cemmap
centre for microdata methods and practice

Generalized instrumental variable models

Andrew Chesher
Adam Rosen

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP04/14

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

An ESRC Research Centre

Institute for Fiscal Studies, 7 Ridgmount Street, London WC1E 7AE **Tel:** 44 (0)20 7291 4800 **Email:** info@cemmap.ac.uk **www.cemmap.ac.uk**

Generalized Instrumental Variable Models*

Andrew Chesher†
UCL and CeMMAP

Adam M. Rosen‡
UCL and CeMMAP

January 20, 2014

Abstract

The ability to allow for flexible forms of unobserved heterogeneity is an essential ingredient in modern microeconometrics. In this paper we extend the application of instrumental variable (IV) methods to a wide class of problems in which multiple values of unobservable variables can be associated with particular combinations of observed endogenous and exogenous variables. In our Generalized Instrumental Variable (GIV) models, in contrast to traditional IV models, the mapping from unobserved heterogeneity to endogenous variables need not admit a unique inverse. The class of GIV models allows unobservables to be multivariate and to enter non-separably into the determination of endogenous variables, thereby removing strong practical limitations on the role of unobserved heterogeneity. Important examples include models with discrete or mixed continuous/discrete outcomes and continuous unobservables, and models with excess heterogeneity where many combinations of different values of multiple unobserved variables, such as random coefficients, can deliver the same realizations of outcomes. We use tools from random set theory to study identification in such models and provide a sharp characterization of the identified set of structures admitted. We demonstrate the application of our analysis to a continuous outcome model with an interval-censored endogenous explanatory variable.

Keywords: instrumental variables, endogeneity, excess heterogeneity, limited information, set identification, partial identification, random sets, incomplete models.

JEL classification: C10, C14, C24, C26.

---

*We thank Federico Bugni and Matt Masten for detailed comments. We have benefitted from discussion with participants at seminar and conference presentations given at Boston College, Chicago, Paris School of Economics, Vanderbilt, the April 2013 conference on Mathematical Statistics of Partially Identified Objects at Oberwolfach, the 2013 Cowles summer econometrics conference, Yale, Virginia, Georgetown, Maryland, Johns Hopkins, Toulouse School of Economics, Oxford, Duke, and Illinois. Financial support from the UK Economic and Social Research Council through a grant (RES-589-28-0001) to the ESRC Centre for Microdata Methods and Practice (CeMMAP) and from the European Research Council (ERC) grant ERC-2009-StG-240910-ROMETA is gratefully acknowledged.

†Address: Andrew Chesher, Department of Economics, University College London, Gower Street, London WC1E 6BT, andrew.chesher@ucl.ac.uk.

‡Address: Adam Rosen, Department of Economics, University College London, Gower Street, London WC1E 6BT, adam.rosen@ucl.ac.uk.

1 Introduction

In this paper we extend the application of instrumental variable (IV) methods to a wide class of models in which multiple values of unobservable variables can be associated with particular combinations of observed endogenous and exogenous variables. Leading examples are models with high dimensional heterogeneity such as random coefficient models, models with discrete outcomes, and models in which observable and unobservable variables are constrained by inequality restrictions.

The Generalized Instrumental Variable (GIV) models we study embody restrictions that (i) constrain the influence of some exogenous variables (instruments) on the determination of outcomes and (ii) limit the covariation of instruments and unobserved variables. Like Blundell and Powell (2003), we use the term "IV model" to refer to models in which the determination of some of the endogenous variables as a function of exogenous variables and unobservables is left unspecified. While we focus mainly on these sorts of models, our methods and results also apply when a more complete specification is given as in the triangular models studied in Chesher (2003), Vytlacil and Yildiz (2007), and Imbens and Newey (2009) and the simultaneous equations models studied in Matzkin (2008) and Chesher and Rosen (2012b).

We provide sharp characterizations of identified sets for structures and structural features delivered by GIV models. We show that these sets can be expressed as systems of moment inequalities to which recently developed inferential procedures are applicable. We give conditions under which some or all of the inequalities are moment equalities. Our characterizations can be refined to obtain point identification in previously studied instrumental variable models, such as those of Koopmans, Rubin, and Leipnik (1950) and Chernozhukov and Hansen (2005).

A distinguishing feature of the models we study is that they admit a residual *correspondence* ρ such that

$$U \in \rho(Y, Z), \tag{1.1}$$

where Y, Z denote observed endogenous and exogenous variables, respectively, and where U denotes unobserved heterogeneity. In many GIV models the endogenous variables $Y \equiv (Y_1, Y_2)$ include variables Y_1 modeled as the outcome of an economic process and variables Y_2 that enter into their determination, i.e. "right-hand-side" endogenous variables which are allowed to be stochastically dependent with unobserved heterogeneity. The exogenous variables $Z \equiv (Z_1, Z_2)$ are restricted in the degree to which they are stochastically dependent with U. Variation in variables Z_1, sometimes referred to as "exogenous covariates", may affect $\rho(Y, Z)$, while variables Z_2 are excluded instruments with respect to which $\rho(Y, Z)$ is restricted to be invariant.

The residual correspondence is defined through a structural relation given by

$$h(Y, Z, U) = 0, \tag{1.2}$$

relating values of observed and unobserved variables. Then $\rho(Y, Z)$ is precisely the set of values of U such that (1.2) holds:

$$\rho(Y, Z) = \mathcal{U}(Y, Z; h) \equiv \{u \in \mathcal{R}_U : (1.2) \text{ holds when } U = u.\}, \tag{1.3}$$

where $\mathcal{R}_U$ denotes the support of U.

By contrast, IV models require that the residual $\rho(Y, Z)$ be unique, so that

$$U = \rho(Y, Z), \tag{1.4}$$

for some *function* $\rho(Y, Z)$. This requires that the structural relation (1.2) produces a unique value of U for almost every realization of (Y, Z). The Generalized Instrumental Variable (GIV) models that give rise to only (1.1) are typically partially identifying although they include as special cases point identifying models, such as classical IV models, for which the solution to (1.2) in U is guaranteed unique. We provide a sharp characterization of the identified set of structures delivered by GIV models, and examples of the sets obtained in particular cases.

The extension of IV methods to models with (1.1) allows for unobservables to be multivariate and to enter nonseparably into the determination of Y. This can be important in practice, as failure to allow for these possibilities places strong limitations on the role of unