $m = \left(h, \mathcal{G}_{U|Z}\right)$ induces a distribution for the random outcome set $\mathcal{Y}\left(U, Z; h\right)$ conditional on $Z = z$, for all $z \in \mathcal{R}_Z$. If $\mathcal{Y}\left(U, Z; h\right)$ is a singleton set with probability one for all admissible h, then the model is complete, and the conditional distribution of $\mathcal{Y}\left(U, Z; h\right)$ given $Z = z$ is simply that of $\{Y\}$ given $Z = z$ for each $z \in \mathcal{R}_Z$. In this case, again for each $z \in \mathcal{R}_Z$, $F_{Y|Z}\left(\cdot|z\right)$ is the *only* conditional distribution of Y given $Z = z$ that is selectionable with respect to the conditional distribution of $\mathcal{Y}\left(U, Z; h\right)$, and the definition of observational equivalence below simplifies to the classical one.

If, on the other hand, the model is incomplete, so that $\mathcal{Y}(U, Z; h)$ is non-singleton with positive probability, then $h(Y, Z, U) = 0$ dictates only that $Y \in \mathcal{Y}(U, Z; h)$, which is insufficient to uniquely determine the conditional distributions $\mathcal{F}_{Y|Z}$. That is, there are for at least some $z \in \mathcal{R}_Z$, multiple $F_{Y|Z}(\cdot|z)$ satisfying $F_{Y|Z}(\cdot|z) \preccurlyeq \mathcal{Y}(U, Z; h)$ given $Z = z$.

The definition of selectionability of $F_{Y|Z}(\cdot|z)$ from the distribution of $\mathcal{Y}(U, Z; h)$ given $Z = z$ for almost every $z \in \mathcal{R}_Z$ characterizes precisely those distributions for which $h(Y, Z, U) = 0$ can hold with probability one for the given structure $(h, \mathcal{G}_{U|Z})$. Those distributions $F_{Y|Z}(\cdot|z)$ that are selectionable with respect to the conditional distribution of $\mathcal{Y}(U, Z; h)$ when $U \sim G_{U|Z}(\cdot|z)$ are precisely those conditional distributions that can be generated by the structure $(h, \mathcal{G}_{U|Z})$.

This leads to the following definitions of potential observational equivalence and of observational equivalence with respect to a particular collection of conditional distributions $\mathcal{F}_{Y|Z}$.

Definition 3 *Under Restrictions A1-A3, two structures $(h, \mathcal{G}_{U|Z})$ and $\left(h', \mathcal{G}'_{U|Z}\right)$ are **potentially observationally equivalent** if there exists a collection of conditional distributions $\mathcal{F}_{Y|Z}$ such that $F_{Y|Z}(\cdot|z) \preccurlyeq \mathcal{Y}(U, z; h)$ when $U \sim G_{U|Z}(\cdot|z)$ and $F_{Y|Z}(\cdot|z) \preccurlyeq \mathcal{Y}(U, z; h')$ when $U \sim G'_{U|Z}(\cdot|z)$ for almost every $z \in \mathcal{R}_Z$. Two structures $(h, \mathcal{G}_{U|Z})$ and $\left(h', \mathcal{G}'_{U|Z}\right)$ are **observationally equivalent** with respect to $\mathcal{F}_{Y|Z} = \left\{F_{Y|Z}(\cdot|z) : z \in \mathcal{R}_Z\right\}$ if $F_{Y|Z}(\cdot|z) \preccurlyeq \mathcal{Y}(U, z; h)$ when $U \sim G_{U|Z}(\cdot|z)$ and $F_{Y|Z}(\cdot|z) \preccurlyeq \mathcal{Y}(U, z; h')$ when $U \sim G'_{U|Z}(\cdot|z)$ for almost every $z \in \mathcal{R}_Z$.*

The closely related definition of the identified set of structures $(h, \mathcal{G}_{U|Z})$ is as follows.

Definition 4 *Under Restrictions A1-A4, the **identified set** of structures $(h, \mathcal{G}_{U|Z})$ with respect to the collection of distributions $\mathcal{F}_{Y|Z}$ are those admissible structures such that the conditional distributions $F_{Y|Z}(\cdot|z) \in \mathcal{F}_{Y|Z}$ are selectionable with respect to the conditional distributions of random set $\mathcal{Y}(U, z; h)$ when $U \sim G_{U|Z}(\cdot|z)$, a.e. $z \in \mathcal{R}_Z$:*

$$\mathcal{I}\left(\mathcal{M}, \mathcal{F}_{Y|Z}, \mathcal{R}_Z\right) \equiv \left\{(h, \mathcal{G}_{U|Z}) \in \mathcal{M} : F_{Y|Z}(\cdot|z) \preccurlyeq \mathcal{Y}(U, z; h) \ \text{when} \ U \sim G_{U|Z}(\cdot|z), \ \text{a.e.} \ z \in \mathcal{R}_Z\right\}. \tag{15}$$

Selectionability of observed conditional distributions from the random outcome set $\mathcal{Y}(U, z; h)$ provides a convenient and extremely general characterization of identified sets in a broad class of econometric models.

The task that remains in an identification analysis of any particular model is to characterize all observable implications of selectionability in a way that is tractable for use in practice. Any collection of observable implications that fully characterize selectionability will suffice. For example, Beresteanu, Molchanov, and Molinari (2011) (referred to later as BMM11) shows how one can cast selectionability in terms of the support function of the Aumann Expectation of the random outcome set in order to characterize identified sets in a particular class of econometric models.

Given Definition 4 of the identified set of structures admitted by a model, we can now define set identification of structural features. As is commonly done, we define a structural feature $\psi(\cdot, \cdot)$ as any functional of a structure $(h, \mathcal{G}_{U|Z})$. Examples include the structural function h itself, $\psi(h, \mathcal{G}_{U|Z}) = h$, the distributions of unobserved heterogeneity, $\psi(h, \mathcal{G}_{U|Z}) = \mathcal{G}_{U|Z}$, and counterfactual probabilities such as the probability that a component of Y exceeds a given threshold conditional on $Z = z$ calculated using $G_{U|Z}(\cdot|z)$.

Definition 5 *The **identified set of structural features** $\psi(\cdot, \cdot)$ under Restrictions A1-A4 is*

$$\Psi \equiv \left\{ \psi\left(h, \mathcal{G}_{U|Z}\right) : \left(h, \mathcal{G}_{U|Z}\right) \in \mathcal{I}\left(\mathcal{M}, \mathcal{F}_{Y|Z}, \mathcal{R}_Z\right) \right\}.$$

Depending on the context, a variety of different features may be of interest. The identified set of structures $\mathcal{I}\left(\mathcal{M}, \mathcal{F}_{Y|Z}, \mathcal{R}_Z\right)$ can be used to ascertain the identified set of any such feature. The identified set of structures $\mathcal{I}\left(\mathcal{M}, \mathcal{F}_{Y|Z}, \mathcal{R}_Z\right)$ is thus the focus of analysis, and unless a particular feature of interest is specified, reference to only the "identified set" without qualification refers to $\mathcal{I}\left(\mathcal{M}, \mathcal{F}_{Y|Z}, \mathcal{R}_Z\right)$.

A key component of econometric models are restrictions on the joint distribution of U and Z. The use of the Aumann Expectation of random outcome set $\mathcal{Y}(U, z; h)$ and associated support function dominance criteria can be convenient in models with conditional mean restrictions, as discussed in section 5 of Beresteanu, Molchanov, and Molinari (2012). In models with $G_{U|Z}(\cdot|z)$ parametrically specified, this approach or a capacity functional characterization of selectionability has been used, see e.g. BMM11 or the related characterization of Galichon and Henry (2011).

In the following Section we show that the characterizations of observational equivalence and the identified set $\mathcal{I}\left(\mathcal{M}, \mathcal{F}_{Y|Z}, \mathcal{R}_Z\right)$ given in Definitions 3 and 4 can equivalently be expressed in terms of selectionability of $G_{U|Z}(\cdot|z)$ relative to the random residual set $\mathcal{U}(Y, Z; h)$. These characterizations in terms of sets on the support of unobserved heterogeneity enable consideration of all conceivable restrictions on the conditional distributions $\mathcal{G}_{U|Z}$.

4.3 Observational Equivalence via Selectionability in U-Space

In CR17 it is shown that the dual relation between the level sets, $\mathcal{Y}(u, z; h)$ and $\mathcal{U}(y, z; h)$ leads to a dual relation between random outcome sets $\mathcal{Y}(U, Z; h)$ and random residual sets $\mathcal{U}(Y, Z; h)$ which leads to Theorem 1 relating selectionability of $F_{Y|Z}(\cdot|z)$ with respect to the distribution of $\mathcal{Y}(U, Z; h)$ and selectionability of $G_{U|Z}(\cdot|z)$ with respect to the distribution of $\mathcal{U}(Y, Z; h)$, all conditional on $Z = z$.

Theorem 1 *Let Restrictions A1-A3 hold. Then for any $z \in \mathcal{R}_Z$, $F_{Y|Z}(\cdot|z)$ is selectionable with respect to the conditional distribution of $\mathcal{Y}(U, Z; h)$ given $Z = z$ when $U \sim G_{U|Z}(\cdot|z)$ if and only*

if $G_{U|Z}(\cdot|z)$ is selectionable with respect to the conditional distribution of $\mathcal{U}(Y, Z; h)$ given $Z = z$ when $Y \sim F_{Y|Z}(\cdot|z)$.

Proof: See CR17 Theorem 1. $\square$

This results follows directly from the duality of the level sets $\mathcal{Y}(U, Z; h)$ and $\mathcal{U}(Y, Z; h)$ set out in Section 2.3. When considering selectionability and in other contexts the roles of observed endogenous Y and its probability distribution and unobserved U and its probability distribution can be interchanged.

With Theorem 1 established, we now characterize the identified set of structures in terms of random variables and sets in the space of unobserved heterogeneity. A key benefit that comes from this is that it allows imposition of restrictions directly on the distributions of unobservable variables $G_{U|Z}$ through specification of the class $\mathcal{G}_{U|Z}$ admitted by the model $\mathcal{M}$. One can then check whether any such $G_{U|Z} \in \mathcal{G}_{U|Z}$ are selectionable with respect to the identified conditional distributions of random set $\mathcal{U}(Y, Z; h)$, given identification of the conditional distributions $\mathcal{F}_{Y|Z}$ under Restriction A2. That is, in the context of any particular model, events concerning this random set can be expressed as events involving observable variables, as we illustrate in the examples of Sections 7 and 8.

Theorem 2 *Let Restrictions A1-A3 hold. Then two structures $\left(h, \mathcal{G}_{U|Z}\right)$ and $\left(h^*, \mathcal{G}^*_{U|Z}\right)$ are observationally equivalent with respect to $\mathcal{F}_{Y|Z}$ if and only if $G_{U|Z}(\cdot|z)$ and $G^*_{U|Z}(\cdot|z)$ are selectionable with respect to the conditional (on $Z = z$) distributions of random sets $\mathcal{U}(Y, Z; h)$ and $\mathcal{U}(Y, Z; h^*)$, respectively, a.e. $z \in \mathcal{R}_Z$. If, additionally, Restriction A4 holds, then the identified set of structures $\left(h, \mathcal{G}_{U|Z}\right)$ are those elements of $\mathcal{M}$ such that $G_{U|Z}(\cdot|z)$ is selectionable with respect to the conditional (on $Z = z$) distribution of random set $\mathcal{U}(Y, Z; h)$, a.e. $z \in \mathcal{R}_Z$.*

Proof: See CR17 Theorem 2. $\square$

In summary: from Definition 4, it follows that any characterization of the set of structures $\left(h, \mathcal{G}_{U|Z}\right)$ such that $F_{Y|Z}(\cdot|z)$ is selectionable with respect to the conditional distribution of $\mathcal{Y}(U, Z; h)$ given $Z = z$ almost surely also characterizes the identified set $\mathcal{I}\left(\mathcal{M}, \mathcal{F}_{Y|Z}, \mathcal{R}_Z\right)$. Theorem 2 uses a duality relation between random outcome sets $\mathcal{Y}(U, Z; h)$ and random residual sets $\mathcal{U}(Y, Z; h)$ to express observational equivalence and characterization of the identified set of structures $\left(h, \mathcal{G}_{U|Z}\right)$ in terms of selectionability with respect to the conditional distribution of $\mathcal{U}(Y, Z; h)$. With this in hand, any conditions that characterize the set of $\left(h, \mathcal{G}_{U|Z}\right)$ such that $G_{U|Z}(\cdot|z)$ is selectionable with respect to the conditional distribution of $\mathcal{U}(Y, Z; h)$ will suffice for characterization of the identified set.

5   Observable Implications of Selectionability

A common starting point for identification analysis is to develop some observable implications of a structural model, which place restrictions on those structures that are feasible given knowledge of the distributions of observed endogenous Y given observed exogenous Z, $\mathcal{F}_{Y|Z}$. Such implications may for example take the form of moment equations or inequality restrictions. They are called *observable* implications because they involve the probability distribution of observable variables, which is identified under a suitable sampling process.

Section 5.1 below lays out some observable implications of a structural model that can be derived from quite elementary considerations, without calling on any results from random set theory. Restrictions produced in this way provide bounds on the set of possible structures $\left(h, \mathcal{G}_{U|Z}\right)$ compatible with $\mathcal{F}_{Y|Z}$.

Specifically we show that if a structure $\left(h, \mathcal{G}_{U|Z}\right)$ is compatible with $\mathcal{F}_{Y|Z}$ then, conditional on $Z = z$, the probability that unobserved U lies in a set $\mathcal{S}$ cannot be smaller than the conditional on $Z = z$ probability of occurrence of those values of Y that can *only* occur when U lies in $\mathcal{S}$. Under Restriction 5 this must hold for all closed sets $\mathcal{S}$ on the support of U and for almost every z on the support of Z.

A key result from random set theory, Artstein's Inequality – see e.g. Artstein (1983), Norberg (1992), and Molchanov (2005, Section 1.4.8) – establishes that suitably rich collections of implications of this form in fact deliver *sharp* bounds on the set of structures $\left(h, \mathcal{G}_{U|Z}\right)$, equivalently, that they characterize the identified set $\mathcal{I}\left(\mathcal{M}, \mathcal{F}_{Y|Z}, \mathcal{R}_Z\right)$. The observable implications that are derived in Section 5.1 are inequalities and equalities involving conditional probabilities of observable events. These are particular kinds of conditional moment inequalities and equalities, since the probability of any event may be expressed as the expectation of the indicator of that event.

Section 5.2 then introduces the notion of core-determining sets. The characterization of an identified set of structures given in Section 5.1 comprises a system of inequalities in probabilities obtained by considering every closed set $\mathcal{S}$ on the support of U. The core determining sets comprise a smaller collection of sets such that if the probability inequalities hold for all core determining sets then they hold for all closed sets on the support of U.

In the next section we start by presenting some observable implications of GIV models which can be deduced from elementary considerations of the relative magnitudes of the probabilities of certain events. Results from random set theory can be deployed to show that sufficiently rich collections of such observable implications characterize identified sets of structures in GIV models. We start with this so that the basic idea underlying this characterization of identified sets can be appreciated by readers who have no knowledge of random set theory.

5.1 Conditional Probability Inequalities

Let $\left(h, \mathcal{G}_{U|Z}\right) \in \mathcal{M}$ denote a particular structure under consideration. Suppose that for some random vector U with conditional distributions $\mathcal{G}_{U|Z}$, $h\left(Y, Z, U\right) = 0$ almost surely. What properties would the identified conditional distributions $\mathcal{F}_{Y|Z}$ then have to satisfy?

On the way to answering this question, consider first what observable variables (Y, Z) imply about the unobservable variables U. This is useful to consider because the structural model places restrictions on the conditional distributions of U given Z. If the data is generated by a process with structure $\left(h, \mathcal{G}_{U|Z}\right)$, then the conditional distributions of U given Z must correspond to those elements of the collection of conditional distributions $\mathcal{G}_{U|Z}$. Given observed (Y, Z), if the structural function is h, then with probability one $h\left(Y, Z, U\right) = 0$, equivalently $U \in \mathcal{U}\left(Y, Z; h\right)$. Conditional on any realization of Z, the realization of the set $\mathcal{U}\left(Y, Z; h\right)$ is determined by the realization of Y. Thus, conditional on $Z = z$, the conditional distribution $F_{Y|Z}\left(\cdot | z\right)$ induces a probability distribution of sets $\mathcal{U}\left(Y, Z; h\right)$ given $Z = z$.

This is illustrated for a case in which Y is discrete and for a particular z in Figure 1. Restricting Y to be discrete is convenient for exposition, but not necessary. The figure illustrates a setting in which $\mathcal{R}_U = \mathbb{R}^2$ and the support of Y conditional on $Z = z$ is given by $\mathcal{R}_{Y|z} = \left\{y^*, y^\$, y^\#, y^+\right\}$. For each such y, the set $\mathcal{U}\left(y, z; h\right)$ is shown in the figure. Additionally, a set $\mathcal{S}$ of values of U is represented as an ellipse in $\mathbb{R}^2$. The sets

$$\left\{\mathcal{U}\left(y, z; h\right) : y \in \mathcal{R}_{Y|z}\right\} = \left\{\mathcal{U}\left(y^*, z; h\right), \mathcal{U}\left(y^\$, z; h\right), \mathcal{U}\left(y^\#, z; h\right), \mathcal{U}\left(y^+, z; h\right)\right\}$$

comprise the support of the random set $\mathcal{U}\left(Y, Z; h\right)$ conditional on $Z = z$. The set $\mathcal{S}$ is a fixed set in $\mathcal{R}_U$, and it could be chosen to be any set at all. Given such a set $\mathcal{S}$, consider those sets $\mathcal{U}\left(y, z; h\right)$ such that $\mathcal{U}\left(y, z; h\right) \subseteq \mathcal{S}$. Here these are the sets $\mathcal{U}\left(y^\#, z; h\right)$ and $\mathcal{U}\left(y^\$, z; h\right)$, and so

$$\mathbb{P}\left[\mathcal{U}\left(Y, Z; h\right) \subseteq \mathcal{S} | z\right] = \mathbb{P}\left[Y \in \left\{y \in \mathcal{R}_Y : \mathcal{U}\left(y, z; h\right) \subseteq \mathcal{S}\right\} | z\right] = \mathbb{P}\left[Y \in \left\{y^\$, y^\#\right\} | z\right].$$

The conditional probability that the event $\mathcal{U}\left(Y, Z; h\right) \subseteq \mathcal{S}$ occurs can be written as the conditional probability that Y belongs to the set of values of y such that $\mathcal{U}\left(y, z; h\right) \subseteq \mathcal{S}$. This is the set of values of Y that can only occur when U takes a value in $\mathcal{S}$. For any structural function h, and for a given fixed set $\mathcal{S}$, this conditional probability is known given knowledge of $\mathcal{F}_{Y|Z}$, which is point identified.

The observation that, for any particular specification of the structural function h and for any given set $\mathcal{S}$, the probability $\mathbb{P}\left[\mathcal{U}\left(Y, Z; h\right) \subseteq \mathcal{S} | z\right]$ is known from knowledge of $\mathcal{F}_{Y|Z}$ holds in general, and does not require a model that gives rise to the particular illustration shown in Figure 1. Regardless of the model under consideration, the set of values y such that $\mathcal{U}\left(y, z; h\right) \subseteq \mathcal{S}$ can be

defined for any h, $\mathcal{S}$, and z as

$$\mathcal{A}\left(\mathcal{S}, z; h\right) \equiv \left\{y \in \mathcal{R}_Y : \mathcal{U}\left(y, z; h\right) \subseteq \mathcal{S}\right\}.$$

Then there is the equivalence

$$\mathbb{P}\left[\mathcal{U}\left(Y, Z; h\right) \subseteq \mathcal{S}|z\right] = \mathbb{P}\left[Y \in \mathcal{A}\left(\mathcal{S}, z; h\right)|z\right].$$

The set $\mathcal{A}\left(\mathcal{S}, z; h\right)$ contains the values of Y that structural function h says can only occur when U takes a value in the set $\mathcal{S}$ when $Z = z$.

What does knowledge of this probability allow one to learn about $\left(h, \mathcal{G}_{U|Z}\right)$? The event $\mathcal{U}\left(Y, Z; h\right) \subseteq \mathcal{S}$ carries an important implication for the realization of U. If $\mathcal{U}\left(Y, Z; h\right) \subseteq \mathcal{S}$ and the structural function is indeed h, then it must be that $U \in \mathcal{U}\left(Y, Z; h\right)$ and so $U \in \mathcal{S}$, that is:

$$\left\{\mathcal{U}\left(Y, Z; h\right) \subseteq \mathcal{S}\right\} \implies \left\{U \in \mathcal{S}\right\}. \tag{16}$$

Define the function

$$C_h\left(\mathcal{S}|z\right) \equiv \mathbb{P}\left[\mathcal{U}\left(Y, Z; h\right) \subseteq \mathcal{S}|z\right] = \mathbb{P}\left[Y \in \mathcal{A}\left(\mathcal{S}, z; h\right)|z\right]$$

which is the conditional probability of the event on the left hand side of (16). If an event $\mathcal{E}_1$ implies another event $\mathcal{E}_2$ then for any measure ν on the measure space on which these events are defined there is the relation $\nu\left(\mathcal{E}_1\right) \leq \nu\left(\mathcal{E}_2\right)$. So, taking $\mathcal{E}_1$ to be the event on the left of (16) and $\mathcal{E}_2$ to be the event on the right of (16) as well as $\nu\left(\cdot\right) = \mathbb{P}\left[\cdot|z\right]$ it follows that

$$C_h\left(\mathcal{S}|z\right) \leq \mathbb{P}\left[U \in \mathcal{S}|z\right] = G_{U|Z}\left(\mathcal{S}|z\right), \tag{17}$$

where the equality follows from Restriction A3. This inequality will be referred to as a (conditional) containment inequality, since $C_h\left(\mathcal{S}|z\right)$ is the conditional probability that $\mathcal{U}\left(Y, Z; h\right)$ is contained in $\mathcal{S}$.[16]

The inequality (17) is an observable implication for structure $\left(h, \mathcal{G}_{U|Z}\right)$. The probability on the left hand side of the above inequality is identified for any h but will differ with h because changing h changes the set of values u for which $h\left(Y, Z, u\right) = 0$, namely $\mathcal{U}\left(Y, Z; h\right)$. If data are generated by a process in which the structure is $\left(h, \mathcal{G}_{U|Z}\right)$, then the inequality (17) must hold. The set $\mathcal{S}$ and the value of the conditioning variable z were both chosen arbitrarily, so that (17) must hold for any $\mathcal{S}$ and almost every $z \in \mathcal{R}_Z$ if the structure $\left(h, \mathcal{G}_{U|Z}\right)$ is embodied in the actual data generation process.

---

[16]In random set theory $C_h(\mathcal{S}|z)$ is the conditional containment functional, giving the probability that a random set (in this case $\mathcal{U}(Y, Z; h)$) is a subset of the set $\mathcal{S}$ conditional on $Z = z$.

There is for any $\mathcal{S}$ another observable implication, namely that the probability that U is in $\mathcal{S}$ is no greater than the probability that the set $\mathcal{U}(Y, Z; h)$ intersects – or hits – $\mathcal{S}$. This is because if h is the structural function and $\mathcal{U}(Y, Z; h)$ does not hit $\mathcal{S}$, then U cannot be in $\mathcal{S}$, since $U \in \mathcal{U}(Y, Z; h)$ almost surely. Using this implication there is

$$\mathbb{P}\left[\mathcal{U}(Y, Z; h) \cap \mathcal{S} = \emptyset | z\right] \leq \mathbb{P}\left[U \notin \mathcal{S} | z\right],$$

where $\emptyset$ denotes the empty set, equivalently

$$G_{U|Z}(\mathcal{S}|z) = \mathbb{P}\left[U \in \mathcal{S} | z\right] \leq \overline{C_h}(\mathcal{S}|z), \tag{18}$$

where

$$\overline{C_h}(\mathcal{S}|z) \equiv \mathbb{P}\left[\mathcal{U}(Y, Z; h) \cap \mathcal{S} \neq \emptyset | z\right]$$

is the conditional probability that $\mathcal{U}(Y, Z; h)$ hits the set $\mathcal{S}$. For any structural function h the conditional hitting probability is, like the conditional containment probability, known given knowledge of $\mathcal{F}_{Y|Z}$. In the example depicted in Figure 1 the implication is that $\mathbb{P}\left[U \in \mathcal{S} | z\right] \leq \mathbb{P}\left[Y \in \left\{y^{\$}, y^{\#}, y^{+}\right\} | z\right]$. This probability, $\mathbb{P}\left[\mathcal{U}(Y, Z; h) \cap \mathcal{S} \neq \emptyset | z\right]$, taken as a function of argument $\mathcal{S}$, is sometimes referred to as a capacity functional for $\mathcal{U}(Y, Z; h)$, and so it is appropriate to refer to the inequality $G_U(\mathcal{S}|z) \leq \overline{C_h}(\mathcal{S}|z)$ in (18) above as a capacity inequality.

With (17) and (18) one can therefore produce observable implications conditional on each z for as many sets $\mathcal{S}$ as one is willing to consider. All such observable implications must hold, but some of them may be redundant. For example, for any set $\mathcal{S}$, the inequality (17) is equivalent to the inequality (18) applied to $\mathcal{S}^c$, the complement of $\mathcal{S}$, since

$$1 - \overline{C_h}(\mathcal{S}^c|z) = C_h(\mathcal{S}|z) \leq G_U(\mathcal{S}|z) = 1 - G_U(\mathcal{S}^c|z).$$

Because of this equivalence between inequalities generated by (17) and (18) using sets $\mathcal{S}$ and $\mathcal{S}^c$, respectively, this chapter focuses without loss of generality on characterizations that use containment probabilities, represented as (17).

Any collection S of sets $\mathcal{S}$ delivers bounds on permissible $(h, \mathcal{G}_{U|Z})$ giving a set

$$\mathcal{M}_0(\mathsf{S}) \equiv \left\{(h, \mathcal{G}_{U|Z}) : \forall \mathcal{S} \in \mathsf{S}, \quad C_h(\mathcal{S}|z) \leq G_U(\mathcal{S}|z) \text{ a.e. } z \in \mathcal{R}_Z\right\}, \tag{19}$$

So far, it has been reasoned that $(h, \mathcal{G}_{U|Z})$ must lie in the sets $\mathcal{M}_0(\mathsf{S})$ produced by any collection of sets S. The question remains: do any of these sets constitute *the identified set*, i.e. sharp bounds on $(h, \mathcal{G}_{U|Z})$?

A result from Artstein (1983) using random set theory answers this question.[17] This result,

<hr/>

[17]See Norberg (1992) for a generalization of Artstein's result, as well as Molchanov (2005, Section 1.4.8).

which we refer to as Artstein's Inequality, can be applied if Restriction A5 also holds, that is if the random set $\mathcal{U}(Y, Z; h)$ is closed. Its application enables a characterization of the identified set $\mathcal{I}(\mathcal{M}, \mathcal{F}_{Y|Z}, \mathcal{R}_Z)$ given in Theorem 2 through the conditional containment probabilities of random set $\mathcal{U}(Y, Z; h)$, which is precisely the conditional probability $C_h(\mathcal{S}|z)$ appearing on the left of the observable implication (17).

Characterization via the containment probabilities produces an expression for $\mathcal{I}(\mathcal{M}, \mathcal{F}_{Y|Z}, \mathcal{R}_Z)$ in the form of inequality restrictions, as given in the following Corollary.

Corollary 1 *Under Restrictions A1-A5 the identified set can be written*

$$\mathcal{I}(\mathcal{M}, \mathcal{F}_{Y|Z}, \mathcal{R}_Z) \equiv \left\{ (h, \mathcal{G}_{U|Z}) \in \mathcal{M} : \forall \mathcal{S} \in \mathsf{F}(\mathcal{R}_U), \; C_h(\mathcal{S}|z) \leq G_{U|Z}(\mathcal{S}|z), \; a.e. \; z \in \mathcal{R}_Z \right\}, \quad (20)$$

where $\mathsf{F}(\mathcal{R}_U)$ *denotes the collection of all closed subsets of* $\mathcal{R}_U$.

Proof: Proven as Corollary 1 of CR17. $\square$

Corollary 1 tells us that the set $\mathcal{M}_0(\mathsf{S})$ defined in (19) with the collection $\mathsf{S} = \mathsf{F}(\mathcal{R}_U)$ comprising all closed sets on $\mathcal{R}_U$ is sharp. Since a conditional probability is simply the conditional expectation of an indicator function, i.e.

$$C_h(\mathcal{S}|z) = \mathbb{P}[\mathcal{U}(Y, Z; h) \subseteq \mathcal{S}|z] = \mathbb{E}[1[\mathcal{U}(Y, Z; h) \subseteq \mathcal{S}]|z],$$

Corollary 1 expresses the selectionability requirement for characterization of the identified set as a collection of conditional moment inequalities. The inequalities in this characterization are for almost every value of the instrument $z \in \mathcal{R}_Z$ as well as all closed *test sets* $\mathcal{S}$ on $\mathcal{R}_U$. The inequality

$$C_h(\mathcal{S}|z) \leq G_{U|Z}(\mathcal{S}|z) \quad (21)$$

follows immediately from the fact that U is, by virtue of $h(Y, Z, U) = 0$, a measurable selection of $\mathcal{U}(Y, Z; h)$. Artstein's inequality establishes that the inequality holding for all $\mathcal{S} \in \mathsf{F}(\mathcal{R}_U)$ guarantees selectionability of $G_{U|Z}(\cdot|z)$ from the conditional distribution of $\mathcal{U}(Y, Z; h)$ given $Z = z$, a.e. $z \in \mathcal{R}_Z$.

If Restriction A6 holds, Artstein's inequality can also be used to establish the alternative formulation

$$\mathcal{I}(\mathcal{M}, \mathcal{F}_{Y|Z}, \mathcal{R}_Z) = \left\{ (h, \mathcal{G}_{U|Z}) \in \mathcal{M} : \forall \mathcal{K} \in \mathsf{K}(\mathcal{R}_Y), \; F_{Y|Z}(\mathcal{K}|z) \leq G_{U|Z}(\mathcal{Y}(U, z; h) \cap \mathcal{K} \neq \emptyset|z), \; \text{a.e.} \; z \in \mathcal{R}_Z \right\},$$
$$(22)$$

where $\mathsf{K}(\mathcal{R}_Y)$ denotes the collection of compact subsets of $\mathcal{R}_Y$. The key implication $F_{Y|Z}(\mathcal{K}|z) \leq G_{U|Z}(\mathcal{Y}(U, z; h) \cap \mathcal{K} \neq \emptyset|z)$ has formed the basis of representations of identified sets for various incomplete models studied in Appendix D.2 of BMM11 and in Galichon and Henry (2011).

Corollary 1 offers a convenient formulation of selectionability for generalized instrumental variable models. The inequality $C_h\left(\mathcal{S}|z\right) \leq G_{U|Z}\left(\mathcal{S}|z\right)$ isolates the conditional distribution of unobservable heterogeneity on the right hand side, with argument given by the fixed set $\mathcal{S}$. This makes it relatively straightforward to impose restrictions on the conditional distribution of unobservable heterogeneity common to instrumental variable models. Incorporation of such restrictions and the simplifications in the characterization of $\mathcal{I}\left(\mathcal{M}, \mathcal{F}_{Y|Z}, \mathcal{R}_Z\right)$ which they can sometimes afford are considered in Section 6. Direct computation or approximation by simulation of probabilities of events determined by the distribution of the random set $\mathcal{Y}\left(U, z; h\right)$ are not required using our approach.

Moreover, depending on the particular restrictions employed, it is often possible to characterize a collection of sets, say Q, which is a subset of the collection of all closed sets on $\mathcal{R}_U$, such that requiring (21) for each set $\mathcal{S} \in \mathsf{Q}$ is sufficient to characterize the identified set. This occurs when the restrictions placed on h can be used to establish that the inequality (21) holding for every $\mathcal{S} \in \mathsf{Q}$ implies that inequality (21) in fact holds for all closed sets $\mathcal{S} \in \mathsf{F}\left(\mathcal{R}_U\right)$. This means that the observable implication (21) for all $\mathcal{S} \in \mathsf{Q}$ in fact implies all observable implications that one could derive. Consequently, the identified set of structures can then be characterized as those $\left(h, \mathcal{G}_{U|Z}\right)$ such that (21) holds for all $\mathcal{S} \in \mathsf{Q}$ rather than all $\mathcal{S} \in \mathsf{F}\left(\mathcal{R}_U\right)$, leading to a more concise characterization with no loss of sharpness. The collection of inequalities involved can be vast, and such reductions can be significant. Such collections of sets Q are referred to as core determining sets. These are the topic of the next subsection.

5.2 Core Determining Collections

Any collection of sets $\mathsf{Q}(h, z)$ such that

$$\left\{\forall \mathcal{S} \in \mathsf{Q}(h, z), \quad C_h\left(\mathcal{S}|z\right) \leq G_{U|Z}\left(\mathcal{S}|z\right)\right\} \implies \left\{\forall \mathcal{S} \in \mathsf{F}\left(\mathcal{R}_U\right), \quad C_h\left(\mathcal{S}|z\right) \leq G_{U|Z}\left(\mathcal{S}|z\right)\right\}$$

is referred to as a *core determining $\mathcal{U}$ collection*. There may be different collections $\mathsf{Q}(h, z)$ and $\mathsf{Q}'(h, z)$ which are both core-determining. As indicated by the notation, these sets may be specific to a given structural function h and value of the conditioning variables z. This section lays out characterizations of such core determining collections. Conditions whereby there exists a unique minimal core-determining collection appear to be not yet established.

The notion of core-determining collections was introduced by Galichon and Henry (2011). Their analysis focused on collections of sets of outcomes, which are subsets of $\mathcal{R}_Y$. To distinguish them from core determining $\mathcal{U}$ collections, such a collection T will be referred to as a *core determining $\mathcal{Y}$ collection*.

Proposition 1 of Galichon and Henry (2011) establishes that if T is a core determining $\mathcal{Y}$

collection for all h and almost every z, then the identified set is given by

$$\mathcal{I}\left(\mathcal{M}, \mathcal{F}_{Y|Z}, \mathcal{R}_Z\right) = \left\{(h, \mathcal{G}_{U|Z}) \in \mathcal{M} : \forall \mathcal{T} \in \mathsf{T},\, F_{Y|Z}\left(\mathcal{T}|z\right) \leq G_{U|Z}\left(\mathcal{Y}\left(U, z; h\right) \cap \mathcal{T} \neq \emptyset|z\right),\, \text{a.e. } z \in \mathcal{R}_Z\right\},$$

which is (22), but with the collection of all compact sets replaced by the core determining $\mathcal{Y}$ collection T. A characterization of such a core determining $\mathcal{Y}$ collection was provided under a monotonicity relation between ordered pairs of values of unobservables (u_1, u_2) and associated sets of feasible outcomes $\mathcal{Y}\left(u_1, Z; h\right)$ and $\mathcal{Y}\left(u_2, Z; h\right)$, and was shown to apply to models of family bargaining and oligopoly entry with two types of firms previously studied by Engers and Stern (2002) and Berry and Tamer (2007), respectively.

Corollary 1 provides an alternative characterization of the identified set using observational equivalence characterized by selectionability with respect to the distribution of $\mathcal{U}\left(Y, Z; h\right)$. This characterization employs a collection of inequalities $C_h\left(\mathcal{S}|z\right) \leq G_{U|Z}\left(\mathcal{S}|z\right)$ for all $\mathcal{S} \in \mathsf{F}\left(\mathcal{R}_U\right)$, which are sets of values in the space of unobservable variables, $\mathcal{R}_U$, rather than than the space of endogenous outcomes, $\mathcal{R}_Y$. To characterize a more concise collection of such inequalities that provide a sharp characterization of $\mathcal{I}\left(\mathcal{M}, \mathcal{F}_{Y|Z}, \mathcal{R}_Z\right)$ requires the development of core determining $\mathcal{U}$ collections, a task taken up in Chesher and Rosen (2017a). Such core determining collections $\mathsf{Q}(h, z)$ are now characterized.

For this development it is useful to define the support of the random set $\mathcal{U}\left(Y, Z; h\right)$ conditional on $Z = z$, and the collection of sets comprising unions of such sets, which are

$$\mathsf{U}\left(h, z\right) \equiv \left\{\mathcal{U} \subseteq \mathcal{R}_U : \exists y \in \mathcal{R}_{Y|z} \text{ such that } \mathcal{U} = \mathcal{U}\left(y, z; h\right)\right\}.$$

and

$$\mathsf{U}^*\left(h, z\right) \equiv \left\{\mathcal{U} \subseteq \mathcal{R}_U : \exists \mathcal{Y} \subseteq \mathcal{R}_{Y|z} \text{ such that } \mathcal{U} = \mathcal{U}\left(\mathcal{Y}, z; h\right)\right\},$$

respectively. For ease of reference, these objects and others used in this section are collected in Table 1. The following slight abuse of notation is used throughout:

$$\forall \mathcal{Y} \subseteq \mathcal{R}_{Y|z}, \qquad \mathcal{U}\left(\mathcal{Y}, z; h\right) \equiv \bigcup_{y \in \mathcal{Y}} \mathcal{U}\left(y, z; h\right).$$

That is, $\mathcal{U}\left(\mathcal{Y}, z; h\right)$ is the union of sets $\mathcal{U}\left(y, z; h\right)$ such that $y \in \mathcal{Y}$.

As stated in Lemma 1, below, in order for the containment inequality (21) to hold for any given (h, z) and all closed $\mathcal{S} \subseteq \mathcal{R}_U$, it suffices that it hold for all $\mathcal{S} \in \mathsf{U}^*\left(h, z\right)$. For any set $\mathcal{S} \subseteq \mathcal{R}_U$ and any (h, z), define

$$\mathsf{U}^{\mathcal{S}}\left(h, z\right) \equiv \left\{\mathcal{U} \in \mathsf{U}\left(h, z\right) : \mathcal{U} \subseteq \mathcal{S}\right\},$$

which are the sets $\mathcal{U} \in \mathsf{U}\left(h, z\right)$ that are contained in $\mathcal{S}$.

Collection	Description
$\mathsf{U}\left(h,z\right)$	Support of $\mathcal{U}\left(Y,Z;h\right)$ conditional on $Z=z$.
$\mathsf{U}^{*}\left(h,z\right)$	Sets that are unions of sets in $\mathsf{U}\left(h,z\right)$.
$\mathsf{U}^{\mathcal{S}}\left(h,z\right)$	Sets in $\mathsf{U}\left(h,z\right)$ that are contained in $\mathcal{S}$.

Table 1: Notation for collections of subsets of $\mathcal{R}_U$ used in the development of core determining sets.

Lemma 1 *Let Restrictions A1-A3 hold. Let $z \in \mathcal{R}_Z$, $h \in \mathcal{H}$, and $\mathcal{S} \subseteq \mathcal{R}_U$. Let $\mathcal{U}_{\mathcal{S}}\left(h,z\right)$ denote the union of all sets in $\mathsf{U}^{\mathcal{S}}\left(h,z\right)$,*

$$\mathcal{U}_{\mathcal{S}}\left(h,z\right) \equiv \bigcup_{\mathcal{U}\in\mathsf{U}^{\mathcal{S}}\left(h,z\right)} \mathcal{U}. \tag{23}$$

If

$$C_h\left(\mathcal{U}_{\mathcal{S}}\left(h,z\right)|z\right) \leq G_{U|Z}\left(\mathcal{U}_{\mathcal{S}}\left(h,z\right)|z\right),$$

then

$$C_h\left(\mathcal{S}|z\right) \leq G_{U|Z}\left(\mathcal{S}|z\right). \tag{24}$$

Proof: Proven as Lemma 1 of CR17. $\square$

Lemma 1 establishes that if the containment inequality (24) holds for all $\mathcal{S} \in \mathsf{U}^{*}\left(h,z\right)$, then it holds for all sets $\mathcal{S} \subseteq \mathcal{R}_U$.

The following result now defines a collection of core-determining test sets $\mathsf{Q}\left(h,z\right)$, which is a refinement of $\mathsf{U}^{*}\left(h,z\right)$. It provides conditions whereby certain sets may be discarded from consideration. The sets which can be excluded have the property that they contain two sets that (i) are both members of the collection $\mathsf{Q}\left(h,z\right)$, and (ii) have the property that all sets in $\mathsf{U}\left(h,z\right)$ that are contained in one of these sets have measure zero intersection $G_{U|Z}\left(\cdot|z\right)$ with all sets in $\mathsf{U}\left(h,z\right)$ that are contained in the other.

Theorem 3 *Let Restrictions A1-A3 hold. For any $\left(h,z\right) \in \mathcal{H} \times \mathcal{R}_Z$, let $\mathsf{Q}\left(h,z\right) \subseteq \mathsf{U}^{*}\left(h,z\right)$, such that for any $\mathcal{S} \in \mathsf{U}^{*}\left(h,z\right)$ with $\mathcal{S} \notin \mathsf{Q}\left(h,z\right)$, there exist nonempty collections of sets $\mathsf{S}_1, \mathsf{S}_2$ that partition $\mathsf{U}^{\mathcal{S}}\left(h,z\right)$ such that*

$$\mathcal{S}_1 \equiv \bigcup_{\mathcal{U}\in\mathsf{S}_1} \mathcal{U}, \ \mathcal{S}_2 \equiv \bigcup_{\mathcal{U}\in\mathsf{S}_2} \mathcal{U}, \ \text{and } G_{U|Z}\left(\mathcal{S}_1 \cap \mathcal{S}_2|z\right) = 0, \tag{25}$$

with $\mathcal{S}_1, \mathcal{S}_2 \in \mathsf{Q}\left(h,z\right)$. Then $C_h\left(\mathcal{S}|z\right) \leq G_{U|Z}\left(\mathcal{S}|z\right)$ for all $\mathcal{S} \in \mathsf{Q}\left(h,z\right)$ implies that $C_h\left(\mathcal{S}|z\right) \leq G_{U|Z}\left(\mathcal{S}|z\right)$ holds for all $\mathcal{S} \subseteq \mathcal{R}_U$, and in particular for $\mathcal{S} \in \mathsf{F}\left(\mathcal{R}_U\right)$, so that the collection of sets $\mathsf{Q}\left(h,z\right)$ is core-determining.

Proof: Proven as Theorem 3 of CR17. $\square$

Note that all sets of the form $\mathcal{U}\left(y,z;h\right)$ with $y \in \mathcal{R}_Y$ are contained in $\mathsf{Q}\left(h,z\right)$, so that all sets in $\mathsf{U}\left(h,z\right)$ are elements of $\mathsf{Q}\left(h,z\right)$. Theorem 3 implies that the identified sets of Theorem

2 are characterized by the set of structures $(h, \mathcal{G}_{U|Z})$ that satisfy the containment inequalities of Corollary 1, but with $\mathsf{Q}(h, z)$ replacing $\mathsf{F}(\mathcal{R}_U)$. If, as is the case in many models, the sets in $\mathsf{U}(h, z)$ are each connected with boundary of Lebesgue measure zero, and $G_{U|Z}(\cdot|z)$ is absolutely continuous with respect to Lebesgue measure, then the condition $G_{U|Z}(\mathcal{S}_1 \cap \mathcal{S}_2|z) = 0$ in (25) is implied if the sets $\mathcal{S}_1$ and $\mathcal{S}_2$ have non-overlapping interiors.

To illustrate the results of Theorem 3 in a relatively simple context consider again Example 2 of Section 3.2, also studied in Chesher and Rosen (2013). In that model recall that $\mathcal{U}(y, z; h) = [0, g(y_2, z_1)]$ when $y_1 = 0$ and $\mathcal{U}(y, z; h) = [g(y_2, z_1), 1]$ when $y_1 = 1$. Consider a fixed z and a conjectured structural function h, characterized by the threshold function g. From Lemma 1 it follows that for the containment function inequality characterization of $\mathcal{I}(\mathcal{M}, \mathcal{F}_{Y|Z}, \mathcal{R}_Z)$ in Corollary 1 we need only consider test sets that are unions of sets of the form $[0, g(y_2, z_1)]$ or $[g(y_2, z_1), 1]$, for $y_2 \in \mathcal{R}_{Y_2}$. The union of any collection of sets $\{[0, g(y_2, z_1)] : y_2 \in \mathcal{Y}_2 \subseteq \mathcal{R}_{Y_2}\}$ is simply $[0, \max_{y_2 \in \mathcal{Y}_2} g(y_2, z_1)]$. Likewise, the union of any collection of sets $\{[g(y_2, z_1), 1] : y_2 \in \mathcal{Y}_2 \subseteq \mathcal{R}_{Y_2}\}$ is $[\min_{y_2 \in \mathcal{Y}_2} g(y_2, z_1), 1]$. Thus, all unions of sets of the form $[0, g(y_2, z_1)]$ or $[g(y_2, z_1), 1]$ can be expressed as

$$\mathcal{S} = [0, g(y_2, z_1)] \cup \left[g(y_2', z_1), 1\right], \text{ for some } y_2, y_2' \in \mathcal{R}_{Y_2}. \tag{26}$$

Now consider test sets $\mathcal{S}$ of the form given in (26). If $g(y_2, z_1) \geq g(y_2', z_1)$, then $\mathcal{S} = \mathbb{R}$. This test set can be trivially discarded because in this case the containment inequality (24) is simply $1 \leq G_{U|Z}(\mathcal{R}_U|z)$, which holds by virtue of $G_{U|Z}(\cdot|z)$ being a probability measure on $\mathcal{R}_U$, so that $G_{U|Z}(\mathcal{R}_U|z) = 1$. If instead $g(y_2, z_1) < g(y_2', z_1)$, then $\mathcal{S} = [0, g(y_2, z_1)] \cup [g(y_2', z_1), 1]$ is such that $G_{U|Z}(\mathcal{S}_1 \cap \mathcal{S}_2|z) = 0$. We can then apply Theorem 3 with $\mathcal{S}_1 = [0, g(y_2, z_1)]$ and $\mathcal{S}_2 = [g(y_2', z_1), 1]$ to conclude that as long as $\mathcal{S}_1$ and $\mathcal{S}_2$ are included in the collection of core-determining sets $\mathsf{Q}(h, z)$, $\mathcal{S}$ need not be included in $\mathsf{Q}(h, z)$. Thus it suffices to consider all $\mathcal{S} \in \mathsf{Q}(h, z)$ given by the collection of intervals of the form $[0, g(y_2, z_1)]$ or $[g(y_2, z_1), 1]$ for some $y_2 \in \mathcal{R}_{Y_2}$.

Corollary 2 below shows that in some models certain of the containment inequalities for core-determining sets can be replaced by equalities.[18] Then the identified set can be written as a collection of conditional moment inequalities and equalities. There are two major classes of model in which we can show that only equality restrictions arise, complete models and models with point-valued residuals. We return to consider these cases at the end of this section.

The strengthening of the containment inequality (24) to an equality occurs for test sets $\mathcal{S} \in \mathsf{Q}(h, z)$ that satisfy either one of two criteria, depending on whether the boundary of $\mathcal{S}$, denoted $\partial \mathcal{S}$, has positive measure $G_{U|Z}(\cdot|z)$.

Consider first sets $\mathcal{S} \in \mathsf{Q}(h, z)$ that have a boundary of measure zero $G_{U|Z}(\cdot|z)$. This occurs for example in models in which $G_{U|Z}(\cdot|z)$ is restricted to be continuous with respect to Lebesgue measure, and in which sets $\mathcal{U}(Y, Z; h)$ are convex with probability one. Many models in econometrics

---

[18] There are however no such inequalities in the model studied in Example 2.

satisfy both requirements. Consider sets $\mathcal{S}$ that belong to the collection

$$\mathsf{Q}_0^E\left(h,z\right) \equiv \left\{\mathcal{S} \in \mathsf{Q}\left(h,z\right) : G_{U|Z}\left(\partial\mathcal{S}|z\right) = 0 \text{ and } \forall y \in \mathcal{R}_Y \text{ either } \mathcal{U}\left(y,z;h\right) \subseteq \mathcal{S} \text{ or } \mathcal{U}\left(y,z;h\right) \subseteq \mathrm{cl}(\mathcal{S}^c)\right\}.$$

For any set $\mathcal{S} \in \mathsf{Q}_0^E\left(h,z\right)$, we have that each set $\mathcal{U}\left(y,z;h\right)$ on the conditional support of $\mathcal{U}\left(Y,Z;h\right)$ given $Z = z$ is either contained in $\mathcal{S}$ or contained in $\overline{\mathcal{S}^c}$, and therefore $C_h\left(\mathcal{S}|z\right) + C_h\left(\overline{\mathcal{S}^c}|z\right) = 1$. Furthermore, it follows from $G_{U|Z}\left(\partial\mathcal{S}|z\right) = 0$ that $G_{U|Z}\left(\mathcal{S}|z\right) + G_{U|Z}\left(\overline{\mathcal{S}^c}|z\right) = 1$, and this combined with the containment inequality (24) for both $\mathcal{S}$ and $\overline{\mathcal{S}^c}$ imply that the weak inequality must hold with equality.

A similar conclusion follows for sets $\mathcal{S} \in \mathsf{Q}\left(h,z\right)$ that need not satisfy the vanishing boundary condition that $G_{U|Z}\left(\partial\mathcal{S}|z\right) = 0$, if the other requirement is strengthened. Define

$$\mathsf{Q}_+^E\left(h,z\right) \equiv \left\{\mathcal{S} \in \mathsf{Q}\left(h,z\right) : \forall y \in \mathcal{R}_Y \text{ either } \mathcal{U}\left(y,z;h\right) \subseteq \mathcal{S} \text{ or } \mathcal{U}\left(y,z;h\right) \subseteq \mathcal{S}^c\right\}.$$

Then for any $\mathcal{S} \in \mathsf{Q}_+^E\left(h,z\right)$ we have that $C_h\left(\mathcal{S}|z\right) + C_h\left(\mathcal{S}^c|z\right) = 1$, and trivially because $G_{U|Z}\left(\cdot|z\right)$ is a probability measure, $G_{U|Z}\left(\mathcal{S}|z\right) + G_{U|Z}\left(\mathcal{S}^c|z\right) = 1$.

The following Corollary collects the implications of these results for the strengthening of moment inequalities to equalities.

Corollary 2 *Define $\mathsf{Q}^E\left(h,z\right) \equiv \mathsf{Q}_0^E\left(h,z\right) \cup \mathsf{Q}_+^E\left(h,z\right)$. Then, under the conditions of Theorem 3, the collection of equalities and inequalities*

$$\begin{aligned} C_h\left(\mathcal{S}|z\right) &= G_{U|Z}\left(\mathcal{S}|z\right), \text{ all } \mathcal{S} \in \mathsf{Q}^E\left(h,z\right), \\ C_h\left(\mathcal{S}|z\right) &\leq G_{U|Z}\left(\mathcal{S}|z\right), \text{ all } \mathcal{S} \in \mathsf{Q}^I\left(h,z\right) \equiv \mathsf{Q}\left(h,z\right) \backslash \mathsf{Q}^E\left(h,z\right). \end{aligned}$$

holds if and only if $C_h\left(\mathcal{S}|z\right) \leq G_{U|Z}\left(\mathcal{S}|z\right)$ for all $\mathcal{S} \in \mathsf{Q}\left(h,z\right)$.

Proof: That the statement holds for sets $\mathcal{S} \in \mathsf{Q}_0^E\left(h,z\right)$ was proven as Corollary 2 of CR17. That it also holds for sets $\mathcal{S} \in \mathsf{Q}_+^E\left(h,z\right)$ follows by application of the same steps in that proof with $\mathcal{S}^c$ in place of $\overline{\mathcal{S}^c}$ subsequent to noting that $C_h\left(\mathcal{S}|z\right) + C_h\left(\mathcal{S}^c|z\right) = 1$ and $G_{U|Z}\left(\mathcal{S}|z\right) + G_{U|Z}\left(\mathcal{S}^c|z\right) = 1$. $\square$

There are two particular kinds of models in which *all* elements of $\mathsf{Q}\left(h,z\right)$ belong to $\mathsf{Q}^E\left(h,z\right)$, so that the characterization of the identified set delivered by the Corollary comprises a collection of *only* conditional moment *equalities*. These are complete models, and models – complete or incomplete – in which all U-level sets are singletons. These are cases for which IV methods are already well-developed with, in many cases, identified sets characterized by collections of moment equalities and estimation proceeding using GMM procedures. The analysis presented in this Chapter delivers these results as special cases. The reduction to equalities in these cases is discussed in Sections 5.2.1 and 5.2.2 below.

5.2.1 Complete Models

Complete models admit structures $(h, \mathcal{G}_{U|Z})$ all of which have the completeness property, namely that, for all values z and u of exogenous Z and unobserved U, observable outcomes Y are single-valued functions of z and u. The Y-level sets of structures admitted by complete models,

$$\mathcal{Y}(u, z; h) \equiv \{y : h(y, z, u) = 0\}$$

are singleton sets. Because each value of (u, z) delivers a single value of Y, for each z the U-level sets of the structures admitted by complete models are disjoint and they partition the conditional support of U given $Z = z$.

Theorem 3 and its corollary apply to complete and incomplete models. However for complete models there is simplification because the system of containment inequalities that characterizes the identified set of structures reduces to a system of *equalities* as shown in the next section.

Moment equalities. For every value z and every structure in the identified set obtained with a complete model there is the *equality*

$$G_{U|Z}(\mathcal{S}|z) = C_h(\mathcal{S}|z)$$

for every set $\mathcal{S}$ in the collection of core determining sets. The following argument delivers this conclusion.

First recall that the core determining test sets that generate the inequalities characterizing the identified set of structures are all unions of the U-level sets that comprise the support of the random U-level sets, $\mathcal{U}(Y, Z; h)$. Each such union is defined by a set of values of Y. Consider a value z of Z and a set $\mathcal{T} \subseteq \mathcal{R}_{Y|z}$ where $\mathcal{R}_{Y|z}$ is the conditional support of Y given $Z = z$. Define the union of U-level sets determined by $\mathcal{T}$ and z as follows.

$$\mathcal{U}(\mathcal{T}, z; h) \equiv \bigcup_{y \in \mathcal{T}} \mathcal{U}(y, z; h)$$

Since for complete models U-level sets are disjoint the set of values of Y such that $\mathcal{U}(y, z; h)$ is a subset of a union $\mathcal{U}(\mathcal{T}, z; h)$ is precisely the set $\mathcal{T}$, that is:

$$\{y : \mathcal{U}(y, z; h) \subseteq \mathcal{U}(\mathcal{T}, z; h)\} = \mathcal{T}$$

and it follows that the containment probability for any test set $\mathcal{U}(\mathcal{T}, z; h)$ is simply equal to $\mathbb{P}[Y \in \mathcal{T}|z]$, that is:

$$C_h(\mathcal{U}(\mathcal{T}, z; h)|z) \equiv \mathbb{P}[\mathcal{U}(y, z; h) \subseteq \mathcal{U}(\mathcal{T}, z; h)|z] = \mathbb{P}[Y \in \mathcal{T}|z] = F_{Y|Z}(\mathcal{T}|z) \qquad (27)$$

Suppose that for a set $\mathcal{T}$, a value z and a structural function h of a structure in the identified set the containment inequality holds with a strict *inequality*, thus

$$G_{U|Z}(\mathcal{U}(\mathcal{T}, z; h)|z) > C_h(\mathcal{U}(\mathcal{T}, z; h)|z).$$

Considering $\mathcal{T}^c$ the complement of $\mathcal{T}$, the weak inequality

$$G_{U|Z}(\mathcal{U}(\mathcal{T}^c, z; h)|z) \geq C_h(\mathcal{U}(\mathcal{T}^c, z; h)|z)$$

must hold since h is an element of a structure in the identified set. Therefore we have:

$$
\begin{aligned}
G_{U|Z}(\mathcal{U}(\mathcal{T}, z; h)|z) &> \mathbb{P}[Y \in \mathcal{T}|z], \\
G_{U|Z}(\mathcal{U}(\mathcal{T}^c, z; h)|z) &\geq \mathbb{P}[Y \in \mathcal{T}^c|z],
\end{aligned}
$$

and summing

$$G_{U|Z}(\mathcal{U}(\mathcal{T}, z; h)|z) + G_{U|Z}(\mathcal{U}(\mathcal{T}^c, z; h)|z) > 1.$$

This cannot be true because in a complete model the sets $\mathcal{U}(\mathcal{T}, z; h)$ and $\mathcal{U}(\mathcal{T}^c, z; h)$ *partition* the support of U given $Z = z$. It follows that for all z and every structure in the identified set of a complete model

$$G_{U|Z}(\mathcal{S}|z) = C_h(\mathcal{S}|z)$$

for every set $\mathcal{S}$ in the collection of core determining sets.

5.2.2 Singleton U-level sets

Many commonly used econometric models have the property that, for all the structures they admit, unobservables can be written as single valued functions of observed outcomes and endogenous variables. This is the case in the classical simultaneous equations models of Koopmans, Rubin, and Leipnik (1950), and the single equation IV models of Theil (1953), Newey and Powell (2003), and Chernozhukov and Hansen (2005) for example, but not in models for discrete outcomes, or with high dimensional unobservables, such as discrete choice random utility models and nonlinear measurement error models, or when models place inequality restrictions on observed and unobserved variables.

In models admitting structures whose U-level sets are always singleton sets, an implication of Corollary 2 is that *all* containment inequalities $G_{U|Z}(\mathcal{S}|z) \geq C_h(\mathcal{S}|z)$ of Corollary 1 become equalities. This is shown by the following argument. When U-level sets are singleton sets, for any set $\mathcal{S}$, either $\mathcal{U}(Y, Z; h)$ is a subset of $\mathcal{S}$ or a subset of $\mathcal{S}^c$, the complement of $\mathcal{S}$. It follows that $C_h(\mathcal{S}|z) + C_h(\mathcal{S}^c|z) = 1$ and this together with $G_{U|Z}(\mathcal{S}|z) + G_{U|Z}(\mathcal{S}^c|z) = 1$ and the requirement that $G_{U|Z}(\mathcal{S}|z) \geq C_h(\mathcal{S}|z)$ and $G_{U|Z}(\mathcal{S}^c|z) \geq C_h(\mathcal{S}^c|z)$ implies that for all sets $\mathcal{S}$ and values z,

$G_{U|Z}(\mathcal{S}|z) = C_h(\mathcal{S}|z)$ when the structural function h is such that U is a single valued function of Y and Z.

6 Restrictions on Unobservable Heterogeneity

In the previous section restrictions on the structural function h incorporated into admissible structures $\mathcal{M}$ were shown to enable construction of core determining collections. This is because the specification of h affects the geometry of the level sets $\mathcal{U}(y, z; h)$ on the support of $\mathcal{U}(Y, Z; h)$. This determines for example whether the level sets are connected, and which level sets have non-empty intersections with each other.

In this section it is shown how restrictions on conditional distributions of unobservable heterogeneity can further simplify characterization of the identified set. Like restrictions on h, restrictions on these conditional distributions are incorporated by way of the specification of admissible structures $\mathcal{M}$, which are pairs $(h, \mathcal{G}_{U|Z})$, where recall that $\mathcal{G}_{U|Z}$ represents a collection of conditional distributions $\{G_{U|Z}(\cdot|z) : z \in \mathcal{R}_Z\}$. Thus, both the selectionability characterization in Theorem 2 and the containment inequality characterization (20) in Corollary 1, namely

$$\mathcal{I}\left(\mathcal{M}, \mathcal{F}_{Y|Z}, \mathcal{R}_Z\right) \equiv \left\{ (h, \mathcal{G}_{U|Z}) \in \mathcal{M} : \forall \mathcal{S} \in \mathsf{F}(\mathcal{R}_U), \, C_h(\mathcal{S}|z) \leq G_{U|Z}(\mathcal{S}|z), \, \text{a.e. } z \in \mathcal{R}_Z \right\} \quad (28)$$

already incorporate such restrictions implicitly by only allowing structures that belong to the specified set $\mathcal{M}$. So for example, a model can specify that the conditional mean of U given Z is zero by taking $\mathcal{M}$ to be a collection of structures in which all collections of conditional distributions $\mathcal{G}_{U|Z}$ have conditional distributions $G_{U|Z}(\cdot|z)$ that satisfy this requirement.

Restrictions of stochastic independence, conditional mean independence, and conditional quantile independence are now set out for consideration in identification analysis. These restrictions are commonly made in structural models, but the isolation of the conditional distribution of unobservables in Theorem 2 could also assist in considering other restrictions. For instance, in models studied in Chesher and Rosen (2015) a blend of conditional independence and marginal independence restrictions are analyzed.

Proofs that the identified sets set out below are those delivered under these restrictions are provided in CR17. The restrictions above cover the most common forms of stochastic, mean, and quantile independence restrictions used in applications, but some of the results in CR17 cover a slightly broader set of cases. For example, a conditional mean may be restricted to be constant across values of the conditioning variables, but that constant is only known to lie in some set of values $\mathcal{C}$ rather than to be equal to zero.[19] We refer to that paper for such cases where further

---

[19] In many models such a constant is simply normalized to zero without loss of generality. But in some models such a normalization may not be desirable, or there may be multivariate unobservables and a possibility that restricting the conditional mean of all of them to zero is not merely a normalization.

generality is desired, focusing here instead on expositing the main results as succinctly as possible.

In keeping with the preceding exposition, the restrictions considered here are stated as restrictions on the collection of conditional distributions of unobservable heterogeneity put forth in Restriction A3. The restrictions are formally set out at the beginning of each subsection that follows.

6.1   Stochastic Independence

Restriction SI (Stochastic Independence): Random vectors U and Z are restricted to be stochastically independent, such that for all $\mathcal{G}_{U|Z} \in \mathsf{G}_{U|Z}$, there exists a function $G_U(\cdot)$ such that $G_{U|Z}(\cdot|z) = G_U(\cdot)$ almost surely P_Z. $\square$

Under this restriction, there is only one conditional distribution for U given Z, rather than a collection of conditional distributions across different values of z. Following Theorem 2, the identified set of structures can now be represented as a collection of $(h, G_U(\cdot))$ pairs such that $G_U(\cdot)$ is selectionable with respect to the conditional distribution of $\mathcal{U}(Y, Z; h)$ given $Z = z$ a.e. $z \in \mathcal{R}_Z$. The containment inequality simplifies to

$$C_h(\mathcal{S}|z) \leq G_{U|Z}(\cdot|z) = G_U(\mathcal{S})$$

and since this must hold for almost every $z \in \mathcal{R}_Z$ there is for each set $\mathcal{S}$

$$\sup_{z \in \mathcal{R}_Z} C_h(\mathcal{S}|z) \leq G_U(\mathcal{S}). \tag{29}$$

This is the impact of the independence restriction. Now the containment inequality must hold with the common distribution $G_U(\cdot)$ in place of $G_{U|Z}(\cdot|z)$, and in order for this to be so it must hold for all conditional containment probabilities $C_h(\cdot|z)$, in particular the largest for each $\mathcal{S}$. Thus, the greater the range of possible values of Z, the more pairs $(h, G_U(\cdot))$ will be found to violate (29), and the smaller will be the identified set.

The following Theorem formalizes this development.

Theorem 4 *Let Restrictions A1-A5 and SI hold. Then:*

$$
\begin{aligned}
\mathcal{I}\left(\mathcal{M}, \mathcal{F}_{Y|Z}, \mathcal{R}_Z\right) &= \left\{(h, G_U) \in \mathcal{M} : G_U(\cdot) \preccurlyeq \mathcal{U}(Y, z; h) \ when \ Y \sim F_{Y|Z}(\cdot|z), \ a.e. \ z \in \mathcal{R}_Z\right\} \\
&= \left\{(h, G_U) \in \mathcal{M} : \forall \mathcal{S} \in \mathsf{F}(\mathcal{R}_U), \quad \sup_{z \in \mathcal{R}_Z} C_h(\mathcal{S}|z) \leq G_U(\mathcal{S})\right\} \tag{30} \\
&= \left\{\begin{array}{c} (h, G_U) \in \mathcal{M} : \forall \mathcal{S}_I \in \mathsf{Q}^I(h, z), \ \forall \mathcal{S}_E \in \mathsf{Q}^E(h, z), \\ C_h(\mathcal{S}_I|z) \leq G_U(\mathcal{S}_I), \ C_h(\mathcal{S}_E|z) = G_U(\mathcal{S}_E), \ a.e. \ z \in \mathcal{R}_Z \end{array}\right\}. \tag{31}
\end{aligned}
$$

Proof: See CR17 Theorem 4. $\square$

48

The second representation given in Theorem 4 applies Corollary 1 with $G_{U|Z}(\mathcal{S}|z) = G_U(\mathcal{S})$ for all z. The third representation makes use of results on core determining sets from Theorem 3 and Corollary 2 to characterize the identified set as those structures satisfying a collection of conditional moment equalities and inequalities. An intermediate representation that could also be used would be identical to (30), but instead of requiring the inequality to hold for all $\mathcal{S} \in \mathsf{F}(\mathcal{R}_U)$, requiring it to hold for all $\mathcal{S}$ in the core determining collection $\mathsf{Q}(h,z)$ comprising the union of $\mathsf{Q}^I(h,z)$ and $\mathsf{Q}^E(h,z)$. By Theorem 3 these inequalities holding for all such sets guarantee they will hold for all closed sets on $\mathcal{R}_U$, and furthermore by Corollary 2 the inequalities in $\mathsf{Q}^E(h,z)$ taken together will in fact imply that they must each hold with equality. All of these representations are equivalent, and any could be used as the researcher sees fit.

In many models satisfying Restriction SI the distribution $G_U(\cdot)$ as well as the structural function h may be parametrically specified according to some finite dimensional parameter vector θ, so that $G_U(\cdot) = G_U(\cdot; \theta)$ and $h = h_\theta$. Then the identified set could be represented as the set of parameter vectors θ satisfying the required moment equalities and inequalities, as in

$$\mathcal{I}\left(\mathcal{M}, \mathcal{F}_{Y|Z}, \mathcal{R}_Z\right) = \left\{ \theta \in \Theta : \forall \mathcal{S} \in \mathsf{Q}(\theta, z), \quad \sup_{z \in \mathcal{R}_Z} C_\theta(\mathcal{S}|z) \leq G_U(\mathcal{S}; \theta) \right\},$$

where $\mathsf{Q}(\theta, z)$ and $C_\theta(\mathcal{S}|z)$ are shorthand for $\mathsf{Q}(h_\theta, z)$ and $C_{h_\theta}(\mathcal{S}|z)$, respectively. One may then embed these inequalities in an objective function that is minimized precisely at those θ in $\mathcal{I}\left(\mathcal{M}, \mathcal{F}_{Y|Z}, \mathcal{R}_Z\right)$, such as

$$D(\theta) = \int_{z \in \mathcal{R}_Z} \sum_{\mathcal{S} \in \mathsf{Q}(\theta, z)} \max\left\{0, C_\theta(\mathcal{S}|z) - G_U(\mathcal{S}; \theta)\right\} w(\theta, z) v(\mathcal{S}) \, dF_Z(z) \tag{32}$$

where $w(\theta, z)$ and $v(\mathcal{S})$ are strictly positive weights.[20] This function aggregates violations of the inequality $C_\theta(\mathcal{S}|z) \leq G_U(\mathcal{S}; \theta)$, taking the value 0 if $\theta \in \mathcal{M}^*$, and otherwise taking positive values. Several estimation and inference methods are based on minimizing a sample analog $\hat{D}(\cdot)$ of such a function. With a parametric specification this can be carried out by minimizing over a finite dimensional, typically compact, parameter space Θ.

In some settings the researcher may wish to impose independence without restricting the distribution of unobserved heterogeneity to a parametric family, even if a parametric specification is maintained for h, with $h = h_\theta$ for some $\theta \in \Theta$, as before. With $G_U(\cdot)$ not parametrically restricted, characterizing the identified set using the containment inequality remains conceptually straightforward. Suppose that independence is maintained, but $G_U(\cdot)$ is completely unrestricted, let $\mathcal{G}_U$

---

[20] If $\mathsf{Q}(\theta, z)$ is not a finite set one could instead consider using a finite collection of sets $\mathcal{S}$ for inference, or integrating over a continuum of such sets with some density function in place of $v(\mathcal{S})$.

denote the set of all monotone set functions that map from $\mathcal{R}_U$ to $[0,1]$. Then

$$\mathcal{I}\left(\mathcal{M}, \mathcal{F}_{Y|Z}, \mathcal{R}_Z\right) = \left\{(\theta, G_U) \in \Theta \times \mathcal{G}_U : \forall \mathcal{S} \in \mathsf{Q}\left(\theta, z\right), \quad \sup_{z \in \mathcal{R}_Z} C_\theta\left(\mathcal{S}|z\right) \leq G_U\left(\mathcal{S}\right)\right\},$$

which is equivalently the set of minimizers of $D\left(\theta, G_U\right)$ on $\Theta \times \mathcal{G}_U$ with $D\left(\cdot, \cdot\right)$ defined identically as (32), but with the second argument G_U replacing $G_U\left(\mathcal{S}; \theta\right)$. In principle one could consider replacing the infinite dimensional object $G_U\left(\cdot\right)$ with an appropriately defined sieve-space for estimation and inference using a sample version of $D\left(\theta, G_U\right)$. This however seems likely to present both theoretical and computational challenges. At present such an approach appears to be unavailable.

An alternative approach considered in the literature on semiparametric modeling is to focus attention on the parameters of the structural function. The containment inequality characterization can in fact be used to produce inequalities on θ from which the unknown and nonparametrically specified distribution of unobservables is absent. To see how this is done, consider the containment inequality for an arbitrary set $\mathcal{S}$, as well as its complement $\mathcal{S}^c$.

$$C_h\left(\mathcal{S}|z\right) \leq G_U\left(\mathcal{S}\right), \qquad C_h\left(\mathcal{S}^c|z\right) \leq G_U\left(\mathcal{S}^c\right). \tag{33}$$

Since $G_U\left(\cdot\right)$ gives the probability that U lies within the given argument, it must be that $G_U\left(\mathcal{S}\right) + G_U\left(\mathcal{S}^c\right) = 1$. It is also easy to see that $C_h\left(\mathcal{S}|z\right) + C_h\left(\mathcal{S}^c|z\right)$ must be less than or equal to one, because $C_h\left(\mathcal{S}|z\right)$ is the conditional probability that a given set, namely $\mathcal{U}\left(Y, Z; h\right)$, is contained in $\mathcal{S}$, while $C_h\left(\mathcal{S}^c|z\right)$ is the conditional probability that the same set is contained in $\mathcal{S}^c$. These events are mutually exclusive, so their probabilities must sum to no more than one.

The statistical independence restriction requires however that (33) holds for almost every $z \in \mathcal{R}_Z$, *for the same* $G_U\left(\cdot\right)$, since U and Z are independent. Therefore, under Restriction SI, there is for any set $\mathcal{S}$,

$$\forall z, z' \in \mathcal{R}_Z, \qquad C_h\left(\mathcal{S}|z\right) \leq G_U\left(\mathcal{S}\right) \text{ and } C_h\left(\mathcal{S}^c|z'\right) \leq G_U\left(\mathcal{S}^c\right).$$

Consequently, it follows that,

$$\forall \mathcal{S} \subseteq \mathcal{R}_U, \quad \forall z, z' \in \mathcal{R}_Z, \qquad C_h\left(\mathcal{S}|z\right) + C_h\left(\mathcal{S}^c|z'\right) \leq 1, \tag{34}$$

as otherwise the preceding inequalities would imply a violation of the requirement that $G_U\left(\mathcal{S}\right) + G_U\left(\mathcal{S}^c\right) = 1$. The role of the instrumental variable Z is apparent because for any fixed z, $C_h\left(\mathcal{S}|z\right) + C_h\left(\mathcal{S}^c|z\right)$ is necessarily less than or equal to one. Variation in z that affects the containment probabilities $C_h\left(\mathcal{S}|z\right)$ and $C_h\left(\mathcal{S}^c|z\right)$ can however lead to violations and thereby provide observable implications that can be used to rule out structural functions h that produce a violation. The greater the variation in $C_h\left(\mathcal{S}|z\right)$ and $C_h\left(\mathcal{S}^c|z\right)$ as z varies on its support, the smaller will be the

set of h that satisfy (34).

The above reasoning justifies the collection of inequalities described in (34), which can be used to characterize bounds on the structural function h. An alternative approach to the same end is to note that (33) taken together with the equality $G_U(\mathcal{S}^c) = 1 - G_U(\mathcal{S})$ imply that

$$\forall z, z' \in \mathcal{R}_Z, \qquad C_h(\mathcal{S}|z) \leq G_U(\mathcal{S}) \leq 1 - C_h(\mathcal{S}^c|z'),$$

which of course implies that for all z and z', $C_h(\mathcal{S}|z) \leq 1 - C_h(\mathcal{S}^c|z')$. Adding $C_h(\mathcal{S}^c|z')$ to both sides of this inequality, which must hold for all sets $\mathcal{S}$ results in (34).

Formally, this delivers the following outer region for structural function h, stated as a Corollary to the previous Theorem.

Corollary 3 *If Restrictions A1-A5 and SI hold, then the set*

$$\overline{\mathcal{H}^*} = \left\{ h \in \mathcal{H} : \sup_{\mathcal{S} \in \mathsf{S}} \sup_{z,z' \in \mathcal{R}_Z} \left(C_h(\mathcal{S}|z) + C_h(\mathcal{S}^c|z') \right) \leq 1 \right\} \tag{35}$$

comprises bounds on h for any collection of test sets S. If, in addition, $G_U(\partial\mathcal{S}) = 0$ for all $\mathcal{S} \in \mathsf{S}$, where $\partial\mathcal{S}$ denotes the boundary $\mathcal{S}$, then there is the refinement

$$\overline{\mathcal{H}^*} = \left\{ h \in \mathcal{H} : \sup_{\mathcal{S} \in \mathsf{S}} \sup_{z,z' \in \mathcal{R}_Z} \left(C_h(\mathrm{cl}(\mathcal{S})|z) + C_h(\mathrm{cl}(\mathcal{S}^c)|z') \right) \leq 1 \right\}, \tag{36}$$

where $\mathrm{cl}(\mathcal{S})$ and $\mathrm{cl}(\mathcal{S}^c)$ denote the closure of $\mathcal{S}$ and $\mathcal{S}^c$, respectively.

Proof: See CR17 Corollary 3. $\square$

Because the inequality $C_h(\mathcal{S}|z) + C_h(\mathcal{S}^c|z') \leq 1$ must hold for almost every z and z' and all $\mathcal{S}$ it must equivalently hold for the supremum over these arguments, as used in the definition of $\overline{\mathcal{H}^*}$. The refinement (36) holds in many practical cases, for example whenever sets $\mathcal{S}$ are convex and $G_U(\cdot)$ is restricted to be absolutely continuous with respect to Lebesgue measure, as is the case in many commonly used econometric models.

The set $\overline{\mathcal{H}^*}$ provides valid bounds on the structural function h, equivalently an outer region. The set may not however comprise sharp bounds, even if the collection of sets S were all subsets of $\mathcal{R}_U$. To understand why, note that $h \in \overline{\mathcal{H}^*}$ is equivalent to inequalities

$$\sup_{z \in \mathcal{R}_Z} C_h(\mathcal{S}|z) \leq \inf_{z' \in \mathcal{R}_Z} \left(1 - C_h(\mathcal{S}^c|z') \right), \tag{37}$$

but even if this inequality holds for all $\mathcal{S} \subseteq \mathcal{R}_U$, this by itself does not ensure that there exists a

probability measure $G_U(\cdot)$ such that

$$\sup_{z \in \mathcal{R}_Z} C_h(\mathcal{S}|z) \leq G_U(\mathcal{S}) \leq \inf_{z' \in \mathcal{R}_Z} \left(1 - C_h(\mathcal{S}^c|z')\right) \tag{38}$$

holds for all $\mathcal{S} \subseteq \mathcal{R}_U$, because (37) doesn't guarantee the existence of an *additive* set function $G_U(\cdot)$ satisfying (38). The existence of conditions under which $\overline{\mathcal{H}^*}$ is in fact sharp remains an open question.

6.2 Mean Independence

Restriction MI (Mean Independence): The collection $\mathsf{G}_{U|Z}$ comprises all collections of conditional distribution functions $\mathcal{G}_{U|Z}$ whose elements $G_{U|Z}(\cdot|z)$ satisfy $E[U|Z=z] = 0$ almost surely P_Z. $\square$

This restriction limits the distribution of unobservable heterogeneity to those satisfying a familiar mean zero restriction conditional on the realization of the exogenous variables Z. It is straightforward to generalize this restriction to allow the conditional mean $E[U|Z=z]$ to be equal to some constant c in a known set $\mathcal{C}$, or to restrict some function of U to have fixed conditional mean for all $z \in \mathcal{R}_Z$. For simplicity of exposition we focus here on the case where the conditional mean of U is simply restricted to zero. More general cases are covered in CR17.

In order to characterize the identified set under this conditional mean restriction, it is helpful to use a particular notion of the expectation of a random set, namely the Aumann expectation. The definition is the following.

Definition 6 *The **Aumann expectation** of random set $\mathcal{A}$ on a finite-dimensional Euclidean space is given by*

$$\mathbb{E}[\mathcal{A}] \equiv \{E[A] : A \in \mathrm{Sel}(\mathcal{A}) \ \text{and} \ E[A] < \infty\}$$

The Aumann expectation of random set $\mathcal{A}$ conditional on $B = b$ is

$$\mathbb{E}[\mathcal{A}|b] \equiv \{E[A|b] : A \in \mathrm{Sel}(\mathcal{A}) \ \text{and} \ E[A|b] < \infty\}.$$

The definition here corresponds to that of Molchanov (2005, p. 151) specialized to a finite dimensional space.[21] Under Restriction MI the selectionability criterion of Theorem 2 can be used to characterize the identified set of structures and structural functions as follows.

Theorem 5 *Let Restrictions A1-A5 and MI hold and suppose that $(\Omega, \mathsf{L}, \mathbb{P})$ is non-atomic. Then the identified set for structural function h is*

$$\mathcal{H}^* = \{h \in \mathcal{H} : 0 \in \mathbb{E}[\mathcal{U}(Y,Z;h)|z], \ a.e. \ z \in \mathcal{R}_Z\},$$

---

[21]Specifically, the standard definition of the Aumann expectation is the closure of the set provided in Definition 6. When the random set under consideration is defined on a finite dimensional space – as it is here – the closure operator is unnecessary, see e.g. Nguyen (2006) p.184.

and the identified set for $\left(h, \mathcal{G}_{U|Z}\right)$ is:

$$\mathcal{I}\left(\mathcal{M}, \mathcal{F}_{Y|Z}, \mathcal{R}_Z\right)^* = \left\{\left(h, \mathcal{G}_{U|Z}\right) \in \mathcal{M} : h \in \mathcal{H}^* \text{ and } G_{U|Z}\left(\cdot|z\right) \lesssim \mathcal{U}\left(Y, Z; h\right) \text{ conditional on } Z = z, \text{ a.e. } z \in \mathcal{R}_Z\right\},$$

where because of Restriction MI, all structures $\left(h, \mathcal{G}_{U|Z}\right) \in \mathcal{I}\left(\mathcal{M}, \mathcal{F}_{Y|Z}, \mathcal{R}_Z\right)$ have $G_{U|Z}\left(\cdot|z\right)$ such that $E\left[U|z\right] = 0$ a.e. $z \in \mathcal{R}_Z$.

Proof: Follows Theorem 5 in CR17 by setting $\mathcal{C} = \{0\}$. $\square$

Theorem 5 characterizes $\mathcal{H}^*$ as those structural functions h such that $0 \in \mathbb{E}\left[\mathcal{U}\left(Y, Z; h\right)|z\right]$ a.e. $z \in \mathcal{R}_Z$. Nonempty $\mathcal{H}^*$ guarantees for each $h \in \mathcal{H}^*$ the *existence* of collections of conditional distributions $\mathcal{G}_{U|Z}$ with elements $G_{U|Z}\left(\cdot|z\right)$ each satisfying the conditional mean restriction MI. The identified set for $\left(h, \mathcal{G}_{U|Z}\right)$ is then simply those pairs of $\left(h, \mathcal{G}_{U|Z}\right)$ such that $0 \in \mathbb{E}\left[\mathcal{U}\left(Y, Z; h\right)|z\right]$, and $G_{U|Z}\left(\cdot|z\right)$ is selectionable with respect to $\mathcal{U}\left(Y, Z; h\right)$ conditional on $Z = z$, a.e. $z \in \mathcal{R}_Z$.

Furthermore, under Restriction MI the random set $\mathcal{U}\left(Y, Z; h\right)$ is integrable, since this restriction guarantees that it has a measurable selection with a finite $\mathbf{L}^1$ norm.[22] Consequently, making use of Molchanov (2005, Theorem 2.1.47-iv, p. 171), the support function can be used to present an equivalent characterization of the condition that $0 \in \mathbb{E}\left[\mathcal{U}\left(Y, Z; h\right)|z\right]$, namely that $0 \in \mathbb{E}\left[\mathcal{U}\left(Y, Z; h\right)|z\right]$ if and only if

$$\inf_{v \in \mathcal{R}_Z : \|v\| = 1} \left\{E\left[m\left(v, \mathcal{U}\left(Y, Z; h\right)\right)|z\right]\right\} \geq 0, \tag{39}$$

where for any set $\mathcal{S}$,

$$m\left(v, \mathcal{S}\right) \equiv \sup\left\{v \cdot s : s \in \mathcal{S}\right\}$$

denotes the support function of $\mathcal{S}$ evaluated at v. This means that under the conditions of Theorem 5 there is also the representation

$$\mathcal{H}^* = \left\{h \in \mathcal{H} : \inf_{v \in \mathcal{R}_Z : \|v\| = 1} \left\{E\left[m\left(v, \mathcal{U}\left(Y, Z; h\right)\right)|z\right]\right\} \geq 0, \text{ a.e. } z \in \mathcal{R}_Z\right\}$$

BMM11 previously employed Molchanov (2005, Theorem 2.1.47-iv, p. 171) in consideration of the conditional Aumann expectation of random *outcome* set $\mathcal{Y}\left(Z, U; h\right)$ in characterizing its selections for identification analysis. The representation here is convenient for nonlinear models in which Restriction MI is imposed with U entering h non-additively. If structural function h is additively separable in U, then the representation of the identified set in Theorem 5 which uses the Aumann expectation of the random set $\mathcal{U}\left(Y, Z; h\right)$ differs from that of BMM11 employing random set $\mathcal{Y}\left(Z, U; h\right)$ only by a trivial location shift.

Alternative types of IV restrictions on conditional means given values of instruments may also be incorporated. An example is the monotone instrumental variable (MIV) restriction introduced by

---

[22] See Molchanov (2005, Definition 1.1, p. 146, and Definition 1.11(ii) pp. 150-151) for the formal definition of an integrable random set.

Manski and Pepper (2000) that requires certain conditional expectations of *potential* outcomes to be monotone in the value of an instrumental variable conditioned upon. Specifically, this restriction requires conditional expectations of individuals' *treatment response functions* to be monotone in the instrumental variable. With some effort, the restrictions can be equivalently expressed in a structural model that incorporates finite dimensional vectors of unobservables as done here. As such the MIV model of Manski and Pepper (2000) may be seen as early inspiration for further expansion of the use of IVs as enabled by GIV analysis.

6.3 Quantile Independence

Restriction QI (Quantile Independence): The support of U is a subset of $\mathbb{R}$ and U-level sets $\mathcal{U}(y, z; h)$ comprise intervals

$$\mathcal{U}(y, z; h) = [\underline{u}(y, z; h), \overline{u}(y, z; h)],$$

with endpoints possibly $\pm\infty$, and for some known $\tau \in (0, 1)$, $\mathsf{G}_{U|Z}$ comprises all collections of conditional distribution functions $\mathcal{G}_{U|Z}$ whose elements $G_{U|Z}(\cdot|z)$ have τ-quantile equal to 0 almost surely $\mathbb{P}_Z$. $\square$

This quantile independence restriction set out above requires that U is univariate, which ensures that the τ conditional quantile restriction $Q_{U|Z}(\tau|z) = 0$ is well-defined. The restriction additionally imposes that the sets $\mathcal{U}(y, z; h)$ are intervals, although once again, alternative restrictions could also be considered.

The following theorem is a restatement of Theorem 6 of CR17, specialized to the case where the conditional quantile of U is restricted to be zero.

Theorem 6 *Let Restrictions A1-A5 and QI hold. Then (i) the identified set for structural function h is*

$$\mathcal{H}^* = \left\{ h \in \mathcal{H} : \sup_{z \in \mathcal{R}_Z} \mathbb{P}\left[\overline{u}(Y, Z; h) < 0 | z\right] \leq \tau \leq \inf_{z \in \mathcal{R}_Z} \mathbb{P}\left[\underline{u}(Y, Z; h) \leq 0 | z\right] \right\}. \tag{40}$$

(ii) If $\underline{u}(Y, Z; h)$ and $\overline{u}(Y, Z; h)$ are continuously distributed conditional on $Z = z$, a.e. $z \in \mathcal{R}_Z$, then equivalently

$$\mathcal{H}^* = \left\{ h \in \mathcal{H} : \sup_{z \in \mathcal{R}_Z} \underline{q}(\tau, z; h) \leq 0 \leq \inf_{z \in \mathcal{R}_Z} \overline{q}(\tau, z; h) \right\}, \tag{41}$$

where $\underline{q}(\tau, z; h)$ and $\overline{q}(\tau, z; h)$ are the τ-quantiles of respectively $\underline{u}(Y, Z; h)$ and $\overline{u}(Y, Z; h)$ conditional on $Z = z$, (iii) The identified set for $\left(h, \mathcal{G}_{U|Z}\right)$ is:

$$\mathcal{I}\left(\mathcal{M}, \mathcal{F}_{Y|Z}, \mathcal{R}_Z\right) = \left\{ \left(h, \mathcal{G}_{U|Z}\right) \in \mathcal{M} : h \in \mathcal{H}^* \text{ and } G_{U|Z}(\cdot|z) \preccurlyeq \mathcal{U}(Y, Z; h) \text{ conditional on } Z = z, \text{ a.e. } z \in \mathcal{R}_Z \right\},$$

where following from Restriction QI, all structures $\left(h, \mathcal{G}_{U|Z}\right) \in \mathcal{I}\left(\mathcal{M}, \mathcal{F}_{Y|Z}, \mathcal{R}_Z\right) \subseteq \mathcal{M}$ are such that $q_{U|Z}\left(\tau|z\right) = 0$, a.e. $z \in \mathcal{R}_Z$.

Proof: This follows from CR17 Theorem 6 with $\mathcal{C} \equiv \{0\}$. $\square$

The inequalities comprising (40) follow from knowledge that $\underline{u}\left(Y, Z; h\right) \leq U \leq \overline{u}\left(Y, Z; h\right)$. These inequalities also arise on applying the containment inequality $C_h\left(\mathcal{S}|z\right) \leq G_{U|Z}\left(\mathcal{S}|z\right)$ to test sets $\mathcal{S} = (-\infty, c]$ and $\mathcal{S} = [c, \infty)$. These implications are straightforward to establish, and the proof of this Theorem further shows that these inequalities in fact exhaust all observable implications. That is, for any h, if the containment inequalities hold for these two test sets, then there exists an admissible collection of conditional distributions $\mathcal{G}_{U|Z}$ such that the containment inequality holds for all closed test sets in $\mathcal{R}_U$. The characterization (40) of structural functions h is thus sharp. The second part of Theorem 6 provides an equivalent characterization in the event that $\underline{u}\left(Y, Z; h\right)$ and $\overline{u}\left(Y, Z; h\right)$ are continuously distributed. The third part of Theorem 6 states that the identified set of structures $\left(h, \mathcal{G}_{U|Z}\right)$ – rather than only structural functions h – are all elements of $\mathcal{H}^*$ paired with distributions $G_{U|Z}\left(\cdot|z\right)$ that are selectionable with respect to the conditional distribution of $\mathcal{U}\left(Y, Z; h\right)$ given $Z = z$, a.e. $z \in \mathcal{R}_Z$, equivalently all elements of $\mathcal{H}^*$ paired with conditional distributions of unobservable variables that could feasibly generate the conditional distributions of observable variables $\mathcal{F}_{Y|Z}$.

Section 7.1.4 considers a conditional median restriction in the context of the interval censored endogenous variable model of Example 5 of Section 3.2 in order to demonstrate application of the results obtained under Restriction QI. Additional research on models of censoring incorporating conditional quantile restrictions includes Hong and Tamer (2003), Khan and Tamer (2009), and Khan, Ponomareva, and Tamer (2011).

7 Application I: Interval Censored Endogenous Variable

In this Section we return to Example 5 of Section 3.2. The model specifies a single equation for an outcome involving an endogenous explanatory variable which is only known to lie in an interval whose lower and upper bounds are observed. The bounds of the interval may be fixed or random variables, and are possibly endogenous. This situation can arise with survey data, for example when endogenous income is top coded or reported in intervals whose bounds may be chosen by a respondent.

This is a substantial generalization of a single equation model with an interval censored *exogenous* variable studied in Manski and Tamer (2002). As in that paper, no restrictions are imposed on the censoring process or on the realization of the censored variable relative to the observed interval. Here the interval censored variable and the endpoints of censoring intervals can be endogenous while in Manski and Tamer (2002) these variables were required to be exogenous.

The identifying power of various restrictions on the dependence of unobservable U and observed instrumental variables Z is examined. Numerical illustrations are provided for particular data generating structures.[23]

7.1 Restrictions and Identified Sets

7.1.1 Structural Function

In the model considered here a structural equation

$$Y_1 = g\left(Y_2^*, Z_1, U\right), \tag{42}$$

relates continuously distributed scalar outcome Y_1, scalar endogenous Y_2^*, exogenous, K-element $Z = (Z_1, Z_2)$, and an unobservable scalar variable U which has a strictly monotone distribution function $\Lambda\left(\cdot\right)$. The support of (Y_1, Y_2^*, Z, U) is a subset of an appropriately dimensioned Euclidean space. The function $g\left(\cdot, \cdot, \cdot\right)$ is restricted to be weakly increasing in its first argument, and strictly monotone in its third argument, normalized increasing. The support of Z is denoted $\mathcal{R}_Z$ which is a subset of $\mathbb{R}^K$. In an example presented later there is a simple linear specification.

It is important to understand that $\Lambda(\cdot)$ is the distribution function of the *marginal* distribution of U. At this point, no restrictions have been imposed on the joint distribution of (U, Z), so that for any $z \in \mathcal{R}_U$, the conditional distribution function of U given $Z = z$ can depend on z. It is straightforward to allow $g\left(y_2^*, z_1, u\right)$ to be either (weakly) increasing or decreasing in y_2^* for all (z_1, u) and to leave the direction of the dependence unspecified, but here, to simplify the exposition, the case in which it is maintained that $g\left(y_2^*, z_1, u\right)$ is weakly increasing in y_2^* is considered.

The model allows the possibility that U and Y_2^* are dependent random variables. The impact of various restrictions on the dependence of U and Z on the identifying power of the model is of central interest.

The endogenous variable Y_2^* may not be observed, but there are observed variables Y_{2l} and Y_{2u}, with $Y_{2l} \leq Y_{2u}$ almost surely, such that

$$Y_2^* = Y_{2l} + W \times \left(Y_{2u} - Y_{2l}\right), \tag{43}$$

for some unobserved random variable $W \in [0, 1]$. There is no restriction on the distribution of W and no restriction on its stochastic relation to observed or unobserved variables.

Since there is no restriction on the censoring process, the unobserved variable W can be sup-

---

[23]Models allowing censored *outcome* variables with *uncensored* endogenous explanatory variables with sufficient conditions for point identification include those in Hong and Tamer (2003) and Khan and Tamer (2009).

pressed by replacing (43) with the equivalent formulation

$$\mathbb{P}\left[Y_{2l} \leq Y_2^* \leq Y_{2u}\right] = 1. \tag{44}$$

We consider cases in which realizations of (Y_1, Y_{2l}, Y_{2u}, Z) are obtained *via* some process which identifies the conditional distribution of (Y_1, Y_{2l}, Y_{2u}) given $Z = z$ for all values z in the support, $\mathcal{R}_Z$, of exogenous Z.

The censoring variables (Y_{2l}, Y_{2u}) could be nonstochastic given Y_2^*, for example defining fixed intervals in which realizations of continuous Y_2^* are binned. But they could be stochastic and they may be endogenous. Interval censoring with stochastic interval endpoints could arise if respondents choose intervals to report or if reported intervals arise after interviewer probing based on an initial interval choice by a respondent or through some selection mechanism. Since realizations of Y_{2l} and Y_{2u} can be equal, the model allows the possibility that some realizations reveal the value of Y_2^*. This situation could arise when there is top-coding of, for example, income data, or when Y_2^* is a duration subject to censoring above or below.

A suitable structural function for this model is

$$h\left(y, z, u\right) = \left|y_1 - g\left(y_{2l}, z_1, u\right)\right|_- + \left|g\left(y_{2u}, z_1, u\right) - y_1\right|_+, \tag{45}$$

and this, with the condition $\mathbb{P}\left[h\left(Y, Z, U\right) = 0\right] = 1$, is equivalent to equations (42) and (44).[24]

The level sets of this structural function in respectively Y-space and U-space, are

$$\mathcal{Y}\left(u, z; h\right) = \left\{y = \left(y_1, y_{2l}, y_{2u}\right) \in \mathcal{R}_{Y|z} : g\left(y_{2l}, z_1, u\right) \leq y_1 \leq g\left(y_{2u}, z_1, u\right)\right\},$$

and

$$\mathcal{U}\left(y, z; h\right) = \left[g^{-1}\left(y_{2u}, z_1, y_1\right), g^{-1}\left(y_{2l}, z_1, y_1\right)\right]. \tag{46}$$

Here g^{-1} denotes the inverse of g in its last argument such that for all a, b, c

$$g(a, b, g^{-1}\left(a, b, c\right)) = c.$$

In some of the development to come, and in the numerical illustrations, the structural function h is restricted such that explanatory variables and the unobservable variable are required to appear in a linear index thus:

$$g\left(y_2^*, z, u\right) = \beta y_2^* + z_1 \gamma + u, \tag{47}$$

where the first element of z_1 is one, and g (and hence h) are now parameterized by $(\beta, \gamma) \in$

---

[24]The notation $\left|\cdot\right|_+$ and $\left|\cdot\right|_-$ indicate respectively the positive and negative part of their arguments, that is: $\left|a\right|_+ \equiv \max(a, 0)$, $\left|a\right|_- \equiv -\min(a, 0)$.

$\mathbb{R}^{\dim(z_1)+1}$. Under the linear index restriction

$$g^{-1}(y_2^*, z, y_1) = y_1 - \beta y_2^* - z_1 \gamma$$

and the U level set becomes

$$\mathcal{U}(y, z; h) = [y_1 - \beta y_{2u} - z_1 \gamma, y_1 - \beta y_{2l} - z_1 \gamma]$$

in which β is restricted to be nonnegative as we are restricting the function $g(y_2^*, z, u)$ to be nondecreasing in y_2^*.

Using results set out in Section 6, the identified set for this model is now characterized under alternative restrictions on the collection of conditional distributions $\mathcal{G}_{U|Z}$. For each restriction considered, the characterization is expressed in terms of conditional moment inequalities that can be used as a basis for estimation and inference.

7.1.2 Stochastic Independence

First consider the restriction requiring U and Z to be independently distributed, $U \perp\!\!\!\perp Z$. Let $G_U(\cdot)$ denote the marginal distribution of U which, recall, has cumulative distribution function $\Lambda(\cdot)$. The function $G_U(\cdot)$ has a set-valued argument with $G_U(\mathcal{S}) = \mathbb{P}[U \subseteq \mathcal{S}]$ so: $\Lambda(s) = G_U((-\infty, s])$.

Each set $\mathcal{U}(y, z; h)$ is a closed interval on $\mathbb{R}$ and hence connected. Theorem 3 can be used to express the identified set of structures (h, G_U) as comprising those structures such that the inequality

$$\mathbb{P}[\mathcal{U}(Y, Z; h) \subseteq \mathcal{S}|Z = z] \leq G_U(\mathcal{S}) \tag{48}$$

is satisfied for almost every $z \in \mathcal{R}_Z$ and for all $\mathcal{S} \in \mathsf{Q}(h, z)$, where $\mathsf{Q}(h, z)$ is the collection of the intervals that can be formed as unions of sets of the form $\left[g^{-1}(y_{2u}, z_1, y_1), g^{-1}(y_{2l}, z_1, y_1)\right]$. If the components of y are continuously distributed with sufficiently rich support the required test sets may constitute *all* intervals on $\mathbb{R}$.[25] Unless g has very restricted structure, the conditions for (48) to hold with equality will not be satisfied for any test set $\mathcal{S}$ in which case $\mathsf{Q}^E(h, z) = \emptyset$ and $\mathsf{Q}^I(h, z) = \mathsf{Q}(h, z)$ which is the collection of all intervals on $\mathbb{R}$, here denoted

$$\mathsf{Q} \equiv \left\{[a, b] \in \mathbb{R}^2 : a \leq b\right\}.$$

Let m denote a structure characterized by h, a structural function, as in (45), and Λ, a distribution function of unobserved U and let $\mathcal{M}$ denote the collection of structures admitted by the

---

[25] If the support of Y_1 is limited, application of Theorem 3 may dictate that not all intervals of $\mathbb{R}$ need to be considered as test sets. Nonetheless, this smaller collection of core-determining sets will differ for different (h, z). A characterization based on all intervals, although employing more test sets than necessary, has the advantage of being invariant to (h, z). Both characterizations - that using the core determining sets of Theorem 3, and that using all intervals on $\mathbb{R}$ - are for the same identified set. That is, both characterizations are sharp.

model. Applying Theorem 4 gives the following expression for the identified set of structures.

$$\mathcal{I}\left(\mathcal{M}, \mathcal{F}_{Y|Z}, \mathcal{R}_Z\right) = \{m \in \mathcal{M} : \forall \left[u_*, u^*\right] \in \mathsf{Q}, \mathbb{P}\left[\mathcal{U}\left(Y, Z; h\right) \subseteq \left[u_*, u^*\right]|z\right] \leq \Lambda\left(u^*\right) - \Lambda\left(u_*\right), \text{ a.e. } z \in \mathcal{R}_Z\}$$

For any structural function h, the probability $\mathbb{P}\left[\mathcal{U}\left(Y, Z; h\right) \subseteq \left[u_*, u^*\right]|z\right]$ is the probability of an event determined entirely by realizations of observed variables, and is thus identified. Specifically, the containment functional inequality appearing in the definition of the identified set, $\mathcal{I}\left(\mathcal{M}, \mathcal{F}_{Y|Z}, \mathcal{R}_Z\right)$, can be equivalently written as

$$\mathbb{P}\left[u_* \leq g^{-1}\left(Y_{2u}, Z_1, Y_1\right) \wedge g^{-1}\left(Y_{2l}, Z_1, Y_1\right) \leq u^*|z\right] \leq \Lambda\left(u^*\right) - \Lambda\left(u_*\right),$$

or, using monotonicity of $g\left(y_2, z_1, u\right)$ in its third argument,

$$\mathbb{P}\left[g\left(Y_{2u}, Z_1, u_*\right) \leq Y_1 \leq g\left(Y_{2l}, Z_1, u^*\right)|z\right] \leq \Lambda\left(u^*\right) - \Lambda\left(u_*\right). \tag{49}$$

With the linear index restriction from (47) this produces the following representation for the identified set.

$$\mathcal{I}\left(\mathcal{M}, \mathcal{F}_{Y|Z}, \mathcal{R}_Z\right) = \left\{ \begin{array}{l} \left(\beta, \gamma, \Lambda\left(\cdot\right)\right) \in \mathcal{M} : \forall \left[u_*, u^*\right] \in \mathsf{Q}, \\ \quad \mathbb{P}\left[u_* + \beta Y_{2u} \leq Y_1 - Z_1\gamma \leq u^* + \beta Y_{2l}|z\right] \leq \Lambda\left(u^*\right) - \Lambda\left(u_*\right), \text{ a.e. } z \in \mathcal{R}_Z \end{array} \right\} \tag{50}$$

If the function $g(\cdot, \cdot, \cdot)$ were restricted to be *decreasing* in its first argument the inequalities defining the event on the left hand side of (49) would be reversed. It is clearly possible to allow for restrictions requiring weak monotonicity. If the function $g(\cdot, \cdot, \cdot)$ were restricted to be monotone with no restriction on the sign of the effect of y_2^* on the value taken by the function then the identified set would comprise the union of the set defined in (50) and the set obtained by reversing the inequalities inside the conditional probability appearing in (50). If one of these components were empty then the sign of β would be identified.

Now consider two particular cases: in one of these the model imposes a parametric restriction on the distribution function, Λ, of unobserved U; in the other the model leaves Λ completely unspecified.

Parametric restriction: Gaussian Unobservable   Here we consider the case in which, in addition to the linear index restriction (47), U is restricted to be normally distributed with mean zero and variance $\sigma^2 > 0$ so that $\Lambda\left(u\right) = \Phi\left(\sigma^{-1}u\right)$, where $\Phi\left(\cdot\right)$ is the standard normal cumulative distribution function. In this case the model is fully characterized by $\theta \equiv \left(\beta, \gamma, \sigma\right)$, and $\mathcal{M}$ can be represented as the parameter space Θ containing admissible values of θ.

The identified set is as in (50) with $\Lambda\left(u^*\right) - \Lambda\left(u_*\right)$ replaced by $\Phi\left(\sigma^{-1}u^*\right) - \Phi\left(\sigma^{-1}u_*\right)$. The change of variables $t^* = \Phi\left(\sigma^{-1}u^*\right)$ and $t_* \equiv \Phi\left(\sigma^{-1}u_*\right)$ delivers the following equivalent represen-

tations.

$$\mathcal{I}\left(\mathcal{M}, \mathcal{F}_{Y|Z}, \mathcal{R}_Z\right) = \left\{ \begin{array}{l} \theta \in \Theta : \forall \left[t_*, t^*\right] \subseteq [0, 1], \\ \quad \mathbb{P}\left[\sigma \Phi^{-1}(t_*) + \beta Y_{2u} \leq Y_1 - Z_1\gamma \leq \sigma \Phi^{-1}(t^*) + \beta Y_{2l}|z\right] \leq t^* - t_*, \text{ a.e. } z \in \mathcal{R}_Z \end{array} \right\}.$$
(51)

$$\mathcal{I}\left(\mathcal{M}, \mathcal{F}_{Y|Z}, \mathcal{R}_Z\right) = \left\{ \begin{array}{l} \theta \in \Theta : \forall \left[t_*, t^*\right] \subseteq [0, 1], \\ \quad \mathbb{P}\left[t_* \leq \Phi\left(\frac{Y_1 - \beta Y_{2u} - Z_1\gamma}{\sigma}\right) \wedge \Phi\left(\frac{Y_1 - \beta Y_{2l} - Z_1\gamma}{\sigma}\right) \leq t^*|z\right] \leq t^* - t_*, \text{ a.e. } z \in \mathcal{R}_Z \end{array} \right\}$$

The identified set can be represented as the set of parameter values θ satisfying the collection of conditional moment inequalities

$$E\left[m\left(\theta; Y, Z, u_*, u^*\right)|z\right] \leq 0, \quad \text{all } u_*, u^* \in \mathbb{R} \text{ s.t. } u_* \leq u^*, \text{ a.e. } z \in \mathcal{R}_Z,$$

with moment function defined as:

$$m\left(\theta; Y, Z, t_*, t^*\right) \equiv 1\left[u_* + \beta Y_{2u} \leq Y_1 - Z_1\gamma \leq u^* + \beta Y_{2l}\right] - \left(\Phi\left(\sigma^{-1}u^*\right) - \Phi\left(\sigma^{-1}u_*\right)\right),$$

equivalently, as the set of parameter values θ satisfying the collection of conditional moment inequalities

$$E\left[m\left(\theta; Y, Z, t_*, t^*\right)|z\right] \leq 0, \quad \text{all } t_*, t^* \text{ s.t. } \left[t_*, t^*\right] \subseteq [0, 1], \text{ a.e. } z \in \mathcal{R}_Z,$$

with moment function defined as:

$$m\left(\theta; Y, Z, t_*, t^*\right) \equiv 1\left[t_* \leq \Phi\left(\frac{Y_1 - \beta Y_{2u} - Z_1\gamma}{\sigma}\right) \wedge \Phi\left(\frac{Y_1 - \beta Y_{2l} - Z_1\gamma}{\sigma}\right) \leq t^*\right] - \left(t^* - t_*\right).$$

No Distributional Restriction. Suppose now that the independence restriction $U \per\!\!\!\perp Z$ and the same additive index structure for g are both imposed, with *no* parametric restriction on the distribution of unobserved heterogeneity, indeed imposing no restriction at all on the distribution of U and Z other than the independence restriction.

If Y_2^* were observed, a location normalization would be required on the distribution of U for identification of the first component of γ, the intercept term in the linear index. Thus it will be prudent to incorporate a location normalization in this model with Y_2^* censored, for example that the median of U given $Z = z$ is zero. Since Y_1 is continuously distributed, there is no scale normalization to be made.[26]

Corollary 3 can be used to obtain an outer region for the structural function, h, equivalently, under the index restriction, for parameters $\theta \equiv (\beta, \gamma)$. To achieve this, start with the condition

---

[26] Using partially identifying models care needs to be taken when considering normalizations. This is because there is scope for the model to carry no information whatsover regarding a parameter, such as the location of additive unobservable U in the censoring model with an intercept present. However, without the normalization imposed, an identified set will still be obtained, but its projection onto the space of such a parameter will be all values in the parameter space!

$$\mathbb{P}\left[u_* + \beta Y_{2u} \le Y_1 - Z_1\gamma \le u^* + \beta Y_{2l}|z\right] \le \Lambda\left(u^*\right) - \Lambda\left(u_*\right), \text{ a.e. } z \in \mathcal{R}_Z, \tag{52}$$

for all $[u_*, u^*] \in \mathsf{Q}$ from (50) above. Noting that $G_U\left(\mathcal{S}\right) = \Lambda\left(u^*\right) - \Lambda\left(u_*\right)$ for any set $\mathcal{S} = [u_*, u^*]$ and following Corollary 3 we also have for all $-\infty < u_* \le u^* < \infty$ and a.e. $z \in \mathcal{R}_Z$,

$$
\begin{aligned}
\Lambda\left(u^*\right) - \Lambda\left(u_*\right) &\le 1 - C_h\left(\mathcal{S}^c|z\right) \tag{53}\\
&= 1 - \mathbb{P}\left[Y_1 - \beta Y_{2l} - Z_1\gamma < u_* \vee Y_1 - \beta Y_{2u} - Z_1\gamma > u^*|z\right]\\
&= \mathbb{P}\left[u_* + \beta Y_{2l} \le Y_1 - Z_1\gamma \le u^* + \beta Y_{2u}|z\right].
\end{aligned}
$$

Define

$$
\begin{aligned}
\underline{G}\left(\theta, u_*, u^*\right) &\equiv \sup_{z \in \mathcal{R}_Z} \mathbb{P}\left[u_* + \beta Y_{2u} \le Y_1 - Z_1\gamma \le u^* + \beta Y_{2l}|z\right],\\
\overline{G}\left(\theta, u_*, u^*\right) &\equiv \inf_{z \in \mathcal{R}_Z} \mathbb{P}\left[u_* + \beta Y_{2l} \le Y_1 - Z_1\gamma \le u^* + \beta Y_{2u}|z\right],
\end{aligned}
$$

each of which are identified for any parameter vector $\theta = (\beta, \gamma)$ from knowledge of $\mathcal{F}_{Y|Z}$ under Restriction A2. Combining (52) and (53) as in Corollary 3, an outer region for the parameter vector θ, is given by

$$\mathcal{I}_\theta^O\left(\mathcal{M}, \mathcal{F}_{Y|Z}, \mathcal{R}_Z\right) = \left\{\theta \in \Theta : \forall [u_*, u^*] \in \mathsf{Q}, \, \underline{G}\left(\theta, u_*, u^*\right) \le \overline{G}\left(\theta, u_*, u^*\right)\right\} \tag{54}$$

where Θ denotes values of θ admitted by model $\mathcal{M}$.

By contrast the identified set for $(\theta, \Lambda\left(\cdot\right))$ can be expressed as

$$\mathcal{I}\left(\mathcal{M}, \mathcal{F}_{Y|Z}, \mathcal{R}_Z\right) = \left\{(\theta, \Lambda\left(\cdot\right)) \in \mathcal{M} : \forall [u_*, u^*] \in \mathsf{Q}, \, \underline{G}\left(\theta, u_*, u^*\right) \le \Lambda\left(u^*\right) - \Lambda\left(u_*\right) \le \overline{G}\left(\theta, u_*, u^*\right)\right\}.$$

In any particular application $\mathcal{I}_\theta^O$ defined in (54) may be the sharp identified set for θ but this cannot be guaranteed. This is because it cannot be guaranteed that for every $\theta \in \mathcal{I}_\theta^O$ there exists a proper distribution function Λ such that the inequality (52) holds for all $u_* \le u^*$, and almost every $z \in \mathcal{R}_Z$. Nevertheless the outer region $\mathcal{I}_\theta^O$ will usually be informative.

Equivalent to (54), the values θ in the outer region $\mathcal{I}_\theta^O$ are those $\theta \in \Theta$ satisfying the moment inequality representation:

$$
\begin{aligned}
E\left[m_1\left(\theta; Y, Z, u_*, u^*\right)|z\right] - E\left[m_2\left(\theta; Y, Z, u_*, u^*\right)|z'\right] &\le 0,\\
\text{all } u_*, u^* &\in \mathbb{R} \text{ s.t. } u_* \le u^*, \text{ a.e. } z, z' \in \mathcal{R}_Z \times \mathcal{R}_Z,
\end{aligned}
$$

where

$$m_1\left(\theta; Y, Z, u_*, u^*\right) \;\equiv\; 1\left[u_* + \beta Y_{2u} \le Y_1 - Z_1\gamma \le u^* + \beta Y_{2l}\right],$$
$$m_2\left(\theta; Y, Z, u_*, u^*\right) \;\equiv\; 1\left[u_* + \beta Y_{2l} \le Y_1 - Z_1\gamma \le u^* + \beta Y_{2u}\right].$$

7.1.3 Mean Independence

Now suppose the linear index restriction (47) remains in place, but the stochastic independence restriction $U \perp\!\!\!\perp Z$ is replaced with the weaker conditional mean restriction $E\left[U|Z = z\right] = 0$ a.e. $z \in \mathcal{R}_Z$, equivalently Restriction MI from Section 6.2.

The random set $\mathcal{U}\left(Y, Z; h\right)$ in this model is given by the interval

$$\mathcal{U}\left(Y, Z; h\right) = \left[Y_1 - Z_1\gamma - \beta Y_{2u},\, Y_1 - Z_1\gamma - \beta Y_{2l}\right],$$

rendering application of Theorem 5 particularly straightforward. The conditional Aumann expectation of $\mathcal{U}\left(Y, Z; h\right)$ given $Z = z$ is simply the interval spanning the conditional expectation of the lower endpoint to the conditional expectation of the upper endpoint. Therefore the condition that

$$0 \in \mathbb{E}\left[\mathcal{U}\left(Y, Z; h\right)|z\right], \text{ a.e. } z \in \mathcal{R}_Z$$

is equivalent to

$$E\left[Y_1 - Z_1\gamma - \beta Y_{2u}|z\right] \le 0 \le E\left[Y_1 - Z_1\gamma - \beta Y_{2l}|z\right] \text{ a.e. } z \in \mathcal{R}_Z.$$

Thus, the identified set for $\theta \equiv (\beta, \gamma)$ is

$$\mathcal{I}_\theta\left(\mathcal{M}, \mathcal{F}_{Y|Z}, \mathcal{R}_Z\right) = \left\{\theta \in \Theta : \underline{\mathbb{E}}\left(\theta\right) \le 0 \le \overline{\mathbb{E}}\left(\theta\right)\right\},$$

where
$$\underline{\mathbb{E}}\left(\theta\right) \equiv \sup_{z \in \mathcal{R}_Z} E\left[Y_1 - Z_1\gamma - \beta Y_{2u}|z\right], \quad \overline{\mathbb{E}}\left(\theta\right) \equiv \inf_{z \in \mathcal{R}_Z} E\left[Y_1 - Z_1\gamma - \beta Y_{2l}|z\right]$$

and as before Θ denotes values admitted by a model $\mathcal{M}$.

7.1.4 Quantile Independence

Finally, consider a model imposing the linear index structure (47) coupled with Restriction QI that the conditional τ-quantile of U given $Z = z$ is zero for almost every $z \in \mathcal{R}_Z$.

It remains under (47) that

$$\mathcal{U}\left(Y, Z; h\right) = \left[Y_1 - Z_1\gamma - \beta Y_{2u},\, Y_1 - Z_1\gamma - \beta Y_{2l}\right],$$

and the identified set for the structural function, h, in this parametric model is the identified set for $\theta \equiv (\beta, \gamma)$. As in Section 7.1.3 once again denote the parameter space and identified set for θ as Θ and $\mathcal{I}_\theta$, respectively. Applying Theorem 6 the identified set of values of the parameter θ is

$$\mathcal{I}_\theta\left(\mathcal{M}, \mathcal{F}_{Y|Z}, \mathcal{R}_Z\right) = \left\{\theta \in \Theta : \sup_{z \in \mathcal{R}_Z} F_{Y|Z}\left[\{y : y_1 < z_1\gamma + \beta y_{2l}\}|z\right] \leq \tau \leq \inf_{z \in \mathcal{R}_Z} F_{Y|Z}\left[\{y : y_1 < z_1\gamma + \beta y_{2u}\}|z\right]\right\},$$

(55)

equivalently,

$$\mathcal{I}_\theta\left(\mathcal{M}, \mathcal{F}_{Y|Z}, \mathcal{R}_Z\right) = \left\{\theta \in \Theta : \sup_{z \in \mathcal{R}_Z}\left(q_{\underline{V_\theta}|Z}(\tau|z) - z_1\gamma\right) \leq 0 \leq \inf_{z \in \mathcal{R}_Z}\left(q_{\overline{V_\theta}|Z}(\tau|z) - z_1\gamma\right)\right\},$$

where $q_{\underline{V_\theta}|Z}(\tau|z)$ and $q_{\overline{V_\theta}|Z}(\tau|z)$ are conditional τ-quantile functions of random variables respectively $\underline{V_\theta} \equiv Y_1 - \beta Y_{2u}$ and $\overline{V_\theta} \equiv Y_1 - \beta Y_{2l}$.

Using (55) the identified set $\mathcal{I}_\theta$ can be represented as the values of θ that satisfy the moment inequalities

$$E\left[m_1(\theta; Y, Z)|z\right] \leq 0 \text{ and } E\left[m_2(\theta; Y, Z)|z\right] \leq 0, \text{ a.e. } z \in \mathcal{R}_Z,$$

where

$$\begin{aligned}
m_1(\theta; Y, Z) &\equiv 1\left[Y_1 < Z_1\gamma + \beta Y_{2l}\right] - \tau, \\
m_2(\theta; Y, Z) &\equiv \tau - 1\left[Y_1 \leq Z_1\gamma + \beta Y_{2u}\right].
\end{aligned}$$

7.2 Numerical Illustrations

In this section numerical illustrations of identified sets are provided for the interval censored endogenous variable model with the linear index restriction of (47). We first consider the identified set obtained under the restriction that $U \sim N(0, \sigma)$ and $U \perp\!\!\!\perp Z$, that is the Gaussian unobservable case above with identified set given by (51). Then we consider a much less restrictive model in which there is no parametric restriction on the distribution of U and instead of the stochastic independence condition there is a condition requiring the conditional median of U given $Z = z$ to be independent of z.

Supplementary Mathematica[27] code used to produce the numerical illustrations reported here is available on-line.[28] Specifically, there are (1) files documenting parameter vectors and associated discrepancy measures of the distance from these parameter vectors to the identified set for each model considered used to produce graphs of identified sets and projections, (2) Mathematica code to check the validity of these discrepancies and to generate discrepancies for new points, and (3) the Mathematica code which draws Figures 2 - 6.

---

[27] Wolfram Research, Inc. (2019).

[28] https://drive.google.com/open?id=110JMgGMJWqWOAvWXHALgl3Cq_zDM2JtO.

7.2.1 Specification of structures

To generate particular probability distributions $\mathcal{F}_{Y|Z}$ of observable variables (Y, Z) to employ in the illustrative calculations it is necessary to specify a particular complete structure. Here a triangular Gaussian structure is employed as follows:

$$
\begin{aligned}
Y_1 &= g + bY_2^* + U, \\
Y_2^* &= d_0 + d_1 Z + V.
\end{aligned}
$$

with a binary scalar instrumental variable $Z \in \mathcal{R}_Z = \{-1, 1\}$, with $(U, V) \perp\!\!\!\perp Z$, and with

$$
\begin{bmatrix} U \\ V \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} s^2 & s_{1v} \\ s_{1v} & s_{vv} \end{bmatrix} \right).
$$

In this structure there are no exogenous covariates in the structural equation for Y_1. The case where Z has binary support makes for easy calculations in the numerical illustration, but richer support for the instrument would provide greater identifying power and smaller identified sets. The Roman symbols (b, g, s) in the definition of the complete structure are chosen to correspond to the Greek symbols (β, γ, σ) that appear in the structural function of the incomplete GIV model with interval censored endogenous explanatory variable.

We specify two alternative censoring processes each revealing to which of a collection of mutually exclusive intervals Y_2^* belongs. Such censoring processes are common in practice, for instance when interval bands are used for income in surveys. Specifically, we specify a sequence of J nonstochastic intervals, $I_1, I_2, \ldots, I_J$ with $I_j \equiv (c_j, c_{j+1}]$ and $c_j < c_{j+1}$ for all $j \in \{1, \ldots, J\}$. The censoring process is such that

$$
\forall j \in \{1, \ldots, J\}, \quad (Y_{2l}, Y_{2u}) = (c_j, c_{j+1}) \Leftrightarrow Y_2^* \in I_j.
$$

In the first set of examples two structures denoted ST1 and ST2 are considered, both with parameter values

$$
g = 0, \quad b = 1, \quad d_0 = 0, \quad d_1 = 1, \quad s^2 = 0.5, \quad s_{1v} = 0.25, \quad s_{vv} = 0.5, \tag{56}
$$

and interval censoring endpoints $c_1, \ldots, c_J$ listed in Table 2. In ST1, Y_2^* is censored into 8 intervals $I_j = (c_j, c_{j+1}]$ with endpoints given by the normal quantile function evaluated at 9 equally spaced values in $[0, 1]$, inclusive of 0 and 1. In ST2, Y_2^* is censored into 12 such intervals with endpoints given by the normal quantile function evaluated at 13 equally spaced values.

	c_1	c_2	c_3	c_4	c_5	c_6	c_7	c_8	c_9	c_{10}	c_{11}	c_{12}	c_{13}
ST1	$-\infty$	-1.15	-0.67	-0.32	0.00	0.32	0.67	1.15	$+\infty$	-	-	-	-
ST2	$-\infty$	-1.38	-0.97	-0.67	-0.43	-0.21	0.00	0.21	0.43	0.67	0.97	1.38	$+\infty$

Table 2: Endpoints of censoring process intervals in DGP1 and DGP2.

7.2.2 Calculation of sets

Using probabilities generated by the structures ST1 and ST2, outer sets containing the identified set were calculated for the parameters $\theta \equiv (\beta, \gamma, \sigma)$ of a GIV model which specifies

$$Y_1 = \gamma + \beta Y_2^* + U$$

with $U \perp\!\!\!\perp Z$, with Y_2^* interval censored into 8 or 12 intervals as specified above and with $U \sim N(0, \sigma^2)$. We refer to these as "outer sets" because we employ a finite selection from the infinite collection of inequalities which characterize the sharp identified set.

To compute the outer sets it is necessary to compute the probability

$$\mathbb{P}\left[\sigma \Phi^{-1}(t_*) + \beta Y_{2u} \leq Y_1 - Z_1 \gamma \leq \sigma \Phi^{-1}(t^*) + \beta Y_{2l} | z\right]$$

$$= \sum_{j=1}^{J} \mathbb{P}\left[\left(\sigma \Phi^{-1}(t_*) + \beta c_{j+1} \leq Y_1 - \gamma \leq \sigma \Phi^{-1}(t^*) + \beta c_j\right) \wedge (Y_2^* \in I_j) | z\right] \quad (57)$$

for a selection of intervals $[t_*, t^*] \subseteq [0, 1]$. This is the probability that appears in the characterization (51) of the identified set. This sum of bivariate Gaussian probabilities is easily calculated in the structures under consideration using the joint distribution of Y_1 and Y_2^* which is as follows.

$$\begin{bmatrix} Y_1 \\ Y_2^* \end{bmatrix} | Z = z \sim N\left(\begin{bmatrix} g + bd_0 + bd_1 z \\ d_0 + d_1 z \end{bmatrix}, \begin{bmatrix} s^2 + 2bs_{1v} + b^2 s_{vv} & s_{1v} + bs_{vv} \\ s_{1v} + bs_{vv} & s_{vv} \end{bmatrix}\right) \quad (58)$$

Using data to produce an estimate of the identified set one would proceed using estimates of the conditional probability (57). For example, with a discrete instrument $Z \in \{\bar{z}_1, \ldots, \bar{z}_S\}$ and R realizations, (y_{1r}, d_{jr}, z_r), $r \in \{1, \ldots, R\}$ where d_{jr} denotes a realization of $D_j \equiv 1[Y_2^* \in I_j]$, which is observable, there is the following estimator of the conditional probability given $Z = \bar{z}_s$.

$$\hat{p}(\theta, \bar{z}_s) \equiv \frac{1}{\sum_{r=1}^{R} 1[z_r = \bar{z}_s]} \sum_{j=1}^{J} \sum_{r=1}^{R} \{1\left[\sigma \Phi^{-1}(t_*) + \beta c_{j+1} \leq y_{1r} - \gamma \leq \sigma \Phi^{-1}(t^*) + \beta c_j\right]$$

$$\times d_{jr} \times 1\left[z_r = \bar{z}_s\right]\}, \quad s \in \{1, \ldots, S\}.$$

With a continuous instrument one could use a prediction calculated from a nonparametric estimator of the regression of $\sum_{j=1}^{J} \left(1[\sigma \Phi^{-1}(t_*) + \beta c_{j+1} \leq Y_{1r} - \gamma \leq \sigma \Phi^{-1}(t^*) + \beta c_j] \times D_j\right)$ on Z.

Computation of the (sharp) identified set delivered by (51) would require consideration of every interval $[t_*, t^*] \subset [0,1]$. Obviously a finite selection must be made. In the calculations reported here we use various combinations of collections Q_N of intervals from the collection of all possible intervals $[t_*, t^*] \subseteq [0,1]$. Each collection of intervals Q_N comprises the $N(N+1)/2 - 1$ super-diagonal elements of the following $(N+1) \times (N+1)$ array of intervals that remain after excluding the interval $[0,1]$. Here $n \equiv 1/N$.

$$
\begin{bmatrix}
[0,0] & [0,n] & [0,2n] & [0,3n] & \cdots & \cdots & & \cdots & & [0,1] \\
- & [n,n] & [n,2n] & [n,3n] & \cdots & \cdots & & \cdots & & [n,1] \\
- & - & [2n,2n] & [2n,3n] & \cdots & \cdots & & \cdots & & [2n,1] \\
- & - & - & [3n,3n] & \cdots & \cdots & & & & [3n,1] \\
\vdots & \vdots & \vdots & \vdots & \ddots & & & & & \vdots \\
\vdots & \vdots & \vdots & \vdots & & \ddots & & & & \vdots \\
- & - & - & - & \cdots & & [(N-1)n,(N-1)n] & & [(N-1)n,1] \\
- & - & - & - & \cdots & & & - & & [1,1]
\end{bmatrix}
$$

The inequalities of (51) applied to the intervals of any collection of test sets Q_N defines an outer set relative to the identified set, with larger collections of test sets providing tighter enveloping of the identified set.

Figure 2 shows three dimensional (3D) plots of outer sets for (β, γ, σ) obtained using structure ST1 in which endogenous Y_2^* is censored into 8 intervals. Outer sets using $N \in \{5,7,9\}$ in the lower part of the Figure are slightly smaller than those using just $N = 5$.[29] There is a small reduction in the size of the outer sets in moving from $N = 5$ to $N = \{5,7\}$, but only a tiny change on including also the inequalities obtained with $N = 9$. Only the outer sets obtained using $N = 5$ and $N \in \{5,7,9\}$ are shown in Figure 2.

Figure 3 shows two dimensional projections of the outer set onto the space of each pair of parameters in turn. Here there is censoring of Y_2^* into 8 intervals and $N \in \{5,7,9\}$. The boundaries and surfaces of sets are drawn as convex hulls of the points found to lie inside the sets and projections considered.[30] We have no proof of the convexity of the outer sets in general, but careful investigation of points found to lie in the sets strongly suggested that in the cases considered the sets are convex.

Figure 4 shows the 3D outer set for structure ST2 in which Y_2^* is censored into 12 intervals The collection of inequalities obtained when $N \in \{5,7,9\}$ is employed, as in the case illustrated in the lower part of Figure 2. Compared with that Figure the outer set is smaller, as expected given the finer granularity of intervals with 12 rather than 8 bins. Figure 5 shows two dimensional projections of this outer set, again projecting onto each pair of parameter components. These

---

[29] The notation $N \in \{N_1, N_2, ..., N_R\}$ corresponds to the use of the collection of test sets $\mathsf{Q}_{N_1} \cup \mathsf{Q}_{N_2} \cdots \cup \cdots \cup \mathsf{Q}_{N_R}$.

[30] The 3D figures were produced using the `TetGenConvexHull` function available *via* the `TetGenLink` package in `Mathematica 9`. 2D figures were drawn using `Mathematica`'s `ConvexHull` function.

projections further illustrate the extent of the reduction in the size of the outer region for ST2 relative to ST1, by comparison with Figure 3.

In a second set of numerical illustrations we employ the same triangular Gaussian structure as in ST1 and ST2 with parameter values as specified in (56). We consider two new censoring processes, where Y_2^* is again observed only to lie in one of a fixed set of bins, but where now these bins are set to be of a fixed width. We consider fixed bins, first with width 0.4:

$$\ldots\ldots, \ (-0.8, -0.4], \ (-0.4, 0.0], \ (0, 0.4], \ (0.4, 0.8], \ \ldots\ldots$$

and then with width 0.2:

$$\ldots\ldots, \ (-0.4, -0.2], \ (-0.2, 0.0], \ (0, 0.2], \ (0.2, 0.4], \ \ldots\ldots$$

We compare the identifying power of alternative restrictions on unobserved heterogeneity, in both cases imposing the linear functional form

$$Y_1 = \gamma + \beta Y_2^* + U$$

with excluded instrumental variable Z. We consider the parametric Gaussian restriction on unobserved heterogeneity under which $U \sim N\left(0, \sigma^2\right)$ and $U \perp\!\!\!\perp Z$, and compare to a case with no parametric restriction on the distribution and instead only a restriction that the conditional median of U given $Z = z$ is zero: $q_{U|Z}\left(0.5|z\right) = 0$, a.e. $z \in \mathcal{R}_Z = \{-1, 1\}$. This semiparametric model has no scale parameter σ, so we focus on the identified set for (β, γ), comparing the identified sets for (β, γ) under the conditional median restriction with the projection onto the (β, γ) plane of the set for (β, γ, σ) obtained using the Gaussian model.

Figure 6 illustrates identified sets obtained for bin widths 0.4 (top panels) and 0.2 (bottom panels), as well as for a weaker instrument case with $d_1 = 1$ (left panels) and a stronger instrument case with $d_1 = 1.5$ (right panels). In the triangular structures employed to generate the probabilities used in the calculations, the parameter d_1 is the coefficient multiplying the instrument Z in the equation determining the value of the censored endogenous variable Y_2^*. With a higher value of d_1 the value of Y_2^* is more sensitive to changes in the instrument Z. As we might expect, identified sets when $d_1 = 1.5$ are smaller than those for the case $d_1 = 1$, as are sets obtained when the bin width is only 0.2 rather than 0.4. Outer sets obtained under the Gaussian restriction are shown in light blue; the larger sets obtained under the conditional median independence restriction alone are shown in dark blue.

Sharp identified sets obtained from the model requiring unobservable U to be Gaussian are necessarily contained in those obtained from a model imposing the less restrictive zero conditional quantile restriction. The identified sets obtained under the quantile restriction are sharp unlike

those obtained under the Gaussian restriction which use a finite selection from the infinity of inequalities that define the sharp set. However it seems unlikely that there is any major refinement to be obtained by considering more inequalities.

In the calculations reported here the difference between the identified sets obtained under these different restrictions is not substantial. A partial explanation for this observation may be the fixed-width binning setup. Indeed, under that censoring process with the given triangular data-generating structure, it can be shown that under a distribution-free independence restriction - that is $U \perp\!\!\!\perp Z$ but with the distribution of U otherwise unrestricted but for a zero median location normalization - the identified set is identical to that obtained under the conditional median restriction alone. This is not generally the case. For other censoring processes (not reported here) outer sets under the distribution-free independence restriction lie well inside the set obtained under the weaker conditional quantile restriction.

In these numerical illustrations only one excluded binary instrumental variable has been employed which is in a sense a worst case scenario. Nevertheless the identified sets that result are quite informative. In cases where more instrumental variables are employed and with richer support we can expect much more informative identified sets.

8   Application II: IV Models for Binary Outcomes

In this section we return to Example 2 of Section 3.2 in which a binary outcome Y_1 takes the value 1 if and only if a continuously distributed random variable U takes a value weakly greater than some function $g(Y_2, Z)$, whose arguments are potentially endogenous Y_2 and exogenous Z.[31]

$$
Y_1 = \begin{cases} 0 & , & 0 \leq & U & \leq g(Y_2, Z) \\[2ex] 1 & , & g(Y_2, Z) \leq & U & \leq 1 \end{cases}
\tag{59}
$$

The model imposes a restriction requiring U and Z to be independently distributed. Under this restriction the distribution of U can be normalized uniformly distributed on $[0, 1]$, a normalization imposed from now on. In practice there will need to be restrictions on the influence of Z on $g(\cdot, \cdot)$, for example exclusion restrictions, if the model is to have any identifying power when Y_2 is endogenous. We will come to this later.

Many parametric and semiparametric econometric models for binary outcomes used in empirical practice fall in this class of models, including probit and logit models and the semiparametric model used in Klein and Spady (2003).[32] The model studied here is distinctive in that it allows the possibility that some explanatory variables are endogenous. Chesher (2010, 2013) provides a

---

[31]Since U is continuously distributed the ambiguity regarding the value of Y_1 when $g(Y_2, Z) = U$ is of no consequence. See Footnote 14.

[32]Example 2* on page 20 makes explicit how specification (59) nests these models.

detailed analysis of this binary outcome model. Chesher and Smolinski (2012) study an ordered outcome extension of this model.[33]

8.1 The Identified Set

Let $\mathcal{I}(\mathcal{M}, \mathcal{F}_{Y|Z}, \mathcal{R}_Z)$ denote the identified set of functions $g(\cdot, \cdot)$, or, with a parametric specification, the identified set of parameters, delivered by a model $\mathcal{M}$ and the conditional distribution of $Y = (Y_1, Y_2)$ conditional on Z for each value z in the support of Z. Characterizations of the set $\mathcal{I}(\mathcal{M}, \mathcal{F}_{Y|Z}, \mathcal{R}_Z)$ are provided under various specifications of the model $\mathcal{M}$.

Throughout this section the restriction that there is a single scalar source of unobservable heterogeneity is maintained. This restriction can be relaxed in a GIV analysis as shown in Chesher and Rosen (2014) which studies a binary outcome random coefficients model admitting multiple sources of heterogeneity.

A structural function capturing the restrictions of the threshold crossing index model is given in (11), and the U-level sets of the structural function are simply closed intervals, as follows.

$$
\mathcal{U}(y, z; g) = \begin{cases} [0, g(y_2, z)] & , \quad y_1 = 0 \\ \\ [g(y_2, z), 1] & , \quad y_1 = 1 \end{cases} \tag{60}
$$

Since the structural function is characterized by the function $g(\cdot, \cdot)$, "g" is used as the argument of the U-level set.

Theorem 3 tells us that core determining sets are connected unions of the level sets that comprise the support of the random set $\mathcal{U}(Y, Z; g)$. All such unions are closed intervals, subsets of the unit interval, $[0, 1]$, with either 0 as a lower endpoint or 1 as an upper endpoint. The collection

$$
\mathsf{Q} = \{[0, u], [u, 1] : u \in [0, 1]\}
$$

is a core determining collection.

The identified set of functions $g(\cdot, \cdot)$ is

$$
\mathcal{I}_g(\mathcal{M}, \mathcal{F}_{Y|Z}, \mathcal{R}_Z) = \{g(\cdot, \cdot) : \forall \mathcal{S} \in \mathsf{Q}, \quad G_U(\mathcal{S}) \geq C_g(\mathcal{S}|z), \quad \text{a.e. } z \in \mathcal{R}_Z\} \tag{61}
$$

where $C_g(\mathcal{S}|z)$ is the conditional containment probability for the random set $\mathcal{U}(Y, Z; g)$ given $Z = z$.

$$
C_g(\mathcal{S}|z) \equiv F_{Y|Z}(\{y : \mathcal{U}(y, z; g) \subseteq \mathcal{S}\}|z)
$$

---

[33]These three papers were completed before the development of the general approach set out in CR17. They deliver the same identified sets as CR17 but provide laborious constructive proofs of sharpness, which CR17 renders unnecessary.

The probability $G_U(\mathcal{S})$ in (61) is simply the length of the interval $\mathcal{S}$ because U is normalized uniform on $[0, 1]$.

Consider the containment functional probability $C_g(\mathcal{S}|z)$ that appears in (61) and let $\mathbb{P}(\mathcal{E}|z)$ denote the conditional probability of an event $\mathcal{E}$ delivered by the distribution $F_{Y|Z}(\cdot|z) \in \mathcal{F}_{Y|Z}$.

The random U-level set $\mathcal{U}(Y, z; g)$ has realizations which are a subset of $[0, u]$ only when $Y_1 = 0$ and $g(Y_2, z) \leq u$. The random set $\mathcal{U}(Y, z; h)$ has realizations which are a subset of $[u, 1]$ only when $Y_1 = 1$ and $g(Y_2, z) \geq u$. There is therefore the following characterization of $\mathcal{I}_g(\mathcal{M}, \mathcal{F}_{Y|Z}, \mathcal{R}_Z)$: the identified set of functions $g(\cdot, \cdot)$ under the restrictions of the model $\mathcal{M}$.

$$\mathcal{I}_g(\mathcal{M}, \mathcal{F}_{Y|Z}, \mathcal{R}_Z) = \mathcal{A}_0 \cap \mathcal{A}_1$$

where

$$\mathcal{A}_0 = \{g(\cdot, \cdot) : \forall u \in (0, 1) \quad u \geq \mathbb{P}[Y_1 = 0 \wedge g(Y_2, z) \leq u|z], \text{ a.e. } z\}$$

$$\mathcal{A}_1 = \{g(\cdot, \cdot) : \forall u \in (0, 1) \quad 1 - u \geq \mathbb{P}[Y_1 = 1 \wedge g(Y_2, z) \geq u|z], \text{ a.e. } z\}.$$

This can be expressed as follows.

$$\mathcal{I}_g(\mathcal{M}, \mathcal{F}_{Y|Z}, \mathcal{R}_Z) = \{g(\cdot, \cdot) : \forall u \in (0, 1)$$
$$1 - \mathbb{P}[Y_1 = 1 \wedge g(Y_2, z) \geq u|z]$$
$$\geq u \geq$$
$$\mathbb{P}[Y_1 = 0 \wedge g(Y_2, z) \leq u|z], \text{ a.e. } z\} \quad (62)$$

If one uses a model employing a parametric specification then these expressions define identified sets of parameters. In an application in Section 8.2 we consider IV probit type models in which $g(y_2, z) = \Phi(\beta_0 + \beta_1 z_1 + \alpha y_2)$ where $\Phi(\cdot)$ is the standard normal distribution function.

8.1.1 The power of instruments

The upper bounding probability in (62) can be written as $P[Y_1 = 0 \vee g(Y_2, z) < u|z]$, from which it is trivial that for any fixed (u, z), the upper bounding probability in (62) is at least equal to the lower bounding probability. It follows that these inequalities can only place restrictions on the threshold function $g(\cdot, \cdot)$ when, for one or more values of u, the restrictions that the model places on the threshold function and the support of Z are such that there exists variation in z for which the bounding probabilities in (62) vary while the function $g(\cdot, z)$ remains unchanged.

A leading case of interest is one in which there is an exclusion restriction so that, with $Z \equiv (Z_1, Z_2)$, Z_2 is excluded from the threshold function which then becomes $g(y_2, z_1)$.

In this case, in (62) at each value of Z_1 only the infimum of the upper bounding probabilities and the supremum of the lower bounding probabilities over the values of excluded Z_2 are relevant

for identification and the identified set is as follows.

$$\mathcal{I}_g(\mathcal{M}, \mathcal{F}_{Y|Z}, \mathcal{R}_Z) = \{g(\cdot, \cdot) : \forall u \in (0,1)$$

$$\inf_{z_2 \in \mathcal{R}_{Z_2|z_1}} (1 - \mathbb{P}[Y_1 = 1 \wedge g(Y_2, z_1) \geq u|z])$$

$$\geq u \geq$$

$$\sup_{z_2 \in \mathcal{R}_{Z_2|z_1}} \mathbb{P}[Y_1 = 0 \wedge g(Y_2, z_1) \leq u|z], \text{ a.e. } z_1\} \quad (63)$$

Here there is one pair of upper and lower bounding probabilities at each value z_1 in the support of the included exogenous variables, Z_1, and $\mathcal{R}_{Z_2|z_1}$ denotes the conditional support of excluded Z_2 given $Z_1 = z_1$.

Richer support for Z_2, and stronger dependence on Z_2 of the bounding conditional probabilities that appear in the definitions lead to smaller identified sets. There will typically be richer support for Z_2 when there are more instruments in this list.

In the application set out in Section 8.2 there are two instrumental variables. One is quite weak and delivers quite large disconnected sets for structural features of interest; the other is strong, delivering a small, connected set. Using both instruments we get just a slight refinement of the set obtained using the strong instrument.

8.1.2 Monotone index restriction

In many models employed in econometric practice the threshold function, $g(y_2, z)$, is restricted to be a monotone function of a linear index, that is $g(y_2, z_1) = s(z_1\beta + y_2\alpha)$ with $s(\cdot)$ strictly monotone, normalized increasing, where z_1 and y_2 denote row-vectors, which are possibly multivariate and are conformable with parameter vectors β and α. This restriction appears for example in probit and logit models.

Under the monotone index restriction the inequalities in (63) can be written

$$\inf_{z_2 \in \mathcal{R}_{Z_2|z_1}} \left(1 - \mathbb{P}[Y_1 = 1 \wedge z_1\beta + Y_2\alpha \geq s^{-1}(u)|z]\right)$$

$$\geq u \geq$$

$$\sup_{z_2 \in \mathcal{R}_{Z_2|z_1}} \mathbb{P}[Y_1 = 0 \wedge z_1\beta + Y_2\alpha \leq s^{-1}(u)|z], \text{ a.e. } z_1$$

and then after the change of variable $t = s^{-1}(u)$ there is the following identified set for $(\alpha, \beta, s(\cdot))$.

$$\mathcal{I}_{(\alpha,\beta,s)}(\mathcal{M}, \mathcal{F}_{Y|Z}, \mathcal{R}_Z) = \{\alpha, \beta, \text{ monotone strictly increasing } s(\cdot) : \forall t \in \mathbb{R}$$

$$\inf_{z_2 \in \mathcal{R}_{Z_2|z_1}} (1 - \mathbb{P}[Y_1 = 1 \wedge z_1\beta + Y_2\alpha \geq t|z])$$

$$\geq s(t) \geq$$

$$\sup_{z_2 \in \mathcal{R}_{Z_2|z_1}} \mathbb{P}[Y_1 = 0 \wedge z_1\beta + Y_2\alpha \leq t|z], \text{ a.e. } z_1\} \quad (64)$$

In practice if $s(\cdot)$ is an unknown strictly monotone function, one will normalize one of the index coefficients, for example setting an *a priori* non-zero element of β equal to 1.

The upper and lower bounding probabilities in (64) are weakly increasing functions of t. If for some value of (α, β) they cross then there is no monotone function that can pass between the bounding probabilities and that value of (α, β) does not lie in the identified set.

If for some value of (α, β) the upper bounding probability is at least equal to the lower bounding probability for all $t \in \mathbb{R}$ then there exists at least one monotone function $s(\cdot)$ that passes between the upper and lower bounding probability functions and that value of (α, β) lies in the identified set of parameter values. There is therefore the identified set of index coefficients which is a projection of the set (64), as follows.

$$\mathcal{I}_{(\alpha,\beta)}(\mathcal{M}, \mathcal{F}_{Y|Z}, \mathcal{R}_Z) = \{\alpha, \beta : \forall t \in \mathbb{R}$$

$$\inf_{z_2 \in \mathcal{R}_{Z_2|z_1}} (1 - \mathbb{P}[Y_1 = 1 \wedge z_1\beta + Y_2\alpha \geq t|z]) \geq$$

$$\sup_{z_2 \in \mathcal{R}_{Z_2|z_1}} \mathbb{P}[Y_1 = 0 \wedge z_1\beta + Y_2\alpha \leq t|z], \text{ a.e. } z_1\}$$

In a parametric model, for example a probit model, the function $s(\cdot)$ is the standard normal distribution function. The identified set of index coefficients $\mathcal{I}_{(\alpha,\beta)}(\mathcal{M}, \mathcal{F}_{Y|Z}, \mathcal{R}_Z)$ is then the set of (α, β) that satisfy the same conditions as in (64), but with $s(\cdot)$ fixed at $\Phi(\cdot)$. When that set is empty there is no value of (α, β) in that set such that the normal distribution function can pass between the upper and lower bounding probability functions. In that case the probit model is misspecified and the identified set of parameter values is empty.

8.1.3 Discrete endogenous variables

Before moving on to an application consider the case in which Y_2, possibly a vector, has finite support, say with K points of support, $(y_2^1, \ldots, y_2^K)$. This arises in the application set out in Section 8.2.

At each value z of Z the function $g(\cdot, \cdot)$ is characterized by a point, $\gamma(z)$, in the unit K-cube.

$$\gamma(z) \equiv (\gamma_1(z), \ldots \gamma_K(z)), \quad \gamma_k(z) \equiv g(y_2^k, z), \quad k \in \{1, \ldots, K\}$$

Define $\gamma \equiv \{\gamma(z) : z \in \mathcal{R}_Z\}$. Consider a particular element of $\gamma(z)$, $\gamma_k(z)$, which is not the largest element in $\gamma(z)$. Let $\bar{\gamma}_k(z)$ denote the smallest amongst the elements of $\gamma(z)$ that are larger than $\gamma_k(z)$, that is:

$$\bar{\gamma}_k(z) \equiv \min_{j \neq k} \{\gamma_j(z) : \gamma_j(z) > \gamma_k(z)\}.$$

For any value $u \in [\gamma_k(z), \bar{\gamma}_k(z))$ the event $g(Y_2, z) \leq u$ occurs if and only if $Y_2 \in \{y_2^j : \gamma_j(z) \leq \gamma_k(z)\}$. So only values $u \in \{\gamma_1(z), \ldots \gamma_K(z))\}$ are instrumental in defining the identified set and (62) can be written as follows.

$$\mathcal{I}_\gamma(\mathcal{M}, \mathcal{F}_{Y|Z}, \mathcal{R}_Z) = \{\gamma : \forall k \in \{1, \ldots, K\}$$
$$1 - \mathbb{P}[Y_1 = 1 \wedge Y_2 \in \{y_2^j : \gamma_j(z) \geq \gamma_k(z)\}|z]$$
$$\geq \gamma_2^k(z) \geq$$
$$\mathbb{P}[Y_1 = 0 \wedge Y_2 \in \{y_2^j : \gamma_j(z) \leq \gamma_k(z)\}|z], \text{ a.e. } z\} \quad (65)$$

The probabilities in (65) are determined by the ordering of the elements of $\gamma(z)$. For each of the $K!$ possible orderings of these elements the inequalities in (62) define an intersection of linear half spaces and thus a convex polytope for each element $\gamma(z)$, $z \in \mathcal{R}_Z$. The identified set for γ is a union of these polytopes, some of which may be empty, and this union may not be convex nor even connected.

The situation is illustrated in Chesher (2013) for an example in which Y_2 has three points of support and there is no exogenous variable affecting the threshold function. The identified set for the three values of the nonparametrically specified threshold function comprises the union of up to 6 convex polytopes in the unit cube. Progressively increasing the predictive power of the instrumental variable causes the convex subsets of the identified set associated with each ordering to become successively empty, eventually leaving just one, showing that in the case considered the ordering of the elements of γ, effectively the shape of the threshold function, can be identified.

Imposing shape restrictions will render some orderings inadmissible. When Y_2 is scalar with support $y_2^1 \leq \cdots \leq y_2^K$ and there is a monotonicity restriction so that for all z either $\gamma(z)$ is an increasing sequence or a decreasing sequence then the identified set of values of γ is as follows.

$$\mathcal{I}_\gamma(\mathcal{M}, \mathcal{F}_{Y|Z}, \mathcal{R}_Z) = \mathcal{A}_\uparrow \cup \mathcal{A}_\downarrow$$

$$\mathcal{A}_\uparrow = \{\gamma \text{ increasing} : \forall k \in \{1, \ldots, K\}$$

$$1 - \mathbb{P}[Y_1 = 1 \wedge Y_2 \geq y_2^k | z]$$

$$\geq \gamma_k(z) \geq$$

$$\mathbb{P}[Y_1 = 0 \wedge Y_2 \leq y_2^k | z], \text{ a.e. } z\}$$

$$\mathcal{A}_\downarrow = \{\gamma \text{ decreasing} : \forall k \in \{1, \ldots, K\}$$

$$1 - \mathbb{P}[Y_1 = 1 \wedge Y_2 \leq y_2^k | z]$$

$$\geq \gamma_k(z) \geq$$

$$\mathbb{P}[Y_1 = 0 \wedge Y_2 \geq y_2^k | z], \text{ a.e. } z\}$$

This also applies in the case with $K = 2$ points of support for Y_2 which arises in the application studied in the next section. With strong enough instruments one of the sets $\mathcal{A}_\uparrow$ and $\mathcal{A}_\downarrow$ may be empty in which case the direction of the effect of Y_2 on the threshold function is identified. Both sets may be empty in which case the monotonicity restriction can be rejected.

8.2 Application

To illustrate the use of GIV models in practice and to explain in detail how estimation and inference proceeds we use the data on female labor force participation and family size employed in Angrist and Evans (1998) (AE98).[34] The data comes from the US 1980 Census Public Use Microsamples giving information on $254,654$ married mothers aged 21-35 in 1980 with 2 or more children and oldest child less than 18.[35]

The binary outcome we study is Y_1, equal to 1 if a woman worked for pay in 1979 and 0 otherwise. The potentially endogenous variable, Y_2, is binary, equal to 1 for women having three or more children, equal to zero for women having two children. All women in the sample have at least two children.

In the first model considered the threshold function depends on just one variable, the potentially endogenous family size indicator. Since this is a binary indicator this is effectively a nonparametric specification. We then consider models in which the threshold function depends on an exogenous variable as well as on the family size indicator. The variable we use is a binary variable indicating whether a mother has more than twelve years of education. A nonparametric specification is employed first and then a probit-type specification. Estimates of identified sets and confidence

---

[34] See also Angrist (2001) and Angrist and Pischke (2009).

[35] The original data source for the data used by Angrist and Evans (1998) is the U.S. Department of Commerce, Bureau of the Census, 1983, currently available from the Harvard Dataverse at `https://dataverse.harvard.edu/api/access/datafile/:persistentId?persistentId=doi:10.7910/DVN/8RYANI/HKEUEN`.

regions for sets and their projections are presented and we explain in detail how these are produced. A combination of R[36] and Mathematica[37] code that delivers the results shown in Tables 3-10 and that draws Figures 7-11 is available on-line.[38]

8.2.1 No Included Exogenous Variables

The first model we employ has no exogenous variables in the threshold function simply specifying

$$Y_1 = \begin{cases} 0 & , & 0 \leq U \leq g(Y_2) \\ \\ 1 & , & g(Y_2) \leq U \leq 1 \end{cases}.$$

The excluded instrumental variables $Z_2 = (Z_{21}, Z_{22})$ we consider are also binary, and are as specified in AE98. Instrumental variable Z_{21} takes the value 1 if at the second birth event a woman had a multiple birth. We refer to this as the "twins instrument".[39] Instrumental variable Z_{22} takes the value 1 if the first two children born to the woman have the same sex and 0 otherwise. We refer to this as the "same-sex instrument".

The variables we use are precisely as in AE98, but the crucial difference in our analysis is that we employ a structural model which respects the discreteness of the outcome variable. By contrast in AE98 binary Y_1 is specified as a linear function of Y_2 and other variables and 2SLS estimates of local average treatment effects are calculated.

To connect with the notation used in Section 8.1.3, here we have

$$\begin{aligned} y_2^1 &= 0 & , & \gamma_1 &= g(0) \\ y_2^2 &= 1 & , & \gamma_2 &= g(1) \end{aligned}$$

with γ_1 and γ_2 unaffected by z. Let $f_{jk}(z)$ denote $P[Y_1 = j \wedge Y_2 = k | z]$.

Applying the results set out in Section 8.1.3 there are the following sharp bounds on the threshold parameters $g(0)$ and $g(1)$. The identified set comprises the union of two sets, denoted $\mathcal{A}_\uparrow$ and $\mathcal{A}_\downarrow$, in which the difference $\Delta \equiv g(0) - g(1)$ is respectively nonpositive and nonnegative.[40] These sets are defined as follows:

$$\begin{aligned} \mathcal{A}_\uparrow &\equiv \{(g(0), g(1)) : g(0) - g(1) \leq 0 \text{ and } (66) \text{ and } (67) \text{ hold}\}, \\ \mathcal{A}_\downarrow &\equiv \{(g(0), g(1)) : g(0) - g(1) \geq 0 \text{ and } (68) \text{ and } (69) \text{ hold}\}, \end{aligned}$$

---

[36] R Core Team (2014).

[37] Wolfram Research, Inc. (2019).

[38] https://drive.google.com/open?id=1wRpQWqIbVioEC4fej9vR69w2yByBd7ew.

[39] This is a slight abuse of language as there are a few instances of triplets.

[40] In keeping with earlier notation, the arrows in $\mathcal{A}_\uparrow$ and $\mathcal{A}_\downarrow$ convey whether γ_k is (weakly) increasing or decreasing in k. The parameter $\Delta \equiv g(0) - g(1)$, which is the average treatment effect, is nonpositive if γ_k is weakly increasing and nonnegative if γ_k is weakly decreasing.

where the inequalities in $\mathcal{A}_\uparrow$ are:

$$\sup_{z\in\mathcal{R}_Z} f_{00}(z) \;\leq\; g(0) \leq \inf_{z\in\mathcal{R}_Z} \left(f_{00}(z) + f_{01}(z)\right), \tag{66}$$

$$\sup_{z\in\mathcal{R}_Z} \left(f_{00}(z) + f_{01}(z)\right) \;\leq\; g(1) \leq \inf_{z\in\mathcal{R}_Z} \left(1 - f_{11}(z)\right), \tag{67}$$

and those in $\mathcal{A}_\downarrow$ are:

$$\sup_{z\in\mathcal{R}_Z} \left(f_{00}(z) + f_{01}(z)\right) \;\leq\; g(0) \leq \inf_{z\in\mathcal{R}_Z} \left(1 - f_{10}(z)\right), \tag{68}$$

$$\sup_{z\in\mathcal{R}_Z} f_{01}(z) \;\leq\; g(1) \leq \inf_{z\in\mathcal{R}_Z} \left(f_{00}(z) + f_{01}(z)\right). \tag{69}$$

The identified set for the pair of threshold values $(g(0), g(1))$ is the union of the sets $\mathcal{A}_\uparrow$ and $\mathcal{A}_\downarrow$. It is possible that (i) both of the sets are empty, (ii) only one of the sets is empty, in which case it is possible to sign the effect of Y_2 on Y_1, or (iii) both sets are non-empty. For example, considering the inequality (66), if $f_{01}(z)$ were close to zero and insensitive to variation in z while $f_{00}(z)$ did vary with z then the interval for $g(0)$ in $\mathcal{A}_\uparrow$ could be empty. Because there is the possibility that both sets are empty the model can potentially be falsified.

Each conditional probability $f_{ij}(z)$ appearing in (66)-(69) can be consistently estimated by its sample analog:

$$\hat{f}_{ij}(z) = \left(\frac{1}{n}\sum_{k=1}^{n} 1\left[z_k = z\right]\right)^{-1} \frac{1}{n}\sum_{k=1}^{n} 1\left[y_{k1} = i \wedge y_{k2} = j \wedge z_k = z\right]$$

where n is the sample size and the index k identifies observations in the sample. If Z were continuous or had many points of support then a kernel or sieve based estimator could be employed.

Using these sample conditional probabilities in place of population quantities produces analog set estimates for $\mathcal{A}_\uparrow$ and $\mathcal{A}_\downarrow$. These are illustrated in Figure 7 using each of the twins and same sex instruments by themselves, and also using them together. When using either the twins instrument on its own or the same sex and twins instruments together, the analog estimate for $\mathcal{A}_\downarrow$ is empty. Put another way, there are no $(g(0), g(1))$ pairs in the analog set estimate that correspond to a positive average treatment effect Δ. Such values of Δ are however admitted when the same sex instrument is used by itself, since estimates of both $\mathcal{A}_\uparrow$ and $\mathcal{A}_\downarrow$ are nonempty in this case.

Whatever instruments are used the estimated sets exclude values of $g(0)$ and $g(1)$ such that $\Delta = 0$. Manipulating (66) and (67) for the case $\Delta \leq 0$ delivers an upper bound on nonpositive Δ equal to $\inf_{z\in\mathcal{R}_Z} \left(f_{00}(z) + f_{01}(z)\right) - \sup_{z\in\mathcal{R}_Z} \left(f_{00}(z) + f_{01}(z)\right)$. This is the negative of the lower bound on nonnegative Δ obtained by manipulating (68) and (69). It follows that values of $g(0)$ and $g(1)$ with $g(0) = g(1)$, that is $\Delta = 0$, can only lie in the identified set if $\inf_{z\in\mathcal{R}_Z} \left(f_{00}(z) + f_{01}(z)\right) =$

$\sup_{z \in \mathcal{R}_Z} (f_{00}(z) + f_{01}(z))$, which can happen only if the conditional probability of working *does not vary* with the value of the instrument Z. This makes sense because under the restrictions of the GIV model the only route for Z to affect Y_1 is through its effect on Y_2 and this route is barred if $g(0) = g(1)$.

The sets shown in Figure 7 are analog estimates of identified sets obtained by replacing the probabilities that appear in the inequalities defining the identified sets by analog estimates of the probabilities. A complete analysis requires consideration of the impact of sampling variation in those estimated probabilities. Figure 8 depicts asymptotically valid 95% confidence sets for $(g(0), g(1))$ using inference methods developed in Chernozhukov, Lee, and Rosen (2013). Both sets $\mathcal{A}_\uparrow$ and $\mathcal{A}_\downarrow$ are defined by a collection of conditional moment inequalities, since each $f_{jk}(z)$ can be expressed as $E[1[Y = (j, k)] | Z = z]$. In the present application the excluded instruments Z_{21} (twins indicator) and Z_{22} (same sex indicator) are discrete, so these sets are represented as pairs $(g(0), g(1))$ that satisfy a finite number of conditional moment inequalities. For any given $(g(0), g(1))$ pair, either the inequalities that define $\mathcal{A}_\uparrow$ or the inequalities that define $\mathcal{A}_\downarrow$ are tested, where which one is chosen depends on the sign of the value of $g(0) - g(1)$ being tested.[41] The confidence set comprises the set of pairs such that the null hypothesis that $(g(0), g(1)) \in \mathcal{A}_\uparrow \cup \mathcal{A}_\downarrow$ is not rejected, where the null hypothesis is equivalent to satisfaction of the corresponding moment inequalities evaluated at $(g(0), g(1))$. The details of this moment inequalities test and those used in calculating all confidence sets reported here are set out in Section 8.3.

The shape of the $(g(0), g(1))$ regions and the main features in Figure 8 are similar to those of the analog set estimates reported in Figure 7. In each figure the light blue region is the confidence region obtained using the same sex instrument alone, the yellow and red region together comprise the confidence region obtained using the twins instrument alone, and the yellow region by itself is the confidence region obtained using both instruments.[42] When the same sex instrument is used the confidence set is very similar to the analog set estimate. This happens because the analog set estimate is quite accurate mainly due to the large number of observations at both values of the instrument. As shown in Figure 7 the estimated identified set obtained using the the twins instrument excludes all $(g(0), g(1))$ in $\mathcal{A}_\downarrow$. These values are also excluded from the 95% confidence region shown in Figure 8. The confidence sets using the twins instrument and both instruments together are considerably thicker than the analog set estimates which are manifolds (lines). This occurs because there are rather few twin births in the data (roughly 1 in 100 in the sample) resulting in a fair degree of sampling variation in estimated probabilities conditional on twin births which

---

[41] Note that when $g(0) = g(1) = \bar{g}$ so that $g(0) - g(1) = 0$, the inequalities defining $\mathcal{A}_\uparrow$ and $\mathcal{A}_\downarrow$ coincide, as they both reduce to $\sup_{z \in \mathcal{R}_Z} (f_{00}(z) + f_{01}(z)) \le \bar{g} \le \inf_{z \in \mathcal{R}_Z} (f_{00}(z) + f_{01}(z))$, which is equivalently that $\bar{g} = \mathbb{P}[Y_1 = 0|z]$ a.e. z. As pointed out previously, in this IV model with Z excluded from the structural function, there must then be no effect of z on $\mathbb{P}[Y_1 = 1|z]$ for the IV and exclusion restrictions to be valid.

[42] When both instruments are used the instrumental variables together have three points of support corresponding to: (i) women in the multiple birth category, (ii) women not in the multiple birth category with same-sex first two children, (iii) women not in the multiple birth category with mixed-sex first two children.

affects the accuracy of the set estimates.

Tables 3, 4 and 5 report several point estimates and confidence intervals obtained using respectively the twins instrument, the same sex instrument, and both instruments at once. Each table considers pairs of $(g(0), g(1))$ such that $g(0) \leq g(1)$ ($\mathcal{A}_\uparrow$) as well pairs such that $g(0) \geq g(1)$ ($\mathcal{A}_\downarrow$). Point estimates of the bounds on the thresholds $g(0)$ and $g(1)$ are reported using analog estimates for probabilities $\hat{f}_{ij}(z)$ for all values z of the instruments. Bold font denotes the greatest lower bound and least upper bound across the support of the instrument. For example, in Table 3, the highest estimate of the lower bounds and the lowest estimate of the upper bounds for $g(0)$ using the twins instrument are seen to be 0.2673 and 0.4713, respectively, in the region of the parameter space in which $g(0) \leq g(1)$. With the twins instrument the probability of more than two children conditional on the second birth being twins ($Z = 1$) is equal to one, from which it follows that $g(1)$ is point identified. This explains why in the table the lower and upper bound estimates on $g(1)$ obtained when conditioning on $Z = 1$ are both equal to 0.5236.

For each set $\mathcal{A}_\uparrow$ and $\mathcal{A}_\downarrow$, and each configuration of instruments (twins, same sex, or both) median bias-corrected interval estimates and 95% confidence intervals for each of $g(0)$, $g(1)$, and Δ are reported. The 95% confidence regions for each parameter are the unions of those obtained under the two cases $\mathcal{A}_\uparrow$ and $\mathcal{A}_\downarrow$, since these sets (and the inequalities that define them) cover disjoint regions of the parameter space. As was done for the joint confidence sets for $(g(0), g(1))$ depicted in Figure 8, these confidence intervals were computed using a procedure described in Chernozhukov, Lee, and Rosen (2013), set out in detail in Section 8.3.

The median-corrected bounds reported in these tables are obtained by computing what Chernozhukov, Lee, and Rosen (2013) refer to as half-median-unbiased estimates for each of the interval endpoints. Each corrected lower (upper) bound estimate is guaranteed to be less (greater) than or equal to the population lower bound with probability at least one half asymptotically. This correction is intended to counteract the now well-known fact that analog estimators for intersection bound endpoints are in general inward biased due to the application of the supremum and infimum operators to multiple bound estimators, as first pointed out by Manski and Pepper (2000, footnote 13, p.12). The median bias-corrected endpoint estimates are simply the endpoints of 0.50 confidence intervals as described in Chernozhukov, Lee, and Rosen (2013). These are easy to compute if one is already computing 0.90 or 0.95 confidence intervals, because they employ the same computational steps, requiring only a modification of the nominal level. Andrews and Shi (2013) also consider half-median-unbiased bound estimates, while Kreider and Pepper (2007) employ an alternative bootstrap bias correction.

The confidence intervals for each individual parameter $g(0)$, $g(1)$, and Δ are got by first manipulating the inequalities (66) and (67) for $\mathcal{A}_\uparrow$ and (68) and (69) for $\mathcal{A}_\downarrow$ into inequalities that

characterize the identified set for each parameter alone. For example, the inequalities

$$\sup_{z \in \mathcal{R}_Z} f_{00}(z) \quad \leq \quad g(0) \leq \inf_{z \in \mathcal{R}_Z} \left(f_{00}(z) + f_{01}(z) \right),$$
$$\sup_{z \in \mathcal{R}_Z} \left(f_{00}(z) + f_{01}(z) \right) \quad \leq \quad \inf_{z \in \mathcal{R}_Z} \left(1 - f_{11}(z) \right)$$

characterize the projection of $\mathcal{A}_\uparrow$ onto $g(0)$. One can test the hypothesis that any given $g(0)$ corresponds to the first component of some element of $\mathcal{A}_\uparrow$ by testing these inequalities. The 95% confidence sets for $g(0)$ reported here are those values not rejected by a size 0.05 test of the hypothesis that either these inequalities or the like inequalities for $\mathcal{A}_\downarrow$ hold. The same approach is taken in computing confidence sets for the parameters $g(1)$ and Δ. In each case confidence regions for each individual parameter across the identified set $\mathcal{A}_\uparrow \cup \mathcal{A}_\downarrow$ are the union of the intervals calculated for each region.[43] The confidence sets do not suffer from coverage inflation due to projection, because they are based on the inequalities that define the projections of the identified set for $(g(0), g(1))$. They are not projections of joint confidence sets for $(g(0), g(1))$.

When the twins instrument is used, by itself as in Table 3 or jointly with the same sex instrument as in Table 5, the sample analog estimates of the bounds on $g(1)$ restricted to the region $\mathcal{A}_\downarrow$ deliver an empty interval. This is seen in Table 3 for example by noting that for the $\mathcal{A}_\downarrow$ region, the highest estimated lower bound on $g(1)$ is 0.5236, which is larger than the lowest estimated upper bound, which is 0.4713. As noted earlier, there exist distributions of outcomes and instrumental variables such that either or both of these regions can be empty. In this application this signifies that there is no value of $(g(0), g(1))$ that satisfies the inequalities that define the set $\mathcal{A}_\downarrow$ using analog estimates of probabilities obtained with our sample. However it is possible that the empty estimated set arises in consequence of sampling variation. In order to investigate this the null hypothesis that all of the inequalities defining $\mathcal{A}_\downarrow$ in (68)-(69) are satisfied for some $(g(0), g(1))$ was tested at the 0.05 level using the parametric intersection bounds test in Chernozhukov, Lee, and Rosen (2013). The hypothesis is rejected when the twins instrument is used and when the twins and same sex instruments are used together. The result is that the bias-corrected interval estimates and confidence intervals for each parameter in the $\mathcal{A}_\downarrow$ region in Tables 3 and 5 are empty. The sample analog of the inequalities defining identified sets in all others cases are non-empty, and the inequalities that define them cannot be rejected at any commonly used significance level. The construction of the confidence intervals is set out in detail in Section 8.3.

---

[43]Note that the rationale for taking a union of confidence intervals across the $\mathcal{A}_\uparrow$ regions $\mathcal{A}_\downarrow$ is the construction of a confidence region by way of test inversion, and not a Bonferroni inequality argument. That is, to test any fixed parameter value under consideration, one needs to employ either the inequalities defining $\mathcal{A}_\uparrow$ or those defining $\mathcal{A}_\downarrow$, depending on the sign of $g(0) - g(1)$ at the parameter value being tested.

	$g(0)$		$g(1)$		Δ
$\mathcal{A}_\uparrow : g(0) \leq g(1)$	lower	upper	lower	upper	
$Z = 0$	**0.2673** (0.0009)	**0.4713** (0.0010)	0.4713 (0.0010)	0.8287 (0.0007)	–
$Z = 1$	0.0000 (0.0000)	0.5236 (0.0109)	**0.5236** (0.0109)	**0.5236** (0.0109)	–
Median-Corrected Bounds	$[0.2668, 0.4719]$		$[0.5163, 0.5309]$		$[-0.2568, -0.0448]$
95% CI	$[0.2656, 0.4733]$		$[0.5021, 0.5450]$		$[-0.2745, -0.0308]$
$\mathcal{A}_\downarrow : g(0) \geq g(1)$	lower	upper	lower	upper	
$Z = 0$	0.4713 (0.0010)	**0.6427** (0.0010)	0.2040 (0.0008)	**0.4713** (0.0010)	–
$Z = 1$	**0.5236** (0.0109)	1.000 (0.000)	**0.5236** (0.0109)	0.5236 (0.0109)	–
Median-Corrected Bounds	*empty*		*empty*		*empty*
95% CI	*empty*		*empty*		*empty*

Table 3: Estimated upper and lower bounds on threshold parameters $g(0)$ and $g(1)$ and Δ using only the twins instrument. The highest lower and lowest upper bound point estimates for $g(0)$ and $g(1)$ across values of the instrument are shown in bold, and standard errors are reported in parentheses. The existence of a $(g(0), g(1))$ pair satisfying $g(0) \geq g(1)$ is rejected at the 0.05 level.

	$g(0)$		$g(1)$		Δ
$\mathcal{A}_\uparrow : g(0) \leq g(1)$	lower	upper	lower	upper	
$Z = 0$	**0.2818** (0.0013)	**0.4665** (0.0014)	0.4665 (0.0014)	0.8437 (0.0010)	–
$Z = 1$	0.2532 (0.0012)	0.4761 (0.0014)	**0.4761** (0.0014)	**0.8140** (0.0011)	–
Median-Corrected Bounds	$[0.2818, 0.4673]$		$[0.4753, 0.8140]$		$[-0.5322, -0.0096]$
95% CI	$[0.2797, 0.4693]$		$[0.4734, 0.8158]$		$[-0.5350, -0.0064]$
$\mathcal{A}_\downarrow : g(0) \geq g(1)$	lower	upper	lower	upper	
$Z = 0$	0.4665 (0.0014)	**0.6228** (0.0014)	0.1847 (0.0011)	**0.4665** (0.0014)	–
$Z = 1$	**0.4761** (0.0014)	0.6621 (0.0013)	**0.2230** (0.0012)	0.4761 (0.0014)	–
Median-Corrected Bounds	$[0.4753, 0.6228]$		$[0.2229, 0.4676]$		$[0.0096, 0.3998]$
95% CI	$[0.4733, 0.6250]$		$[0.2210, 0.4699]$		$[0.0064, 0.4028]$

Table 4: Estimated upper and lower bounds on threshold parameters $g(0)$ and $g(1)$ and Δ using only the same sex instrument. The highest lower and lowest upper bound point estimates for $g(0)$ and $g(1)$ across values of the instrument are shown in bold, and standard errors are reported in parentheses. Observations in which there were twins at the second birth event were removed in these calculations, since in this case the sex of the first two children has no bearing on the decision to have a third child.

	$g(0)$		$g(1)$		Δ
$\mathcal{A}_\uparrow : g(0) \leq g(1)$	lower	upper	lower	upper	
$Z = 0$	**0.2818**	**0.4665**	0.4665	0.8437	–
	(0.0013)	(0.0014)	(0.0014)	(0.0010)	
$Z = 1$	0.2532	0.4761	0.4761	0.8140	–
	(0.0012)	(0.0014)	(0.0014)	(0.0011)	
$Z = 2$	0.0000	0.5236	**0.5236**	**0.5236**	–
	(0.0000)	(0.0109)	(0.0109)	(0.0109)	
Median-Corrected Bounds	$[0.2810, 0.4677]$		$[0.5162, 0.5309]$		$[-0.2429, -0.0470]$
95% CI	$[0.2792, 0.4695]$		$[0.5022, 0.5452]$		$[-0.2609, -0.0334]$
$\mathcal{A}_\downarrow : g(0) \geq g(1)$	lower	upper	lower	upper	
$Z = 0$	0.4665	**0.6228**	0.1847	**0.4665**	–
	(0.0014)	(0.0014)	(0.0011)	(0.0014)	
$Z = 1$	0.4761	0.6621	0.2230	0.4761	–
	(0.0014)	(0.0013)	(0.0012)	(0.014)	
$Z = 2$	**0.5236**	1.000	**0.5236**	0.5236	–
	(0.0109)	(0.0000)	(0.0109)	(0.0109)	
Median-Corrected Bounds	*empty*		*empty*		*empty*
95% CI	*empty*		*empty*		*empty*

Table 5: Estimated upper and lower bounds on threshold parameters $g(0)$ and $g(1)$ and Δ using the same sex instrument and the twins instrument simultaneously. The highest lower and lowest upper bound point estimates for $g(0)$ and $g(1)$ across values of the instrument are shown in bold, and standard errors are reported in parentheses. Here $Z = 0$ indicates the first two children were not the same sex and at the second birth event there were not twins, $Z = 1$ indicates that they were the same sex and at the second birth event there were not twins, and $Z = 2$ signifies that there were twins at the second birth event.

8.2.2 Included Exogenous Variables

Analysis similar to that above can be carried out when additional variables are included as exogenous explanatory variables. To demonstrate we include a dummy variable Z_1 equal to 1 if the mother has more than 12 years of education and equal to 0 otherwise. More generally, Z_1 could be a vector of included exogenous variables and some of these could be continuously distributed. The analysis conducted here exploits the discreteness of the included exogenous variable that we use. We consider the identifying power of both nonparametric models and parametric models that incorporate exogenous explanatory variables.

Nonparametric Specification. A nonparametric specification incorporating the included exogenous variables is

$$
Y_1 = \begin{cases} 0 & , & 0 \leq \quad U \quad \leq g(Y_2, Z_1) \\ \\ 1 & , & g(Y_2, Z_1) \leq \quad U \quad \leq 1 \end{cases}
$$

with Z_2 used to denote the excluded instruments, namely the same sex and twins instruments used separately or together. Under a nonparametric specification the dependence of the threshold on Z_1 is left unspecified with no restrictions placed on the relation between the threshold functions $g(\cdot, z_1)$ and $g(\cdot, z_1')$ for differing values of z_1 and z_1' on $\mathcal{R}_{Z_1}$. Estimation and inference is conducted by partitioning the data into distinct subsets according to the value of Z_1, and then applying the techniques used earlier.[44]

The resulting analog set estimates, median-corrected bound estimates, and 95% confidence intervals at each value of the exogenous variable and using each specification of excluded instruments are reported in Tables 6, 7 and 8. Using the same sex instrument the set estimates and confidence set for $(g(0, z_1), g(1, z_1))$ for both values of z_1 comprise the union of regions in which $g(0, z_1) \leq g(1, z_1)$ and $g(0, z_1) \geq g(1, z_1)$. When the twins instrument is used or when both instruments are used together, there are no values with $g(0, z_1) \geq g(1, z_1)$ in the confidence sets, and the hypothesis that this region is nonempty is rejected at the 0.05 level. Bound estimates and confidence intervals for $\Delta(z_1)$ correspondingly contain only negative values when the twins instrument is used with or without the same sex instrument, but using only the same sex instrument regions for $\Delta(z_1)$ contain both positive and negative values. Using the same sex instrument alone some of these parameter sets with $z_1 = 1$ are unions of overlapping intervals. These simplify to intervals but are expressed here as unions of the intervals on which $\Delta(z_1) = g(0, z_1) - g(1, z_1)$ has opposite signs for ease of comparison with other results.

The analog estimates in Table 6 illustrate precisely the bounds obtained by applying the same

---

[44]With continuous included exogenous variables one could either discretize them and proceed in a similar fashion, or, alternatively, estimate bounding probabilities conditional on the exogenous variables nonparametrically by series or kernel methods.

Instrument	$z_1 : \text{edu} > 12$	$g(0, z_1)$	$g(1, z_1)$	$\Delta(z_1)$
same sex	0	$\cup$ [0.2837, 0.4848] [0.4962, 0.6503]	$\cup$ [0.4962, 0.8052] [0.2428, 0.4848]	$\cup$ [−0.5215, −0.0115] [0.0115, 0.4075]
same sex	1	$\cup$ [0.2779, 0.4299] [0.4360, 0.5678]	$\cup$ [0.4360, 0.8314] [0.1835, 0.4299]	$\cup$ [−0.5535, −0.0061] [0.0061, 0.3843]
twins	0	[0.2684, 0.4906]	[0.5450, 0.5450]	[−0.2766, −0.0544]
twins	1	[0.2651, 0.4330]	[0.4847, 0.4847]	[−0.2196, −0.0517]
both	0	[0.2837, 0.4848]	[0.5450, 0.5450]	[−0.2613, −0.0602]
both	1	[0.2779, 0.4299]	[0.4847, 0.4847]	[−0.2068, −0.0548]

Table 6: Analog bound estimates using the same sex and twins instruments with a nonparametric specification for the inclusion of the exogenous variable Z_1 indicating mother's years of education > 12.

function that defines the unknown population bounds to the empirical distribution of the data. Because these are intersection bounds, in which interval endpoints are obtained as the minima and maxima of a collection of points, they are typically inward biased. The median-corrected estimates in Table 7 incorporate the same type of correction as described in the previous section for the model used here, and Table 8 again reports 95% confidence intervals using intersection bound inference as in Chernozhukov, Lee, and Rosen (2013), see Section 8.3 for further detail. In general the median-corrected interval estimates are wider than the analog interval estimates, and the 95% confidence sets are wider. However, in a handful of cases analog and median-corrected endpoint estimates coincide. This can happen when at the boundary of the analog set only one inequality survives the selection procedure described in the third step of the algorithm detailed in Section 8.3. In these situations, the tightest endpoint estimate is sufficiently far from the others relative to sampling variation to convey high probability that it is the uniquely binding inequality.

A benefit of including observed exogenous variables is that heterogenous effects of endogenous variables on outcomes can be measured. The confidence sets obtained with this nonparametric specification admit the possibility of homogeneous or heterogeneous average treatment effects across values of z_1, because the reported intervals overlap, but are not identical. More stringent parametric restrictions constrain the way in which Z_1 impacts the determination of Y_1. Such restrictions will result in (weakly) tighter identified sets for each $\Delta(z_1)$, and possibly different conclusions with regard to heterogenous effects.

A Parametric Specification. We now report results obtained with a parametric model including the exogenous explanatory variable Z_1 equal to 1 if the mother has more than 12 years of education and equal to 0 otherwise. The parametric model uses a threshold function such as is found in a probit model:

$$g(y_2, z) = \Phi(\beta_0 + \beta_1 z_1 + \alpha y_2) \tag{70}$$

Instrument	$z_1 : \mathrm{edu} > 12$	$g(0, z_1)$	$g(1, z_1)$	$\Delta(z_1)$
same sex	0	$\cup \begin{array}{c}[0.2837, 0.4857]\\ [0.4953, 0.6503]\end{array}$	$\cup \begin{array}{c}[0.4953, 0.8053]\\ [0.2428, 0.4861]\end{array}$	$\cup \begin{array}{c}[-0.5216, -0.0108]\\ [0.0115, 0.4075]\end{array}$
same sex	1	$\cup \begin{array}{c}[0.2779, 0.4312]\\ [0.4347, 0.5678]\end{array}$	$\cup \begin{array}{c}[0.4347, 0.8314]\\ [0.1835, 0.4319]\end{array}$	$\cup \begin{array}{c}[-0.5535, -0.0038]\\ [0.0038, 0.3843]\end{array}$
twins	0	$[0.2678, 0.4913]$	$[0.5358, 0.5542]$	$[-0.2774, -0.0451]$
twins	1	$[0.2643, 0.4340]$	$[0.4686, 0.4971]$	$[-0.2202, -0.0393]$
both	0	$[0.2828, 0.4862]$	$[0.5331, 0.5541]$	$[-0.2628, -0.0476]$
both	1	$[0.2767, 0.4320]$	$[0.4659, 0.4970]$	$[-0.2199, -0.0347]$

Table 7: Bound estimates with median-corrected interval endpoints, equivalently 50% confidence sets, using the same sex and twins instruments with a nonparametric specification for the inclusion of the exogenous variable Z_1 indicating mother's years of education > 12.

Instrument	$z_1 : \mathrm{edu} > 12$	$g(0, z_1)$	$g(1, z_1)$	$\Delta(z_1)$
same sex	0	$\cup \begin{array}{c}[0.2811, 0.4882]\\ [0.4929, 0.6530]\end{array}$	$\cup \begin{array}{c}[0.4929, 0.8075]\\ [0.2404, 0.4891]\end{array}$	$\cup \begin{array}{c}[-0.5250, -0.0075]\\ [0.0067, 0.4112]\end{array}$
same sex	1	$\cup \begin{array}{c}[0.2743, 0.4347]\\ [0.4313, 0.5718]\end{array}$	$\cup \begin{array}{c}[0.4314, 0.8344]\\ [0.1804, 0.4361]\end{array}$	$\cup \begin{array}{c}[-0.5582, 0.0000]\\ [0.0000, 0.3894]\end{array}$
twins	0	$[0.2663, 0.4930]$	$[0.5186, 0.5719]$	$[-0.2992, -0.0274]$
twins	1	$[0.2621, 0.4364]$	$[0.4456, 0.5203]$	$[-0.2504, -0.0154]$
both	0	$[0.2806, 0.4885]$	$[0.5164, 0.5716]$	$[-0.2851, -0.0308]$
both	1	$[0.2735, 0.4352]$	$[0.4439, 0.5208]$	$[-0.2393, -0.0130]$

Table 8: 95% confidence sets using the same sex and twins instruments with a nonparametric specification for the inclusion of the exogenous variable Z_1 indicating mother's years of education > 12.

where Φ denotes the standard normal distribution function.[45]

As in the earlier analyses, the identified set comprises the union of two regions $\mathcal{A}_\uparrow$ and $\mathcal{A}_\downarrow$, which now define sets of values of parameters $\theta \equiv (\alpha, \beta_0, \beta_1)$, as follows:

$$
\begin{aligned}
\mathcal{A}_\uparrow &\equiv \{(\alpha, \beta_0, \beta_1) : \alpha \geq 0 \text{ and } (71) \text{ and } (72) \text{ hold a.e. } z \in \mathcal{R}_Z\}, \\
\mathcal{A}_\downarrow &\equiv \{(\alpha, \beta_0, \beta_1) : \alpha \leq 0 \text{ and } (73) \text{ and } (74) \text{ hold a.e. } z \in \mathcal{R}_Z\},
\end{aligned}
$$

where the inequalities defining $\mathcal{A}_\uparrow$ are

$$
f_{00}(z) \leq \Phi(\beta_0 + \beta_1 z_1) \leq f_{00}(z) + f_{01}(z), \tag{71}
$$

$$
f_{00}(z) + f_{01}(z) \leq \Phi(\beta_0 + \beta_1 z_1 + \alpha) \leq 1 - f_{11}(z), \tag{72}
$$

and those defining $\mathcal{A}_\downarrow$ are

$$
f_{00}(z) + f_{01}(z) \leq \Phi(\beta_0 + \beta_1 z_1) \leq 1 - f_{10}(z), \tag{73}
$$

$$
f_{01}(z) \leq \Phi(\beta_0 + \beta_1 z_1 + \alpha) \leq f_{00}(z) + f_{01}(z). \tag{74}
$$

These are the same inequalities as (66)-(69), but for the replacement of the thresholds with the parametric specification (70).

Estimated sets and confidence regions for the full parameter vector $\theta \equiv (\alpha, \beta_0, \beta_1)$ can be obtained using the methods set out in Section 8.2.1 where estimated sets for $(g(0), g(1))$ are reported. Figures 9, 10 and 11 show the projections of the analog set estimate of θ onto the space of each pair of its component parameters. The estimated sets for θ obtained using only the same sex instrument are disconnected, comprising two convex polyhedra whose 2D projections are convex polygons. As in the case with no included exogenous variable, the identified sets contain no values of θ in which $\alpha = 0$. The estimated identified sets obtained using the twins instrument alone or in combination with the same sex instrument contain only positive values of α. These sets are 2D manifolds whose projections onto the space of any two of the components of θ are lines. This reduction in dimension occurs because using the twins instrument the coefficient on the included exogenous variable, β_1, is point identified.[46] The estimated set obtained using both instruments is slightly smaller than the set obtained using the twins instrument alone as the projections demonstrate.

In applications in which parameter vectors have many components it is usually desirable to focus attention on one or a few parameter components or interesting functions of parameters. Calculating confidence regions in this situation is often not straightforward when parameters are

---

[45] The analysis can be just as easily carried out using any strictly increasing CDF in place of Φ, such as the logit CDF. Indeed, logit estimates not reported here were also computed. These were unsurprisingly found to deliver results similar to those using the normal CDF.

[46] To see why consider studying the subsample in which the exogenous twins instrument is equal to 1. In this subsample $Y_2 = 1$ and using this subsample a consistent estimate of β_1 and $\beta_0 + \alpha$ can be obtained.

partially identified because of the difficulty of drawing inferences on subvectors of partially identified parameter vectors. This is an active area of research, with recent important contributions that include those of Bugni, Canay, and Shi (2017), Chen, Christensen, and Tamer (2018), Kaido, Molinari, and Stoye (forthcoming), and Belloni, Bugni, and Chernozhukov (2018).

Here we are able to avoid these difficulties and give confidence intervals for individual parameter components. To do this we exploit the structure of the inequalities (71) – (74) to obtain characterizations of identified sets for individual parameters which comprise systems of moment inequalities that can be tested directly using the methods employed to produce the results reported earlier.

Start by applying the standard normal quantile function transformation to (71) – (74) giving the inequalities

$$\Phi^{-1}\left(f_{00}\left(z\right)\right) \leq \beta_0 + \beta_1 z_1 \leq \Phi^{-1}\left(f_{00}\left(z\right) + f_{01}\left(z\right)\right), \tag{75}$$

$$\Phi^{-1}\left(f_{00}\left(z\right) + f_{01}\left(z\right)\right) \leq \beta_0 + \beta_1 z_1 + \alpha \leq \Phi^{-1}\left(1 - f_{11}\left(z\right)\right), \tag{76}$$

for the set $\mathcal{A}_\uparrow$ and

$$\Phi^{-1}\left(f_{00}\left(z\right) + f_{01}\left(z\right)\right) \leq \beta_0 + \beta_1 z_1 \leq \Phi^{-1}\left(1 - f_{10}\left(z\right)\right), \tag{77}$$

$$\Phi^{-1}\left(f_{01}\left(z\right)\right) \leq \beta_0 + \beta_1 z_1 + \alpha \leq \Phi^{-1}\left(f_{00}\left(z\right) + f_{01}\left(z\right)\right). \tag{78}$$

for the set $\mathcal{A}_\downarrow$. The identified set for θ is given by $\mathcal{A}_\uparrow \cup \mathcal{A}_\downarrow$ where

$$\mathcal{A}_\uparrow \equiv \left\{(\alpha, \beta_0, \beta_1) : \alpha \geq 0 \text{ and (75) and (76) hold a.e. } z \in \mathcal{R}_Z\right\}, \tag{79}$$

$$\mathcal{A}_\downarrow \equiv \left\{(\alpha, \beta_0, \beta_1) : \alpha \leq 0 \text{ and (77) and (78) hold a.e. } z \in \mathcal{R}_Z\right\}. \tag{80}$$

Each constituent set is an intersection of linear half-spaces. The sets $\mathcal{A}_\uparrow$ and $\mathcal{A}_\downarrow$ are therefore both convex so the projections of each of the sets $\mathcal{A}_\uparrow$ and $\mathcal{A}_\downarrow$ onto the space of individual parameter components are intervals or empty.

Because the constituent inequalities of each set taken over all $z \in \mathcal{R}_Z$ are linear in parameters, Fourier-Motzkin elimination (FME) can be applied to obtain the inequalities comprising the identified set for each individual component.[47] To express these projections succinctly, define the functions

$$b_u^+\left(0, z_1, z_2\right) \equiv \Phi^{-1}\left(f_{00}\left(z_1, z_2\right) + f_{01}\left(z_1, z_2\right)\right), \quad b_u^+\left(1, z_1, z_2\right) \equiv \Phi^{-1}\left(1 - f_{11}\left(z_1, z_2\right)\right),$$

$$b_l^+\left(0, z_1, z_2\right) \equiv \Phi^{-1}\left(f_{00}\left(z_1, z_2\right)\right), \quad b_l^+\left(1, z_1, z_2\right) \equiv \Phi^{-1}\left(f_{00}\left(z_1, z_2\right) + f_{01}\left(z_1, z_2\right)\right),$$

---

[47]FME eliminates each parameter in turn until a system of inequalities defining a projection onto the space of one or more parameters of interest is obtained. In the step at which a parameter θ_i is eliminated the inequalities not involving θ_i are passed through to the next step, upper and lower bounds on θ_i are identified, and the inequalities passed on to the next step are those obtained by requiring that each upper bound on θ_i is at least equal to each lower bound on θ_i.

and

$$b_u^- (0, z_1, z_2) \equiv \Phi^{-1} (1 - f_{10} (z_1, z_2)), \quad b_u^- (1, z_1, z_2) \equiv \Phi^{-1} (f_{00} (z_1, z_2) + f_{01} (z_1, z_2)),$$
$$b_l^- (0, z_1, z_2) \equiv \Phi^{-1} (f_{00} (z_1, z_2) + f_{01} (z_1, z_2)), \quad b_l^- (1, z_1, z_2) \equiv \Phi^{-1} (f_{01} (z_1, z_2)),$$

where with slight abuse of notation $f_{jk} (z_1, z_2)$ is written in place of $f_{jk} ((z_1, z_2))$.

The characterizations of projections of the sets $\mathcal{A}_\uparrow$ and $\mathcal{A}_\downarrow$ onto the space of individual components of θ delivered by FME are the same for each of these sets except that functions b_l^+ and b_u^+ define $\mathcal{A}_\uparrow$ while functions b_l^- and b_u^- define $\mathcal{A}_\downarrow$. The projections are now expressed in terms of inequalities in which the "+" and "−" superscripts on the b_u and b_l functions are omitted. Appropriate superscripts "+" and "−" must be applied when definitions of respectively $\mathcal{A}_\uparrow$ and $\mathcal{A}_\downarrow$ are required.

Isolating α by FME produces the following inequalities required to hold for each pair $z_2, z_2' \in \mathcal{R}_{Z_2}$:

$$\forall y_2 \in \{0, 1\}, \quad z_1 \in \{0, 1\} : \quad b_u (y_2, z_1, z_2) - b_l (y_2, z_1, z_2') \geq 0, \tag{81}$$

and

$$\forall z_1 \in \{0, 1\}, \quad \begin{array}{l} \alpha - b_l (1, z_1, z_2) + b_u (0, z_1, z_2') \geq 0, \\ b_u (1, z_1, z_2) - b_l (0, z_1, z_2') - \alpha \geq 0. \end{array} \tag{82}$$

The inequalities (81) do not involve any parameters, but they are inequalities that must hold in order for the set $\mathcal{A}_\uparrow$ or $\mathcal{A}_\downarrow$ (depending on whether b_l^+ and b_u^+ or b_l^- and b_u^- are used) to be nonempty. Recall the identified set for the parameters is the union of $\mathcal{A}_\downarrow$ and $\mathcal{A}_\uparrow$, one or both of which may be empty. The projection of the identified set onto a parameter axis is the union of the projections of $\mathcal{A}_\downarrow$ and $\mathcal{A}_\uparrow$.

For pairs $z_2 = z_2'$, the inequalities (81) hold automatically, and can be dropped. Thus, when Z_2 has K points of support there are $4K(K-1)$ inequalities of this form. The inequalities (81) appear in all characterizations of projections of the sets $\mathcal{A}_\uparrow$ and $\mathcal{A}_\downarrow$. When the twins instrument is used, or when both the twins and same sex instrument are used together, these inequalities for the set $\mathcal{A}_\downarrow$ are rejected at the 0.0001 level. With the same sex instrument, inequalities (81) for $\mathcal{A}_\downarrow$ are not rejected at any conventional level, and they are never rejected for the set $\mathcal{A}_\uparrow$ for any instrument specification. This is easily seen by noting that both sample analog sets are nonempty.

To construct the confidence sets for α that are reported here, the inequalities (81) and (82) are jointly tested using the parametric inference procedure from Chernozhukov, Lee, and Rosen (2013), described in Section 8.3.

Application of FME yields similar characterizations for parameters β_0 and β_1. The sets for β_0

are characterized by the inequalities (81) for all $z_2 \neq z_2'$ and additionally

$$\forall z_2 \in \mathcal{R}_z : \quad \begin{array}{l} \beta_0 - b_l\left(0, 0, z_2\right) \geq 0, \\ b_u\left(0, 0, z_2\right) - \beta_0 \geq 0, \end{array}$$

and

$$\forall z_2, z_2', \tilde{z}_2 \in \mathcal{R}_z : \quad \begin{array}{l} \beta_0 - b_l\left(0, 1, z_2\right) - b_l\left(1, 0, z_2'\right) + b_u\left(1, 1, \tilde{z}_2\right) \geq 0, \\ b_u\left(0, 1, z_2\right) + b_u\left(1, 0, z_2'\right) - b_l\left(1, 1, \tilde{z}_2\right) - \beta_0 \geq 0. \end{array}$$

The inequalities defining the projection of the identified set for θ onto the space of β_1 comprise (81) along with

$$\forall z_2, z_2' \in \mathcal{R}_z : \quad \begin{array}{l} \beta_1 - b_l\left(0, 1, z_2\right) + b_u\left(0, 0, z_2'\right) \geq 0, \\ \beta_1 - b_l\left(1, 1, z_2\right) + b_u\left(1, 0, z_2'\right) \geq 0, \\ b_u\left(0, 1, z_2\right) - b_l\left(0, 0, z_2'\right) - \beta_1 \geq 0, \\ b_u\left(1, 1, z_2\right) - b_l\left(1, 0, z_2'\right) - \beta_1 \geq 0. \end{array}$$

Altogether, there are $4K\left(K-1\right) + 2K + 2K^3$ inequalities for β_0 and $4K\left(K-1\right) + 4K^2$ inequalities for β_1 for each of $\mathcal{A}_\uparrow$ and $\mathcal{A}_\downarrow$.

Table 9 contains analog estimates, median-corrected estimates, and 95% confidence intervals for projections of the identified set onto the space of individual parameters obtained using the twins instrument, and using the same sex and the twins instrument together. Table 10 presents the results obtained when only the same sex instrument is used. Using the twins instrument the inequalities that characterize $\mathcal{A}_\downarrow$ ($\alpha \leq 0$) are rejected at conventional levels. This is not the case when the same sex instrument is used alone, and the two columns of results reported in Table 10 show contributions to confidence regions and set estimates delivered by the inequalities for $\mathcal{A}_\uparrow$ and $\mathcal{A}_\downarrow$. The estimated sets and confidence intervals for each parameter are given by the union of the intervals shown in the two columns.

In both the cases in which the twins instrument is used the set estimates and confidence intervals for parameters α, β_0 and β_1 are much more informative than when the same sex instrument is used alone. The use of the twins and same sex instrument together narrows the bound estimates slightly relative to the use of the twins instrument alone. When the twins instrument is used the confidence intervals for α lie entirely above zero, and those for β_0 and β_1 lie below zero.

Also reported are interval estimates and confidence intervals for the average treatment effects $\Delta\left(z_1\right)$ with z_1 fixed at each of its values. In this parametric model, these are

$$\Delta\left(0\right) \equiv \Phi\left(\beta_0\right) - \Phi\left(\beta_0 + \alpha\right), \quad \Delta\left(1\right) \equiv \Phi\left(\beta_0 + \beta_1\right) - \Phi\left(\beta_0 + \beta_1 + \alpha\right).$$

The identified set for each $\Delta\left(z_1\right)$, $z_1 \in \{0, 1\}$, is given by the set of values of $\Phi\left(\beta_0 + z_1\beta_1\right) - \Phi\left(\beta_0 + z_1\beta_1 + \alpha\right)$ such that either (i) the $\mathcal{A}_\uparrow$ inequalities (71)-(72) hold or (ii) the $\mathcal{A}_\downarrow$ inequalities (73)-(74) hold. Analog estimates of these sets are computed by minimizing and maximizing the

expression for $\Delta(z_1)$ by choice of θ subject to the $\mathcal{A}_\uparrow$ and $\mathcal{A}_\downarrow$ inequalities in turn with the probabilities in these inequalities replaced by estimates. For example, to compute the analog estimate of the lower bound on $\Delta(z_1)$ on $\mathcal{A}_\uparrow$ for either $z_1 \in \{0, 1\}$, the expression

$$\min_{(\alpha, \beta_0, \beta_1) \in \widehat{\mathcal{A}_\uparrow}} \Phi(\beta_0 + \beta_1 z_1) - \Phi(\beta_0 + \beta_1 z_1 + \alpha)$$

is computed, where $\widehat{\mathcal{A}_\uparrow}$ is the set (79), except with analog estimates $\hat{f}_{ij}(z)$ in place of population probabilities $f_{ij}(z)$. The analog estimate of the upper bound is obtained similarly, except the maximum of this expression over the region $\widehat{\mathcal{A}_\uparrow}$ is taken instead of the minimum. Bound estimates for the $\mathcal{A}_\downarrow$ are obtained in the same way, replacing $\widehat{\mathcal{A}_\uparrow}$ with $\widehat{\mathcal{A}_\downarrow}$, the analog estimate of $\mathcal{A}_\downarrow$. These estimated sharp bounds are reported as the $\Delta(0)$ and $\Delta(1)$ estimates in Tables 9 and 10.

The inequalities in (79) and (80) that provide the identified set for θ are linear in θ but $\Delta(0)$ and $\Delta(1)$ are non-linear functions of θ. This precludes characterization of the identified set for each of $\Delta(0)$ and $\Delta(1)$ by way of inequalities that are linear in these parameters, and consequently FME cannot be used to obtain sharp bounds on each of these quantities individually. However, on differencing the upper and lower bounds on each of $\Phi(\beta_0 + z_1\beta_1)$ and $\Phi(\beta_0 + z_1\beta_1 + \alpha)$ individually in (71)-(74), valid non-sharp bounds on each of $\Delta(0)$ and $\Delta(1)$ are obtained. For instance, a valid lower bound on $\Delta(0) = \Phi(\beta_0) - \Phi(\beta_0 + \alpha)$ in the $\mathcal{A}_\uparrow$ region is obtained by plugging in the lower bound on $\Phi(\beta_0)$ in (71) and the upper bound on $\Phi(\beta_0 + \alpha)$ in (72), both with $z_1 = 0$.

Analog estimates of these outer bounds for each of $\mathcal{A}_\uparrow$ and $\mathcal{A}_\downarrow$ and for $z_1 \in \{0, 1\}$ are reported in Tables 9 and 10 in the rows labelled "$\Delta(z_1)$ differencing". The difference between these outer bounds and the sharp bounds are small. The inequalities that define these outer bounds were used for construction of the median-corrected bound estimates and 95% confidence sets reported in these tables.

8.3 Inference

The moment inequalities that define the identified sets described in this section all depend on smooth functions of population moments that are consistently estimable by standard methods. The asymptotic variance of estimators of these moments can be obtained by way of the delta method, and inference can be carried out following recently developed approaches for conducting inference on parameters that are restricted by moment inequalities. Here we describe how the confidence sets reported here were obtained using a procedure introduced in Chernozhukov, Lee, and Rosen (2013). Interval estimates with median-corrected endpoints, i.e. half-median-unbiased estimates, were computed by setting $\alpha = 0.5$.

The inference method considers a vector of moments

$$m(W, \theta) \equiv (m_1(W, \theta), ..., m_J(W, \theta))',$$

89

Table 9: Estimates and confidence sets for each parameter and conditional ATEs at each value of z_1 using the twins instrument and simultaneously the twins and same sex instruments. The included exogenous variable is binary: education of the mother exceeds 12 years. Corrected estimates are interval estimates employing median-corrected estimators of interval endpoints, equivalently 50% confidence intervals, as described in the text.

	Twins only	Twins and same sex
α analog estimate	$[0.1367, 0.5893]$	$[0.1512, 0.5506]$
α corrected estimate	$[0.1006, 0.6165]$	$[0.1041, 0.5832]$
α 95% CI	$[0.0597, 0.6821]$	$[0.0661, 0.6465]$
β_0 analog estimate	$[-0.4763, -0.0237]$	$[-0.4376, -0.0382]$
β_0 corrected estimate	$[-0.5153, -0.0208]$	$[-0.4795, -0.0332]$
β_0 95% CI	$[-0.5661, -0.0168]$	$[-0.5300, -0.0278]$
β_1 analog estimate	-0.1514	-0.1514
β_1 corrected estimate	$[-0.2041, -0.1281]$	$[-0.2067, -0.1269]$
β_1 95% CI	$[-0.2781, -0.0952]$	$[-0.2800, -0.0940]$
$\Delta(0)$ analog estimate	$[-0.2281, -0.0544]$	$[-0.2613, -0.0602]$
$\Delta(0)$ differencing	$[-0.2766, -0.0544]$	$[-0.2613, -0.0602]$
$\Delta(0)$ corrected estimate	$[-0.2841, -0.0425]$	$[-0.2702, -0.0456]$
$\Delta(0)$ 95% CI	$[-0.3033, -0.0257]$	$[-0.2892, -0.0291]$
$\Delta(1)$ analog estimate	$[-0.2196, -0.0542]$	$[-0.2068, -0.0599]$
$\Delta(1)$ differencing	$[-0.2196, -0.0517]$	$[-0.2068, -0.0548]$
$\Delta(1)$ corrected estimate	$[-0.2299, -0.0355]$	$[-0.2241, -0.0326]$
$\Delta(1)$ 95% CI	$[-0.2557, -0.0126]$	$[-0.2444, -0.0115]$

Table 10: Estimates and confidence sets for each parameter and conditional ATEs at each value off z_1 using the same sex instrument. The included exogenous variable is binary: education of the mother exceeds 12 years. Corrected estimates are interval estimates employing median-corrected estimators of interval endpoints, equivalently 50% confidence intervals, as described in the text.

	$\alpha \geq 0$	$\alpha \leq 0$
α analog estimate	$[0.0287, 1.432]$	$[-1.073, -0.0287]$
α corrected estimate	$[0.0223, 1.432]$	$[-1.078, -0.0223]$
α 95% CI	$[0.0150, 1.444]$	$[-1.091, -0.0150]$
β_0 analog estimate	$[-0.5719, -0.0382]$	$[-0.0095, 0.3862]$
β_0 corrected estimate	$[-0.5719, -0.0358]$	$[-0.0119, 0.3862]$
β_0 95% CI	$[-0.5795, -0.0297]$	$[-0.0179, 0.3935]$
β_1 analog estimate	$[-0.5508, 0.3953]$	$[-0.5472, 0.1801]$
β_1 corrected estimate	$[-0.5561, 0.3967]$	$[-0.5506, 0.1848]$
β_1 95% CI	$[-0.5672, 0.4016]$	$[-0.5619, 0.1937]$
$\Delta(0)$ analog estimate	$[-0.5215, -0.0115]$	$[0.0115, 0.4042]$
$\Delta(0)$ differencing	$[-0.5215, -0.0115]$	$[0.0115, 0.4075]$
$\Delta(0)$ corrected estimate	$[-0.5216, -0.0098]$	$[0.0098, 0.4075]$
$\Delta(0)$ 95% CI	$[-0.5250, -0.0066]$	$[0.0067, 0.4112]$
$\Delta(1)$ analog estimate	$[-0.5226, -0.0113]$	$[0.0113, 0.3843]$
$\Delta(1)$ differencing	$[-0.5535, -0.0061]$	$[0.0061, 0.3843]$
$\Delta(1)$ corrected estimate	$[-0.5535, -0.0038]$	$[0.0038, 0.3843]$
$\Delta(1)$ 95% CI	$[-0.5582, 0.0000]$	$[0.0000, 0.3894]$

where the model implies that at any of the value of θ that could have generated the distribution of observable variables W, the moment inequalities $m(W, \theta) \geq 0$ hold.[48] An asymptotic $1 - \alpha$ confidence set for θ is constructed by the set of θ that are not rejected by an α-level test of the null hypothesis $H_0 : m(W, \theta) \geq 0$ against the alternative that for some j, $m_j(W, \theta) < 0$. The null and alternative hypotheses are equivalently formulated as

$$H_0 : \min_{j=1,\ldots,J} m_j(W, \theta) \geq 0, \quad H_1 : \min_{j=1,\ldots,J} m_j(W, \theta) < 0. \tag{83}$$

Formulation (83) motivates the intersection bound test of Chernozhukov, Lee, and Rosen (2013). The analog estimator for $\min_{j=1,\ldots,J} m_j(W, \theta)$, namely $\min_{j=1,\ldots,J} \hat{m}_j(W, \theta)$ will exhibit downward bias due to application of the minimum. To account for this, the idea behind Chernozhukov, Lee, and Rosen (2013) is to adjust each estimated moment $\hat{m}_j(W, \theta)$ upward by its standard error s_j times a critical value $k_{1-\alpha}$ appropriately calibrated in order to achieve correct asymptotic size. The null hypothesis in (83) is then rejected if and only if the minimum of these *precision-corrected* moments $\hat{m}_j(W, \theta) + k_{1-\alpha} s_j$ is less than zero.

Properly choosing the critical value $k_{1-\alpha}$ for the intersection bound test involves two steps. The first step is an adaptive inequality selection procedure that determines which components of $\hat{m}_j(W, \theta)$ are suitably close to or below zero. Let $\widehat{\mathcal{J}} \subseteq \{1, \ldots, J\}$ denote the indices of the selected moments, which we refer to as the contact set estimator. The contact set estimator is constructed in such a way as to guarantee that when H_0 holds all moments $m_j(W, \theta)$ equal to zero are selected with probability tending to one asymptotically. Intuitively, moments that exceed zero by a wide margin can be safely ignored. The second step is to set $k_{1-\alpha}$ to approximately the $1 - \alpha$ quantile of the maximum of the studentized version of the selected moments, i.e. the maximum of

$$\frac{\hat{m}_j(W, \theta) - m_j(W, \theta)}{s_j}$$

across the components of j selected in step 1. These studentized moments are asymptotically multivariate normally distributed with variance equal to the correlation matrix of the selected components of $\hat{m}(W, \theta)$. Thus $k_{1-\alpha}$ is set to the $1 - \alpha$ quantile of the maximum of such a multivariate normal random vector, computed by simulation. The precision-corrected estimator for each moment is then set to $\hat{m}_j(W, \theta) + k_{1-\alpha} s_j$, and the minimum of these is computed across all $j = 1, \ldots, J$ and compared to zero.

To describe implementation of such a test formally, consider inference on the threshold parameters $(g(0), g(1))$ in the model described in Section 8.2.1 in which there are no included exogenous variables.

With instruments Z having finite support $\mathcal{R}_Z = \{z_1, \ldots, z_L\}$, define for each $\ell = 1, \ldots, L$ the 6

---

[48] Vector $a \geq 0$ only if each element of a is nonnegative.

element vector

$$\zeta\left(z_\ell\right) \equiv \left(E\left[1_{z_\ell}\right], E\left[\mathcal{E}_{00}1_{z_\ell}\right], E\left[\mathcal{E}_01_{z_\ell}\right], E\left[!\mathcal{E}_{11}1_{z_\ell}\right], E\left[!\mathcal{E}_{10}1_{z_\ell}\right], E\left[\mathcal{E}_{01}1_{z_\ell}\right]\right),$$

where

$$1_{z_\ell} \equiv 1\left[Z = z_\ell\right], \quad \mathcal{E}_{jk} \equiv 1\left[Y_1 = j \wedge Y_2 = k\right], \quad \mathcal{E}_j \equiv 1\left[Y_1 = j\right], \quad !\mathcal{E}_{jk} \equiv 1 - \mathcal{E}_{jk}.$$

We refer to the elements of the vector

$$\zeta \equiv \left(\zeta\left(z_1\right), ..., \zeta\left(z_L\right)\right)$$

as inference parameters to distinguish them from structural parameters (such as $g(0)$ and $g(1)$).

Since each of the components of $\zeta\left(z_\ell\right)$ is the mean of a function of observable variables, the parameter vector ζ can be consistently estimated by $\hat{\zeta}$ obtained by replacing expectations with sample means. Let Ω denote the asymptotic variance of $\sqrt{n}\left(\hat{\zeta} - \zeta\right)$ and let $\hat{\Omega}$ be an asymptotic variance estimator constructed so that $\left|\hat{\Omega} - \Omega\right| \xrightarrow{p} 0$.

The inequalities (66) and (67) that define the set $\mathcal{A}_\uparrow$ can be written as functions of the inference parameters:

$$
\begin{align}
m_1^\uparrow\left(g\left(0\right), g\left(1\right), \zeta\left(z_\ell\right)\right) &\equiv \zeta_3\left(z_\ell\right)/\zeta_1\left(z_\ell\right) - g(0) \geq 0, \tag{84} \\
m_2^\uparrow\left(g\left(0\right), g\left(1\right), \zeta\left(z_\ell\right)\right) &\equiv g(0) - \zeta_2\left(z_\ell\right)/\zeta_1\left(z_\ell\right) \geq 0, \tag{85} \\
m_3^\uparrow\left(g\left(0\right), g\left(1\right), \zeta\left(z_\ell\right)\right) &\equiv \zeta_4\left(z_\ell\right)/\zeta_1\left(z_\ell\right) - g\left(1\right) \geq 0, \tag{86} \\
m_4^\uparrow\left(g\left(0\right), g\left(1\right), \zeta\left(z_\ell\right)\right) &\equiv g\left(1\right) - \zeta_3\left(z_\ell\right)/\zeta_1\left(z_\ell\right) \geq 0. \tag{87}
\end{align}
$$

Let

$$m\left(g\left(0\right), g\left(1\right), \zeta\right) \equiv \left(m_1^\uparrow\left(g\left(0\right), g\left(1\right), \zeta\left(z_\ell\right)\right), ..., m_4^\uparrow\left(g\left(0\right), g\left(1\right), \zeta\left(z_\ell\right)\right) : \ell = 1, ..., L\right)'$$

denote the $4L$ vector of moment functions that define the set $\mathcal{A}_\uparrow$ of $\left(g\left(0\right), g\left(1\right)\right)$ pairs with $g\left(0\right) \leq g\left(1\right)$. This is a vector of known functions of the inference parameters ζ, consistently estimated by $m\left(g\left(0\right), g\left(1\right), \hat{\zeta}\right)$ since $\hat{\zeta} \xrightarrow{p} \zeta$ and the moment functions are continuous in ζ.

Let p denote the $4L \times 6L$ matrix of partial derivatives of $m\left(g\left(0\right), g\left(1\right), \zeta\right)$ with respect to ζ. Application of the delta method gives

$$\sqrt{n}\left(m\left(g\left(0\right), g\left(1\right), \hat{\zeta}\right) - m\left(g\left(0\right), g\left(1\right), \zeta\right)\right) \xrightarrow{d} N\left(0, V\right), \quad V \equiv p\Omega p',$$

with the asymptotic variance of studentized moments V consistently estimated by $\hat{V} \equiv \hat{p}\hat{\Omega}\hat{p}'$, where

$\hat{p}$ is the consistent estimator for p in which the vector of inference parameters ζ is replaced by $\hat{\zeta}$.

To test the hypothesis that $(g(0), g(1)) \in \mathcal{A}_\uparrow$, equivalently that $m(g(0), g(1), \zeta) \geq 0$, the following steps are performed, closely following Algorithm 1 on pages 708-709 of Chernozhukov, Lee, and Rosen (2013).

1. Compute $\hat{V} \equiv \hat{p}\hat{\Omega}\hat{p}'$ and set $\hat{\Sigma}$ to be the corresponding correlation matrix. Set $s = \sqrt{diag\left(\hat{V}\right)/n}$, the vector of standard errors of each of the sample moment functions in $m\left(g(0), g(1), \hat{\zeta}\right)$. For inference on $(g(0), g(1))$ in the example of Section 8.2.1, with a discrete instrument with K points of support, there are $4K$ such moment functions, K for each of (84)-(87).

2. Simulate a large number, R, of draws of a vector W from the multivariate $N\left(0, \hat{\Sigma}\right)$ distribution.

3. Compute the contact set estimator $\widehat{\mathcal{J}}$ which estimates the identity of the elements of $m(g(0), g(1), \zeta)$ that are violated or close to binding. This is done by setting $\tilde{\gamma}_n = 1 - 0.1 \times (\log n)^{-1}$, computing $\overline{k}$ to be the $\tilde{\gamma}_n$ quantile of $\max(W_1, ..., W_J)$ over the simulation draws, and then taking the indexes, j, of the sample moments $m_j\left(g(0), g(1), \hat{\zeta}\right)$ that satisfy

$$m_j\left(g(0), g(1), \hat{\zeta}\right) \leq \min_{\ell \in \{1, ..., J\}} \left\{m_\ell\left(g(0), g(1), \hat{\zeta}\right) + \overline{k}s_\ell\right\} + 2\overline{k}s_j, \tag{88}$$

so that $\widehat{\mathcal{J}}$ is the set of indices j that fulfill (88). Here $\overline{k}$ is a high level quantile of $\max(W_1, ..., W_J)$, so as to guarantee that $\widehat{\mathcal{J}}$ contains all j such that $m_j(g(0), g(1), \zeta) = 0$ with probability approaching one as $n \to \infty$.

4. From the simulation draws of W, select those components $\tilde{W}$ whose indexes appear in the set $\widehat{\mathcal{J}}$. Compute $k_{1-\alpha}$ which is the $1 - \alpha$ quantile of $\max\left(\tilde{W}\right)$ and compute the precision corrected minimum of *all* the elements of $m\left(g(0), g(1), \hat{\zeta}\right)$ as

$$t = \min_{\ell \in \{1, ..., J\}} \left\{m_\ell\left(g(0), g(1), \hat{\zeta}\right) + k_{1-\alpha}s_\ell\right\}.$$

This quantity t is the minimum of the precision-corrected moments, each adjusted upward by an amount proportional to their standard error, specifically with scale factor given by the critical value $k_{1-\alpha}$. The hypothesis $(m(g(0), g(1), \zeta) \geq 0$ is rejected at level α if $t < 0$.

The same steps can be used to test the hypothesis that any $(g(0), g(1))$ belongs to the set $\mathcal{A}_\downarrow$, using the inequalities (68) and (69). The set of $(g(0), g(1))$ pairs that belong to the confidence set are those with $g(0) \leq g(1)$ that are not rejected by the inequality test for $\mathcal{A}_\uparrow$, together with those with $g(0) \geq g(1)$ that are not rejected by the inequality test for $\mathcal{A}_\downarrow$.

All confidence sets reported in this Section were computed by inverting moment inequality hypothesis tests following the above steps, using the particular collections of inequalities appearing in each characterization in place of the moment inequalities (84)-(87) with parameter of interest θ replacing $(g(0), g(1))$, and the corresponding appropriately defined inference parameters. In each case the inference parameters are population expectations of indicator functions of observable variables, so that inference parameters $\hat{\zeta}$ and the corresponding asymptotic variance matrix $\hat{\Omega}$ were easy to compute.

In order to draw Figure 8, each point in a dense collection of points covering a large neighborhood around the analog set estimate for $(g(0), g(1))$ was tested for inclusion in the identified set using each of the three different configurations of instruments. The sets in the figure are plots of all points not rejected by these tests.

In all tables in which median-corrected bounds and confidence intervals are reported on a univariate parameter of interest, only a one-dimensional search was required. To find the upper and lower bounds of these intervals reported in the tables, points were tested in increments of 0.0001 between each boundary of the analog set estimate and a very distant value outside the analog set. In many cases this distant point was a logical bound on the value of the parameter, such as 0 and 1 for threshold values $g(0)$ and $g(1)$. The search for lower and upper median-corrected bounds and confidence interval endpoints was then refined by a further search at points evenly spaced by 0.00001 between the closest points rejected and furthest points not rejected. The tables report the furthest value from the analog set endpoints not rejected that were obtained by this refined search, rounded to four significant digits. Further details can be found in the replication files posted on-line.[49]

9   Future Directions and Concluding Remarks

The GIV models set out in this chapter often deliver set identification of parameters with sets that can be characterized by systems of moment equalities and inequalities. There is a large and developing literature on the topic of estimation and inference under partial identification. A review of the partial identification literature and alternative approaches to identification analysis and methods for inference are offered in Molinari (2019). Ho and Rosen (2017) gives a recent survey of applications and Canay and Shaikh (2017) surveys inference methods.

We first consider some aspects of the research frontier on inference using partial identifying models that are especially pertinent to GIV models. We conclude with comments on the use of GIV models and methods in econometric research.

---

[49]https://drive.google.com/open?id=1wRpQWqIbVioEC4fej9vR69w2yByByBd7ew.

9.1 Estimation and inference

The application of Section 8 employs parametric inference procedures introduced in Chernozhukov, Lee, and Rosen (2013), thus illustrating one approach for estimation and inference by way of example. The characterization of identified sets in that application comprised moment inequalities that featured conditional probabilities with discretely supported conditioning variables. That led to inference based on a finite collection of unconditional moment inequalities. Other approaches to inference can be used in such contexts, such as those of Chernozhukov, Hong, and Tamer (2007), Rosen (2008), Romano and Shaikh (2010), Andrews and Soares (2010), Bugni (2010), and Canay (2010). As in the analysis reported here, many of the test statistics used in these procedures incorporate studentized versions of sample moments in order to achieve good performance. When using studentized sample moments, estimates of the asymptotic variance of sample conditional moments play an important role. These sample conditional moments can be written as functions of estimates of easily estimated inference parameters, and the required asymptotic variances can be computed as here, by application of the delta method. When conditioning variables are continuous, methods for inference with conditional moment inequalities can be used, for example by using the nonparametric procedures of Chernozhukov, Lee, and Rosen (2013), or other approaches for conditional moment inequalities such as those of Andrews and Shi (2013, 2014), Armstrong (2015), Armstrong and Chan (2016), Chetverikov (2018), and Lee, Song, and Whang (2018). For a more nuanced treatment of different approaches for estimation and inference with partial identification we refer to Molinari (2019) and Canay and Shaikh (2017).

An important direction in current research, one to which the applications of Section 8 point, is that of inference on projections. In economic applications interest frequently lies in performing inference on particular, often low-dimensional, functionals of partially identified structures, such as components of parameter vectors, partial effects, average treatment effects, or counterfactual probabilities. However most inference methods for set identified parameters deliver joint confidence regions for entire partially identified parameter vectors. Taking projections of such regions can result in valid inference on functionals, but at great cost in terms of providing conservative inference – sometimes referred to as coverage inflation – and hence decreased power. Research on inference on projections aims to solve this problem and the results are also potentially useful in alleviating the computational burden attendant on computing the identified set for a high dimensional parameter vector.

In the binary outcome application set out in Section 8.2 we conduct inference on projections. In the context of that model with discrete covariates, sharp inequality characterizations of projections onto individual parameter components can be obtained by Fourier-Motzkin elimination when inequalities are linear in parameters. That enables inference without coverage inflation using a standard moment inequality inference procedure. In many settings the lower dimensional parameters of interest do not feature linearly in the characterization of the identified set. There may be

non-linear dependencies across inequalities, as there is for example in the characterization of the identified sets for average treatment effects at fixed values of z_1, $\Delta(z_1)$, in the parametric model considered in Section 8.2.2. This issue can arise even in quite relatively simple cases such as this, with discrete endogenous and exogenous variables. Recent important work on general approaches to the problem of subvector inference with partial identification that avoid coverage inflation include Bugni, Canay, and Shi (2017), Kaido, Molinari, and Stoye (forthcoming), Chen, Christensen, and Tamer (2018), and Belloni, Bugni, and Chernozhukov (2018).

Another important focus in current research is inference in partially identifying models featuring large numbers of moment inequalities. GIV models – like some others in the partial identification literature – often provide characterizations of identified sets comprising a huge number of moment inequalities and equalities. A model that delivers a rich set of observable implications is attractive from the standpoint of identification, as more observable implications can lead to smaller identified sets. But the complexity of such characterizations can pose a challenge for estimation and inference, and there may be complex trade-offs between the identifying power of moment inequalities and the precision with which the constituent moments can be estimated. Papers that address some of the issues presented by a large number of moment inequalities and equalities include Menzel (2009), Chernozhukov, Chetverikov, and Kato (forthcoming), Andrews and Shi (2017), and Bugni, Caner, Kock, and Lahiri (2016). In addition, Belloni, Bugni, and Chernozhukov (2018) study the aforementioned problem of subvector inference, with methods designed specifically to deal with many moment inequalities.

In the context of GIV models, cases with many moment inequalities are obtained when a model's collection of core-determining sets is large. While a core-determining collection defines a collection of moment inequalities that are sufficient for sharp characterization of the identified set, there is also a possibility that inequalities obtained by sets that are not core-determining could be estimated more precisely in finite samples, so that these are nonetheless useful for conducting inference. There may be cases in which the best quality of inference about particular functionals of structures is obtained using a small selection of moment inequalities. More research aimed at guiding that choice is needed.

9.2 Generalized instrumental variable models in econometric research

In this chapter we have described the extension of the scope of classical IV methods to models in which unobserved variables are related to observed variables by correspondences. This arises in many applications that feature in 21st century econometrics, for example, when models admit high dimensional heterogeneity, when endogenous outcomes are discrete while unobserved variables are continuous, and when the restrictions of economics enter models as inequality restrictions, for example as positive profit conditions.

GIV models can be complete or incomplete. Incomplete models arise when aspects of the

process determining endogenous outcomes are not completely specified. This arises for example when models of a process admit multiple equilibria leaving equilibrium selection unspecified. In econometrics there is a long history of the use of incomplete models that arise because some of the simultaneous equations determining the unique values of outcomes delivered by a process are not specified. We have described how these "single equation" and "limited information" models can be used in econometric analysis of data when the relationship between unobserved variables and observed variables is a correspondence.

The identification analysis set out here has been cast in the GIV framework of Chesher and Rosen (2017a). This builds on the classical modelling framework set out in various papers in the Cowles Commission Monograph 10 [Koopmans (1950)] with an extension to that framework that permits identification analysis of incomplete models. An essential element here is the extension of the concept of observational equivalence to cover cases in which a structure can deliver a non-singleton set of values of endogenous outcomes.

Classical IV restrictions limit the covariation of structural unobservable variables and instruments through statistical independence restrictions or through weaker conditional mean and conditional quantile independence restrictions. These and other restrictions on the distributions of unobserved variables are easily incorporated in identification analysis in the GIV framework because of its focus on the sets of values of unobservables that are compatible with values of observed variables.

GIV models can be point or partially identifying, and in some cases identified sets may be large. Smaller identified sets, and sometimes point identification, may be achievable by imposing additional or different restrictions. Building a complete model or bringing a conditional independence restriction on board are common devices. But the maxim on pages 169-170 in Koopmans and Reiersøl (1950) is relevant.

> Scientific honesty demands that the specification of a model be based on prior knowledge of the phenomenon being studied and possibly on criteria of simplicity, but not on the desire for identifiability of characteristics in which the researcher happens to be interested.

The sharp identified set delivered by a GIV model and a particular distribution of observed variables contains all sets and points identified by restricted versions of the model using that distribution, for example all the complete models nested within an incomplete GIV model. So an important element of the information obtained when using a possibly partially identifying GIV model is the knowledge imparted about the sensitivity of inference to the choice of alternative more restrictive, maybe complete models.

In many situations in which economists are asked for policy advice there are no widely credible complete models and incomplete models must be used. The analysis presented in this chapter

enables econometric model construction and application using credible incomplete models suitable for use in modern econometric analysis in which there is careful attention paid to the modeling and impact of complex forms of across individual heterogeneity.

The resulting econometric models are typically set-identifying. The identification analysis of these types of model presented in this Chapter delivers sharp characterizations of the identified sets of structures delivered by incomplete models with complex specifications of heterogeneity. The task now is to continue to develop the computational and inferential procedures to bring these methods into every-day econometric practice.

References

ANDREWS, D. W. K., AND X. SHI (2013): "Inference Based on Conditional Moment Inequalities," *Econometrica*, 81(2), 609–666.

——— (2014): "Nonparametric Inference Based on Conditional Moment Inequalities," *Journal of Econometrics*, 179(1), 31–45.

——— (2017): "Inference Based on Many Conditional Moment Inequalities," *Journal of Econometrics*, 196(2), 275–287.

ANDREWS, D. W. K., AND G. SOARES (2010): "Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection," *Econometrica*, 78(1), 119–157.

ANDREWS, D. W. K., AND J. H. STOCK (2007): "Inference with Weak Instruments," in *Advances in Economics and Econometrics, Theory and Applications: Ninth World Congress of the Econometric Society, Vol. III*, ed. by R. Blundell, W. K. Newey, and T. Persson. Cambridge University Press.

ANGRIST, J. (2001): "Estimation of Limited-Dependent Variable Models with Dummy Endogenous Regressors: Simple Strategies for Empirical Practice," *Journal of Business and Economic Statistics*, 19(1), 2–14.

ANGRIST, J., AND W. N. EVANS (1998): "Children and Their Parents' Labor Supply: Evidence from Exogenous Variation in Family Size," *American Economic Review*, 88(3), 450–487.

ANGRIST, J., G. IMBENS, AND D. RUBIN (1996): "Identification of Causal Effects Using Instrumental Variables," *Journal of the American Statistical Association*, 91(434), 444–455.

ANGRIST, J. D., AND J.-S. PISCHKE (eds.) (2009): *Mostly Harmless Econometrics*. Princeton University Press, Princeton, NJ.

ARMSTRONG, T. B. (2015): "Asymptotically Exact Inference in Conditional Moment Inequality Models," *Journal of Econometrics*, 186(1), 51–65.

ARMSTRONG, T. B., AND H. P. CHAN (2016): "Multiscale adaptive inference on conditional moment inequalities," *Journal of Econometrics*, 194(1), 92–116.

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

BARSEGHYAN, L., M. COUGHLIN, F. MOLINARI, AND J. C. TEITELBAUM (2017): "Heterogeneous Consideration Sets and Preferences," In preparation.

BELLONI, A., F. BUGNI, AND V. CHERNOZHUKOV (2018): "Subvector Inference in PI Models with with Many Moment Inequalities," arXiv:1806.11466.

BERESTEANU, A., I. MOLCHANOV, AND F. MOLINARI (2011): "Sharp Identification Regions in Models with Convex Moment Predictions," *Econometrica*, 79(6), 1785–1821.

——— (2012): "Partial Identification Using Random Set Theory," *Journal of Econometrics*, 166(1), 17–32.

BERRY, S., AND E. TAMER (2007): "Identification in Models of Oligopoly Entry," in *Advances in Economics and Econometrics: Theory and Applications, Ninth World Congress, Volume II*, ed. by R. Blundell, W. Newey, and T. Persson, pp. 46–85. Cambridge University Press.

BERRY, S. T., AND G. COMPIANI (2019): "An Instrumental Variable Approach to Dynamic Models," working paper, University of California at Berkeley and Yale University.

BERRY, S. T., AND P. A. HAILE (2018): "Identification of Nonparametric Simultaneous Equations Models with a Residual Index Structure," *Econometrica*, 86(1), 289–315.

BLUNDELL, R., AND R. MATZKIN (2014): "Control Functions in Nonseparable Simultaneous Equations Models," *Quantitative Economics*, 5(2), 271–295.

BLUNDELL, R., AND J. L. POWELL (2003): "Endogeneity in Nonparametric and Semiparametric Regression Models," in *Advances in Economics and Econometrics: Theorey and Applications, Eighth World Congress*, ed. by M. Dewatripont, L. P. Hansen, and S. J. Turnovsky, vol. 2, pp. 312–357. Cambridge University Press.

BOWDEN, R. (1973): "The Theory of Parametric Identification," *Econometrica*, 41(6), 1069–1074.

BREUSCH, T. (1986): "Hypothesis Testing in Unidentified Models," *Review of Economic Studies*, 53(4), 635–651.

BROWN, B. W. (1983): "The Identification Problem in Systems Nonlinear in the Variables," *Econometrica*, 51(1), 175–196.

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

BUGNI, F., I. CANAY, AND X. SHI (2015): "Specification Tests for Partially Identified Models Defined by Moment Inequalities," *Journal of Econometrics*, 185(1), 259–282.

——— (2017): "Inference for Subvectors and Other Functions of Partially Identified Parameters in Moment Inequality Models," *Quantitative Economics*, 8(1), 1–38.

BUGNI, F., M. CANER, A. B. KOCK, AND S. LAHIRI (2016): "Inference in Partially Identified Models with Many Moment Inequalities Using Lasso," arXiv:1604.02309.

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

CANAY, I., AND A. SHAIKH (2017): "Practical and Theoretical Advances for Inference in Partially Identified Models," in *Advances in Economics and Econometrics: Eleventh World Congress, Volume II*, ed. by B. Honore, A. Pakes, M. Piazzesi, and L. Samuelson. Cambridge University Press.

CHEN, X., T. M. CHRISTENSEN, AND E. TAMER (2018): "Monte Carlo Confidence Sets for Identified Sets," *Econometrica*, 86(6), 1965–2018.

CHEN, X., H. HONG, AND D. NEKIPELOV (2011): "Nonlinear Models of Measurement Error," *Journal of Economic Literature*, 49(4), 901–937.

CHERNOZHUKOV, V., D. CHETVERIKOV, AND K. KATO (forthcoming): "Inference on Causal and Structural Parameters Using Many Moment Inequalities," *Review of Economic Studies*.

CHERNOZHUKOV, V., AND C. HANSEN (2005): "An IV Model of Quantile Treatment Effects," *Econometrica*, 73(1), 245–261.

CHERNOZHUKOV, V., H. HONG, AND E. TAMER (2007): "Estimation and Confidence Regions for Parameter Sets in Econometric Models," *Econometrica*, 75(5), 1243–1284.

CHERNOZHUKOV, V., S. LEE, AND A. M. ROSEN (2013): "Intersection Bounds: Estimation and Inference," *Econometrica*, 81(2), 667–737.

CHESHER, A. (2003): "Identification in Nonseparable Models," *Econometrica*, 71(5), 1405–1441.

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

——— (2013): "Semiparametric Structural Models of Binary Response: Shape Restrictions and Partial Identification," *Econometric Theory*, 29(2), 231–266.

CHESHER, A., AND A. M. ROSEN (2013): "What Do Instrumental Variable Models Deliver With Discrete Dependent Variables?," *American Economic Review: Papers and Proceedings*, 103(3), 557–562.

——— (2014): "An Instrumental Variable Random Coefficients Model for Binary Outcomes," *Econometrics Journal*, 17(2), S1–S19.

———— (2015): "Counterfactual Worlds," CeMMAP working paper CWP22/15.

———— (2017a): "Generalized Instrumental Variable Models," *Econometrica*, 85(3), 959–989.

CHESHER, A., AND A. M. ROSEN (2017b): "Incomplete English Auction Models with Heterogeneity," CeMMAP working paper CWP27/17.

CHESHER, A., A. M. ROSEN, AND K. SMOLINSKI (2013): "An Instrumental Variable Model of Multiple Discrete Choice," *Quantitative Economics*, 4(2), 157–196.

CHESHER, A., AND K. SMOLINSKI (2012): "IV Models of Ordered Choice," *Journal of Econometrics*, 166(1), 33–48.

CHETVERIKOV, D. (2018): "Adaptive Tests of Conditional Moment Inequalities," *Econometric Theory*, 34(1), 186–227.

CILIBERTO, F., AND E. TAMER (2009): "Market Structure and Multiple Equilibria in Airline Markets," *Econometrica*, 77(6), 1791–1828.

COLLARD-WEXLER, A. (2013): "Demand Fluctuations in the Ready-Mix Concrete Industry," *Econometrica*, 81(3), 1003–1037.

D'HAULTFOEUILLE, X., AND P. FEVRIER (2015): "Identification of Nonseparable Triangular Models With Discrete Instruments," *Econometrica*, 83(3), 1199–1210.

ENGERS, M., AND S. STERN (2002): "Family Bargaining and Long Term Care," *International Economic Review*, 43(1), 73–114.

FLORENS, J.-P., J. J. HECKMAN, C. MEGHIR, AND E. VYTLACIL (2008): "Identification of Treatment Effects Using Control Functions in Models with Continuous, Endogenous Treatment and Heterogeneous Effects," *Econometrica*, 76(5), 1191–1206.

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

HAAVELMO, T. (1943): "The Statistical Implications of a System of Simultaneous Equations," *Econometrica*, 11(1), 1–12.

———— (1944): "The Probability Approach in Econometrics," *Econometrica*, 12(Supplement), iii–vi and 1–115.

HAILE, P. A., AND E. TAMER (2003): "Inference with an Incomplete Model of English Auctions," *Journal of Political Economy*, 111(1), 1–51.

HANSEN, L. P. (1982): "Large Sample Properties of Generalized Method of Moments Esitmators," *Econometrica*, 50(4), 1029–1054.

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

——— (2010): "Building Bridges Bridges Between Structural and Program Evaluation Approaches to Evaluating Policy," *Journal of Economic Literature*, 48, 356–398.

HECKMAN, J. J., AND R. ROBB (1985): "Alternative Methods for Evaluating the Impact of Interventions," in *Longitudinal Analysis of Labor Market Data*, ed. by J. J. Heckman, and B. Singer, pp. 156–245. Cambridge University Press.

HO, K., AND A. M. ROSEN (2017): "Partial Identification in Applied Research: Benefits and Challenges," in *Advances in Economics and Econometrics: Eleventh World Congress, Volume II*, ed. by B. Honore, A. Pakes, M. Piazzesi, and L. Samuelson. Cambridge University Press.

HONG, H., AND E. TAMER (2003): "Inference in Censored Models with Endogenous Regressors," *Econometrica*, 71(3), 905–932.

HOROWITZ, J. L. (2011): "Applied Nonparametric Instrumental Variables Estimation," *Econometrica*, 79(2), 347–394.

HOROWITZ, J. L., AND C. F. MANSKI (1998): "Censoring of outcomes and regressors due to survey nonresponse: Identification and estimation using weights and imputations," *Journal of Econometrics*, 84(1), 37–58.

——— (2000): "Nonparametric Analysis of Randomized Experiments with Missing Covariate and Outcome Data," *Journal of the American Statistical Association*, 95(449), 77–84.

HURWICZ, L. (1950): "Generalization of the Concept of Identification," in *Statistical Inference in Dynamic Economic Models*, ed. by C. C. for Research in Economics, pp. 245–257. John Wiley and Sons.

IMBENS, G. (2014): "Instrumental Variables: An Econometrician's Perspective," *Statistical Science*, 29(3), 323–358.

IMBENS, G., AND C. F. MANSKI (2004): "Confidence Intervals for Partially Identified Parameters," *Econometrica*, 72(6), 1845–1857.

IMBENS, G., AND W. NEWEY (2009): "Identification and Estimation of Triangular Simultaneous Equations Models with Additivity," *Econometrica*, 77(5), 1481–1512.

JOVANOVIC, B. (1989): "Observable Implications of Models with Multiple Equilibria," *Econometrica*, 57(6), 1431–1437.

KAIDO, H., F. MOLINARI, AND J. STOYE (forthcoming): "Inference for Projections of Identified Sets," *Econometrica*.

KHAN, S., M. PONOMAREVA, AND E. TAMER (2011): "Sharpness in Randomly Censored Linear Models," *Economics Letters*, 113(1), 23–25.

KHAN, S., AND E. TAMER (2009): "Inference on Endogenously Censored Regression Models Using Conditional Moment Inequalities," *Journal of Econometrics*, 152(2), 104–119.

KIM, D. (forthcoming): "Partial identification in Nonseparable Count Data IV Models," *Econometrics Journal*.

KLEIN, R. W., AND R. H. SPADY (2003): "An Efficient Semiparametric Estimator for Binary Response Models," *Econometrica*, 61(2), 387–421.

KOOPMANS, T. (1949): "Identification Problems in Economic Model Construction," *Econometrica*, 17(2), 125–144.

KOOPMANS, T., AND O. REIERSØL (1950): "The Identification of Structural Characteristics," *Annals of Mathematical Statistics*, 21(2), 165–181.

KOOPMANS, T., H. RUBIN, AND R. LEIPNIK (1950): "Measuring the equation systems of dynamic economics," in *Statistical Inference in Dynamic Economic Models*, ed. by T. C. Koopmans, pp. 53–237. John Wiley and Sons.

KOOPMANS, T. C. E. (1950): *Statistical Inference in Dynamic Economic Models*. John Wiley and Sons, New York, Cowles Commission Monograph 10.

KREIDER, B., AND J. PEPPER (2007): "Disability and Employment: Reevaluating the Evidence in Light of Reporting Errors," *Journal of the American Statistical Association*, 102(478), 432–441.

LEAMER, E. E. (1983): "Let's Take the Con out of Econometrics," *American Econmic Review*, 73(1), 31–43.

LEE, S. (2007): "Endogeneity in quantile regression models: A control function approach," *Journal of Econometrics*, 141(2), 1131–1158.

LEE, S., K. SONG, AND Y.-J. WHANG (2018): "Testing for a General Class of Functional Inequalities," *Econometric Theory*, 34(5), 1018–1064.

MANSKI, C. F. (2003): *Partial Identification of Probability Distributions*. Springer-Verlag, New York.

MANSKI, C. F., AND J. PEPPER (2000): "Monotone Instrumental Variables: With an Application to the Returns to Schooling," *Econometrica*, 68(4), 997–1010.

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

MARSCHAK, J. (1950): "Statistical Inference in Economics: An Introduction," in *Statistical Inference in Dynamic Economic Models*, ed. by T. C. Koopmans, pp. 1–52. John Wiley and Sons, Cowles Commission for Research in Economics Monograph 10.

MATZKIN, R. (2007): "Nonparametric Identification," in *The Handbook of Econometrics*, ed. by J. J. Heckman, and E. E. Leamer, vol. 6b, pp. 5307–5368. North-Holland.

——— (2008): "Identification in Nonparametric Simultaneous Equations Models," *Econometrica*, 76(5), 945–978.

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

MENZEL, K. (2009): "Consistent Estimation with Many Moment Inequalities," *Journal of Econometrics*, 182(2), 329–350.

MOLCHANOV, I., AND F. MOLINARI (2018): *Random Sets in Econometrics*. Cambridge University Press, Cambridge, U.K.

MOLCHANOV, I. S. (2005): *Theory of Random Sets*. Springer Verlag, London.

MOLINARI, F. (2019): "Econometrics with Partial Identification," in *The Handbook of Econometrics*, ed. by S. Durlauf, L. P. Hansen, H. J. J., and R. Matzkin, vol. 7a. Elsevier.

MOURIFIE, I., M. HENRY, AND R. MEANGO (2017): "Sharp Bounds for the Roy Model," working paper, University of Toronto.

NEWEY, W., AND D. MCFADDEN (1994): "Large Sample Estimation and Hypothesis Testing," in *The Handbook of Econometrics*, ed. by R. F. Engle, and D. L. McFadden, vol. 4, pp. 193–281. North-Holland.

NEWEY, W. K., AND J. L. POWELL (2003): "Instrumental Variable Estimation of Nonparametric Models," *Econometrica*, 71(5), 1565–1578.

NEWEY, W. K., J. L. POWELL, AND F. VELLA (1999): "Nonparametric Estimation of Triangular Simultaneous Equations Models," *Econometrica*, 67(3), 565–603.

NGUYEN, H. T. (2006): *An Introduction to Random Sets*. Chapman and Hall, Boca Raton.

NORBERG, T. (1992): "On the Existence of Ordered Couplings of Random Sets – with Applications," *Israel Journal of Mathematics*, 77(3), 241–264.

R CORE TEAM (2014): *R: A Language and Environment for Statistical Computing* R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0.

ROEHRIG, C. S. (1988): "Conditions for Identification in Nonparametric and Parametric Models," *Econometrica*, 56(2), 433–447.

ROMANO, J. P., AND A. M. SHAIKH (2010): "Inference for the Identified Set in Partially Identified Econometric Models," *Econometrica*, 78(1), 169–211.

ROSEN, A. M. (2008): "Confidence Sets for Partially Identified Parameters that Satisfy a Finite Number of Moment Inequalities," *Journal of Econometrics*, 146(1), 107–117.

ROTHENBERG, T. J. (1971): "Identification in Parametric Models," *Econometrica*, 39(3), 577–591.

SCHENNACH, S. (2016): "Recent Advances in the Measurement Error Literature," *Annual Review of Economics*, 8.

——— (2019): "Mismeasured and Unobserved Variables," in *The Handbook of Econometrics*, ed. by S. Durlauf, L. P. Hansen, H. J. J., and R. Matzkin, vol. 7a. Elsevier.

STATACORP (2017): "Stata Statistical Software: Release 15," College Station, TX: StataCorp LLC.

STOCK, J. H., M. YOGO, AND J. WRIGHT (2002): "A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments," *Journal of Business and Economic Statistics*, 20, 518–529.

TAMER, E. (2003): "Incomplete Simultaneous Discrete Response Models with Multiple Equilibria," *Review of Economic Studies*, 70(1), 147–167.

TEBALDI, P., A. TORGOVITSKY, AND H. YANG (2019): "Nonparametric Estimates of Demand in the California Health Insurance Exchange," NBER Working Paper No. 25827.

THEIL, H. (1953): *Repeated Least Squares Applied to Complete Equation Systems*. Central Planning Bureau, The Hague.

TORGOVITSKY, A. (2015): "Identification of Nonseparable Models Using Instruments With Small Support," *Econometrica*, 83(3), 1185–1197.

U.S. Department of Commerce, Census of population and housing [United States] (1980): "Public use microdata sample, 5-percent sample [computer file]," Washington, DC: U.S. Department of Commerce, Bureau of the Census [producer], 1983.

Wald, A. (1950): "Remarks on the Estimation of Unknown Parameters in Incomplete Systems of Equations," in *Statistical Inference in Dynamic Economic Models*, ed. by C. Koopmans, Tjalling, pp. 305–310. John Wiley and Sons, Cowles Commission for Research in Economics Monograph 10.

Wolfram Research, Inc. (2019): "Mathematica, Version 12.0," Champaign, IL.
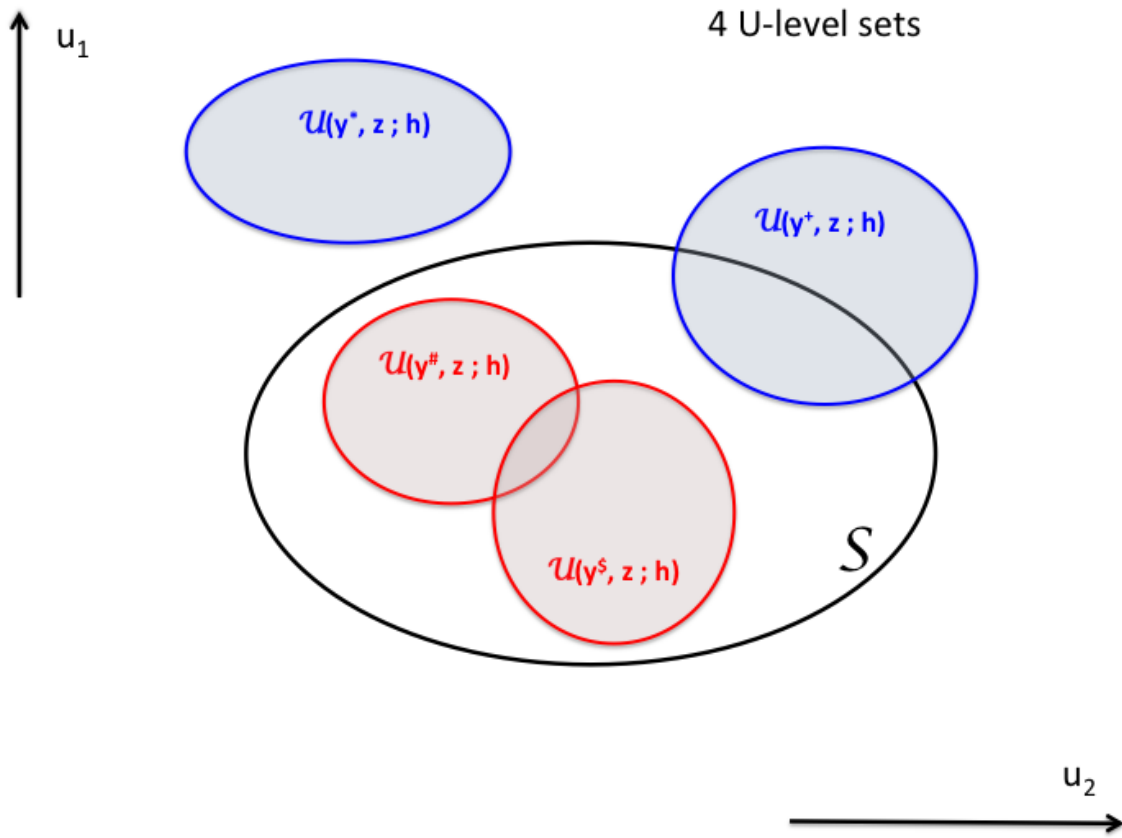
Figure 1: This Figure represents a setting in which the support of U is a subset of $\mathbb{R}^2$, and the support of $\mathcal{U}(Y, Z; h)$ given $Z = z$ has four elements as shown. The set $\mathcal{S}$ is an arbitrary fixed set in $\mathbb{R}^2$. The conditional containment and hitting probabilities $\mathbb{P}[\mathcal{U}(Y, Z; h) \subseteq \mathcal{S}|z]$ and $1 - \mathbb{P}[\mathcal{U}(Y, Z; h) \subseteq \mathcal{S}^c|z]$ are equal to $\mathbb{P}[Y \in \{y^\#, y^\$\}|z]$ and $\mathbb{P}[Y \in \{y^\#, y^\$, y^+\}|z]$, respectively.

Figure 2: Outer sets for parameters (β, γ, σ) for probability generating structure ST1 with 8 bins using the 14 inequalities generated with $N = 5$ (top pane) and the 85 inequalities generated with $N \in \{5, 7, 9\}$ (bottom pane). Dashed green lines intersect at the value of the parameters in the data generating structure.
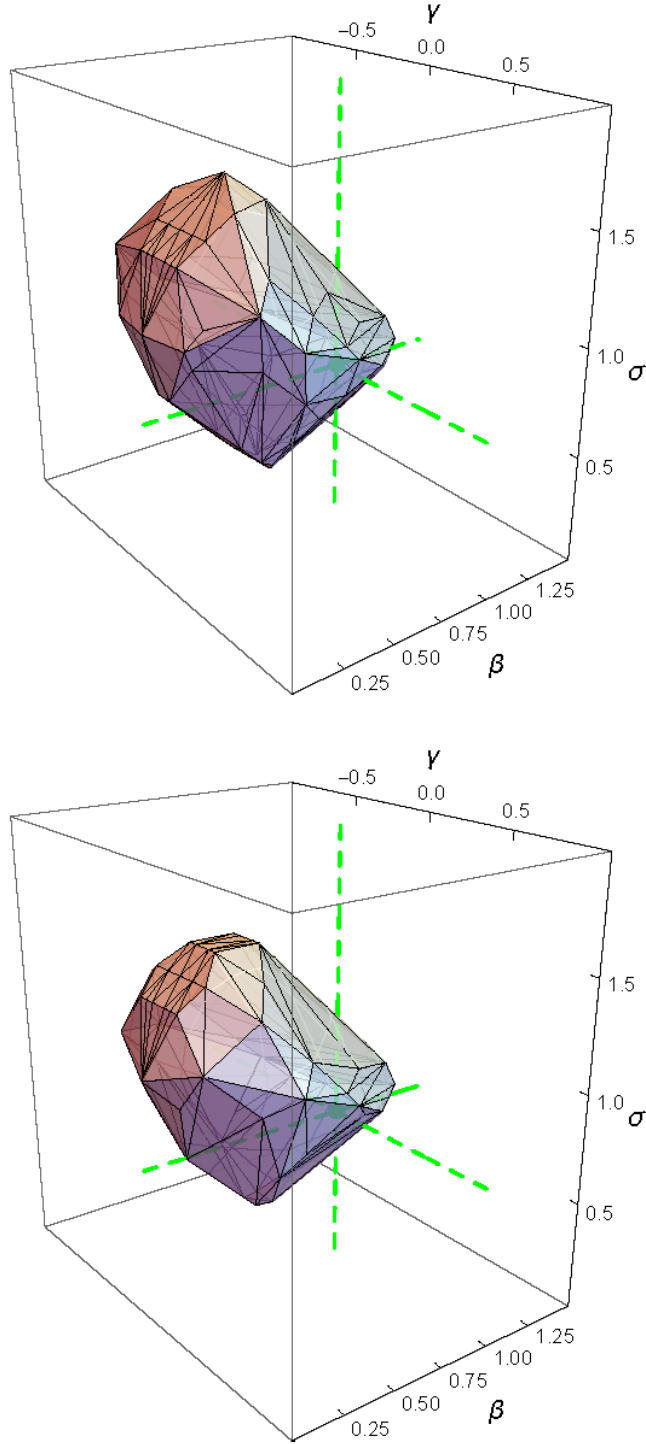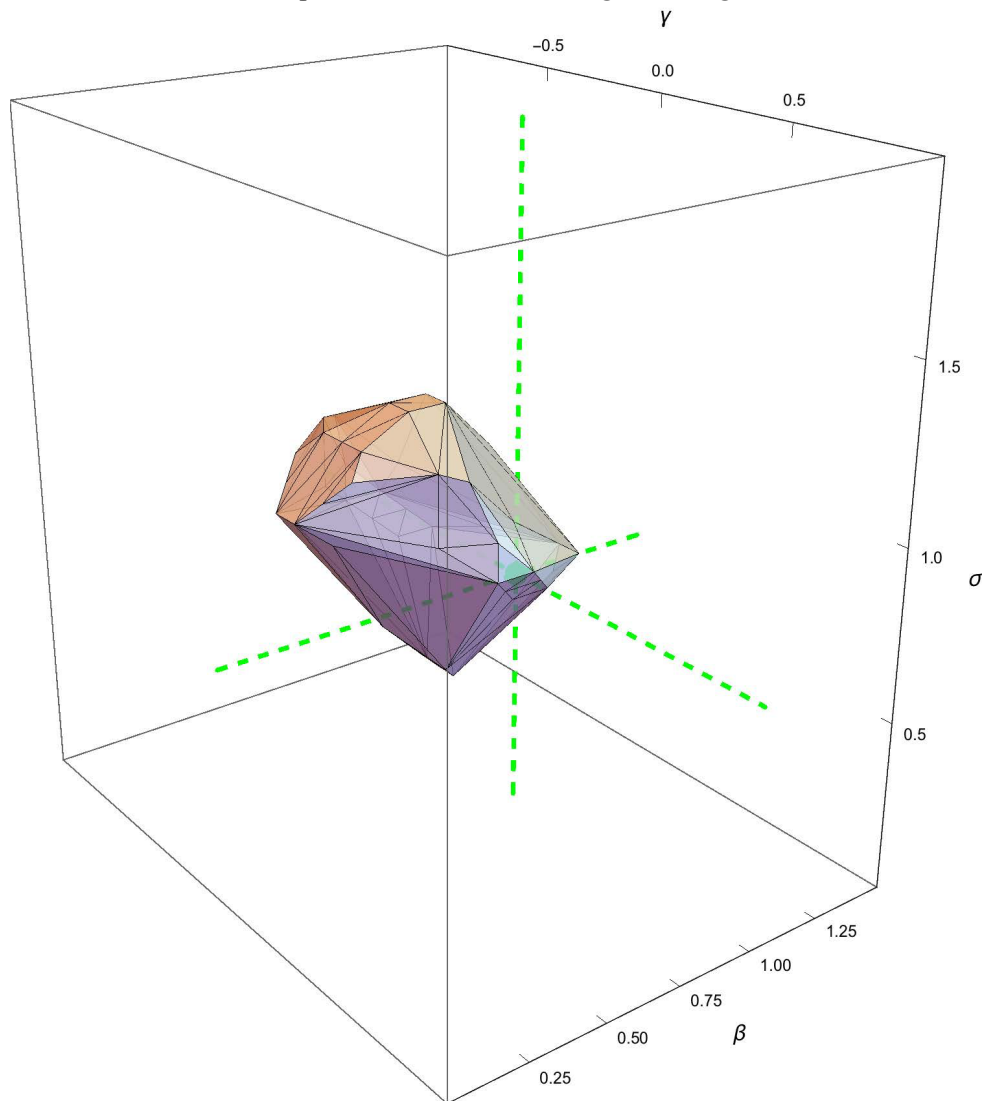
Figure 3: Outer region projections for probability generating structure ST1 onto the (β, γ), (γ, σ), and (β, σ) planes, with endogenous Y_2^* censored into 8 intervals using inequalities generated with $N \in \{5, 7, 9\}$. The red point marks the parameter values in the probability generating structure.
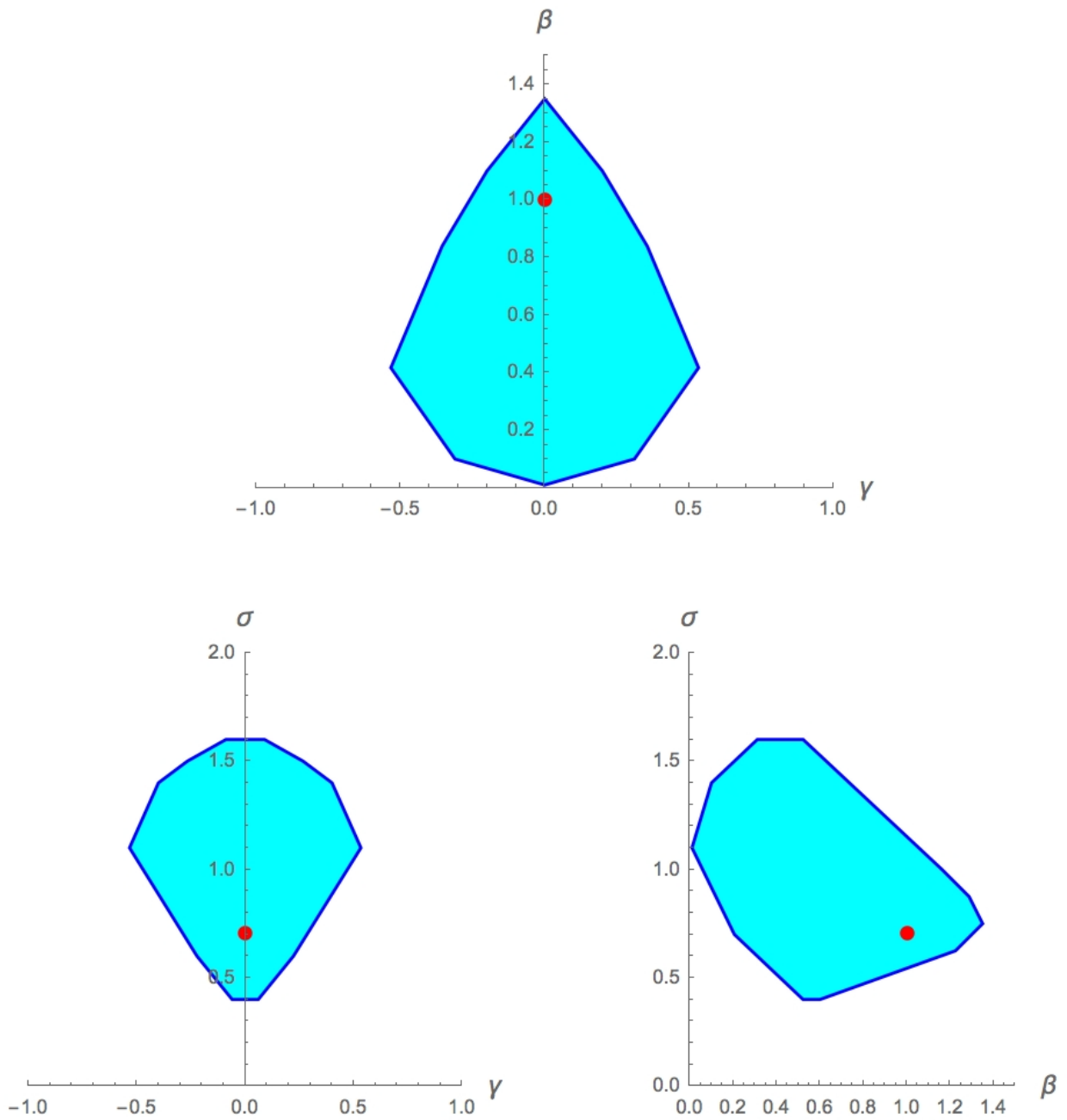
Figure 4: Outer set for parameters (β, γ, σ) for probability generating structure ST2 with Y_2^* censored into 12 bins, calculated using inequalities generated with $N \in \{5, 7, 9\}$. Dashed green lines intersect at the value of the parameters in the data generating structure.

Figure 5: Outer region projections for probability generating structure ST2 onto the (β, γ), (γ, σ), and (β, σ) planes, with endogenous Y_2^* censored into 12 intervals using inequalities generated with $N \in \{5, 7, 9\}$. The red point marks the parameter values in the probability generating structure.
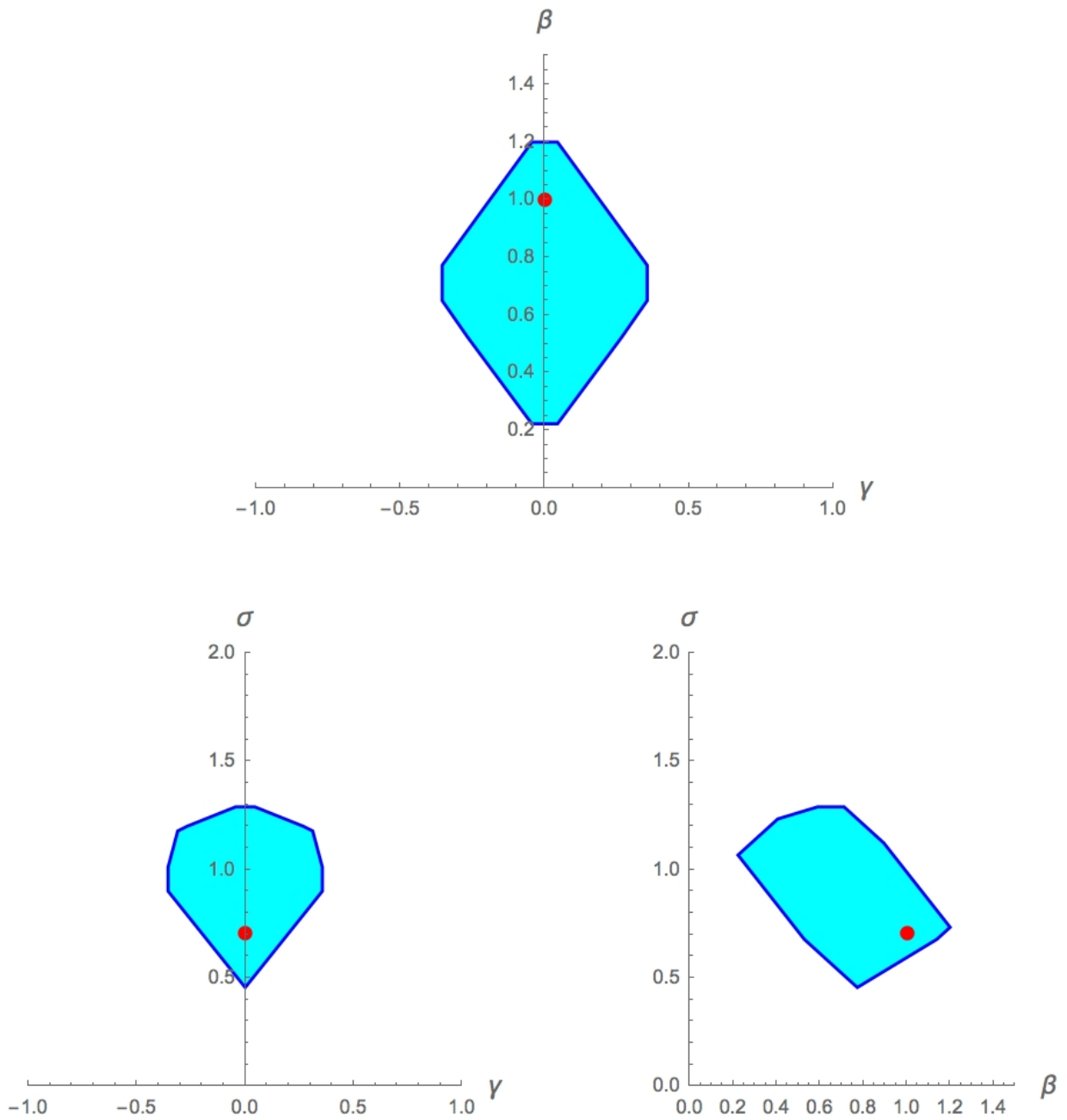
Figure 6: Outer sets for (β, γ). The top panels display sets for censoring intervals of width 0.4 and the bottom panels display sets for intervals of width 0.2. In the panels on the left $d_1 = 1$ (weaker instrument) and on the right $d_1 = 1.5$ (stronger instrument). The dark blue lines indicate boundaries of outer sets obtained with the conditional median restriction $q_{U|Z}(0.5|z) = 0$, while the inner light blue lines indicate boundaries of outer sets when U is restricted to be Gaussian, independent of Z.
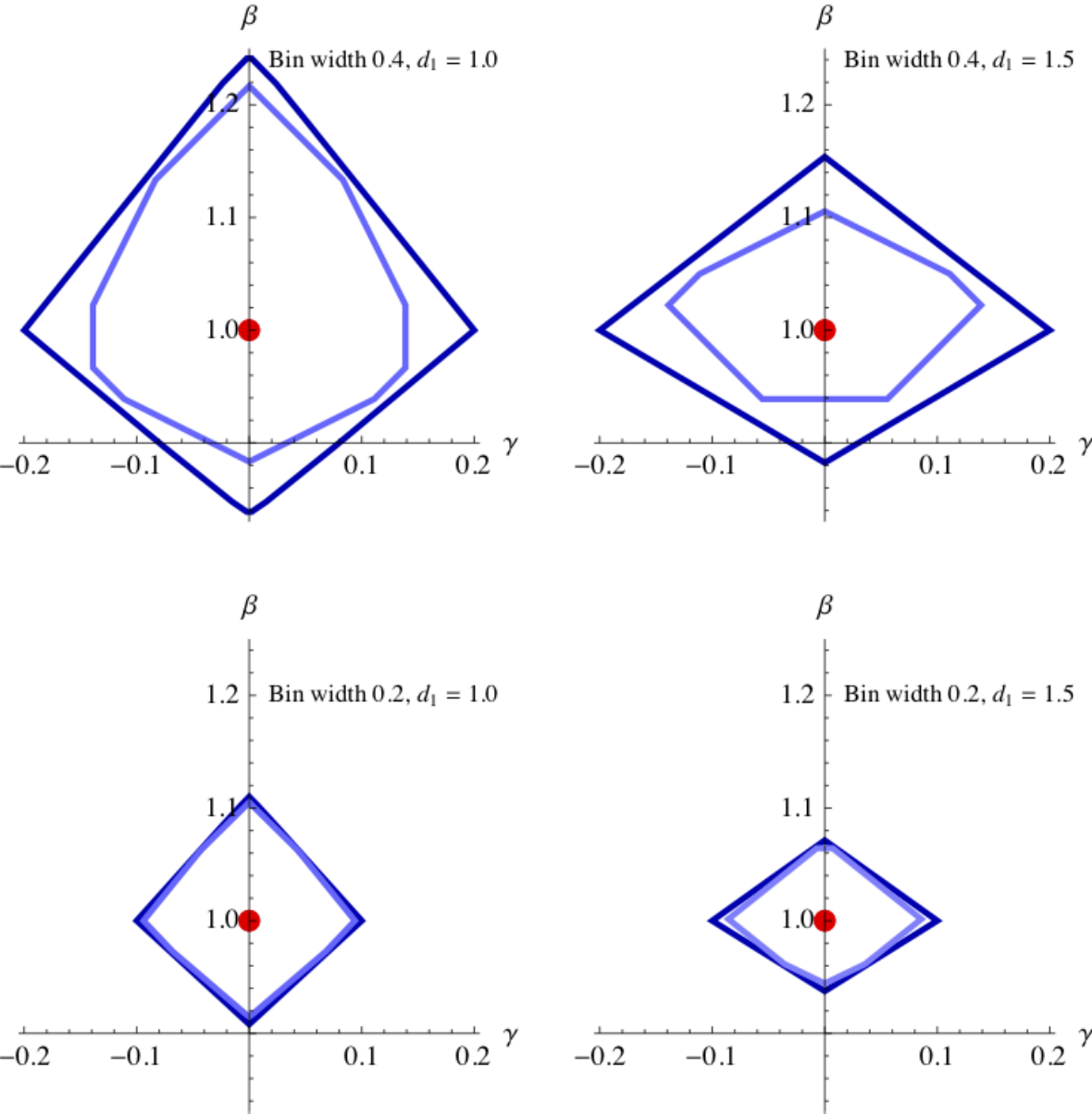
Figure 7: Analog estimates for identified sets of values of thresholds $g(0)$ (two children) and $g(1)$ (three or more children) in the GIV binary outcome model for female labor force participation. Data source: Angrist and Evans (1998).
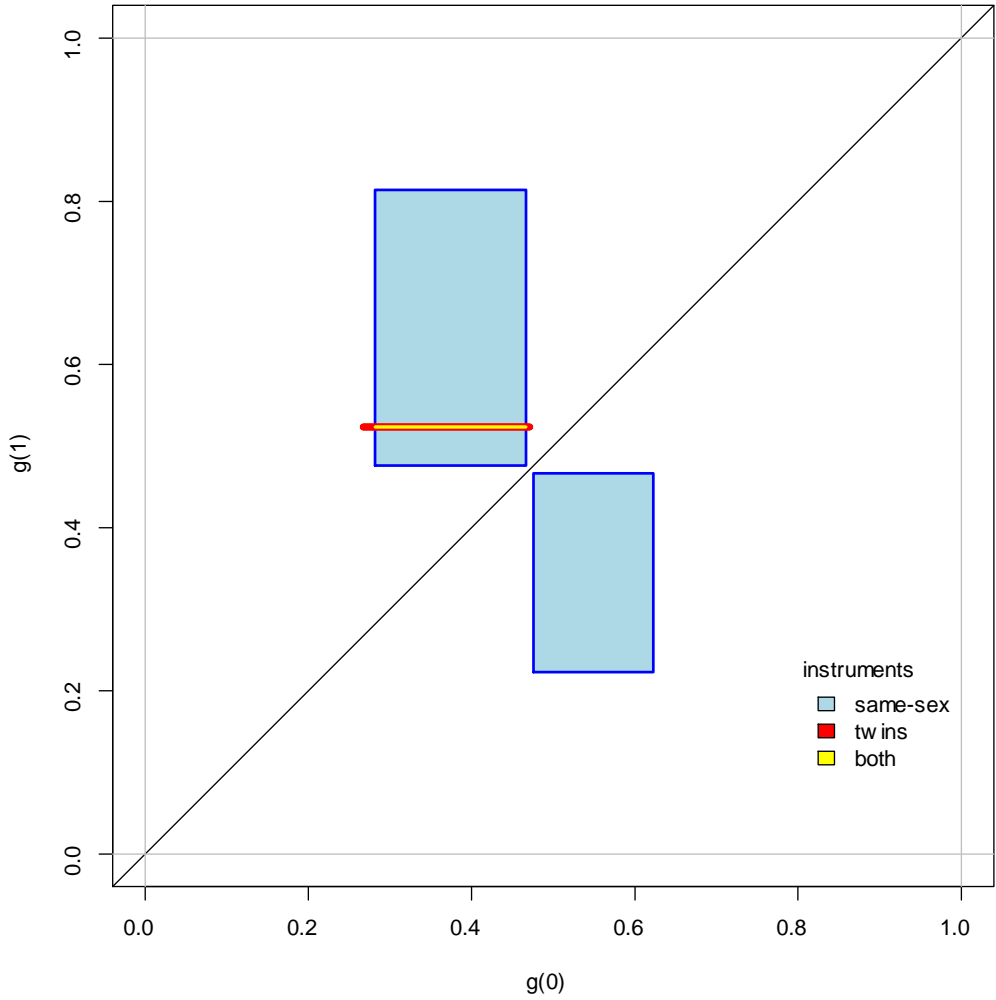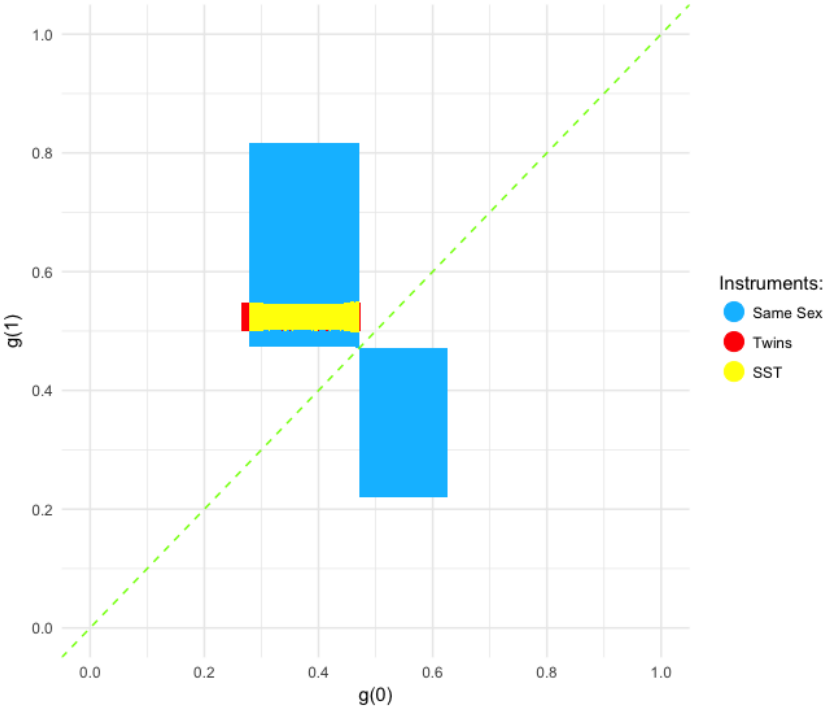
Figure 8: 95% joint confidence sets for values of thresholds $g(0)$ (two children) and $g(1)$ (three or more children) in the GIV binary outcome model for female labor force participation. Data source: Angrist and Evans (1998).

Figure 9: Analog estimates of projections of identified sets onto the space of α and β_1. Identified sets using the twins instruments and both instruments are lines, drawn here with nonnegligible thickness so that it is possible to see the reduction in the size of the projection on using the same sex instrument as well as the twins instrument.
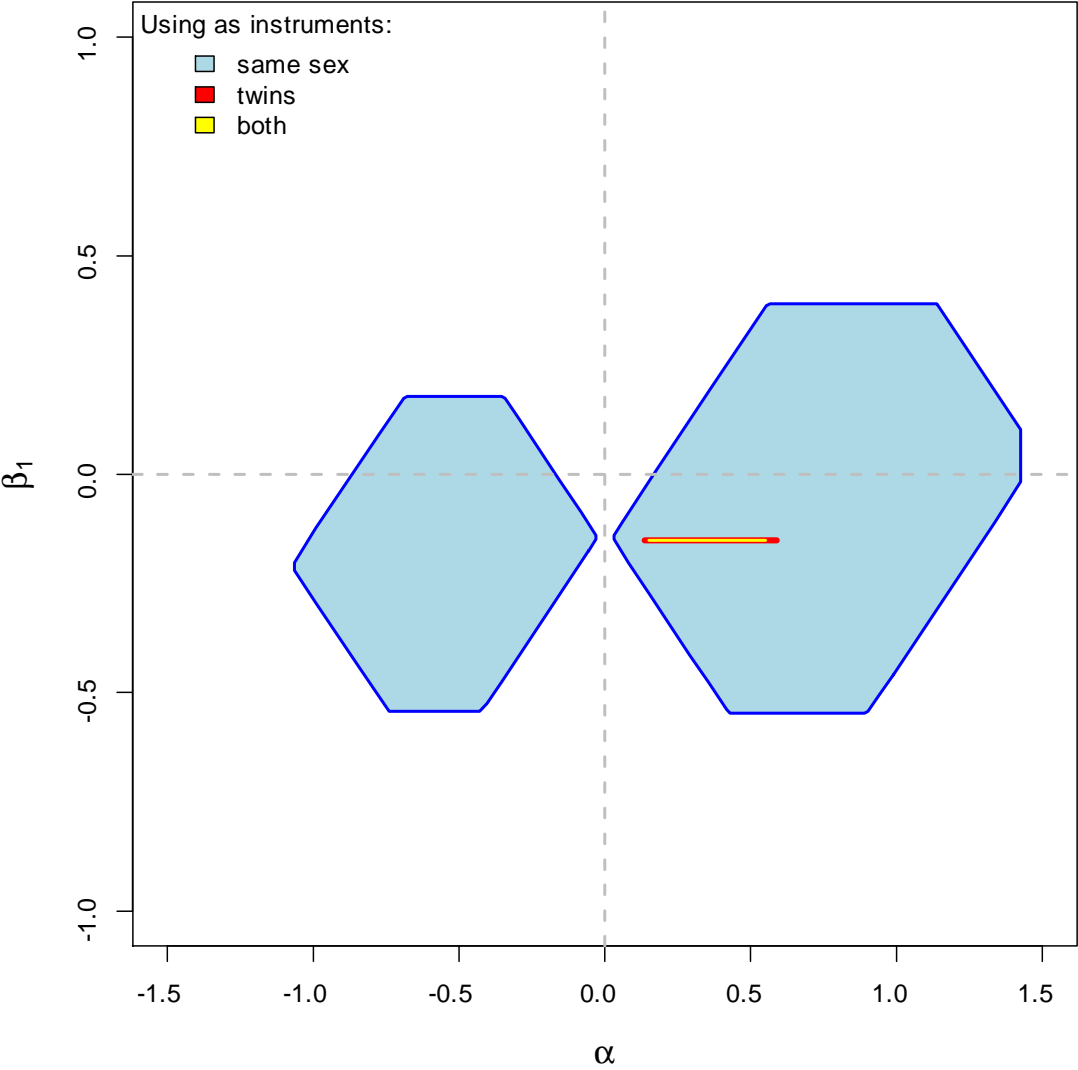
Figure 10: Analog estimates of projections of identified sets onto the space of β_0 and β_1. Identified sets using the twins instruments and both instruments are lines, drawn here with nonnegligible thickness so that it is possible to see the reduction in the size of the projection on using the same sex instrument as well as the twins instrument.
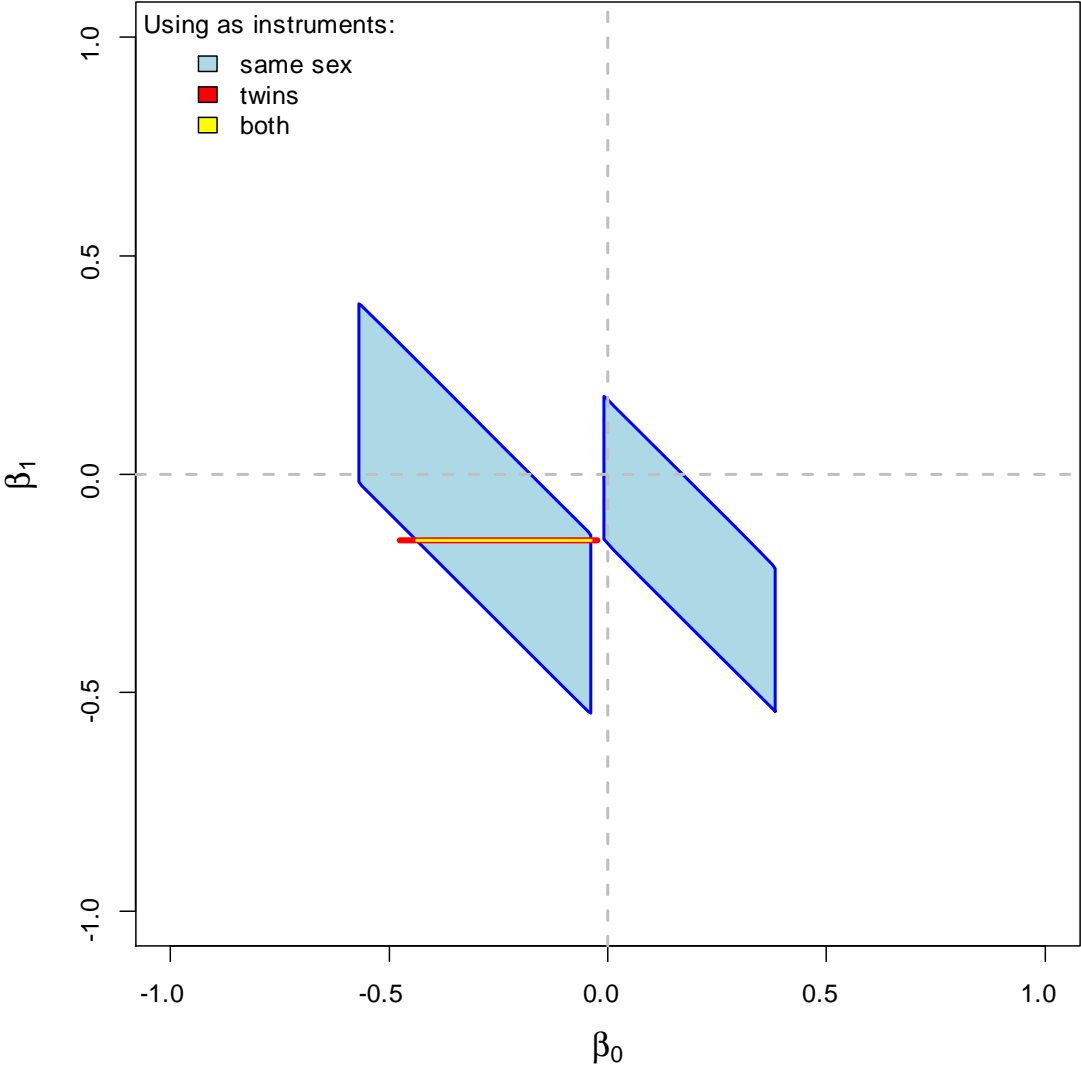
Figure 11: Analog estimates of projections of identified sets onto the space of α and β_0. Identified sets using the twins instruments and both instruments are lines, drawn here with nonnegligible thickness so that it is possible to see the reduction in the size of the projection on using the same sex instrument as well as the twins instrument.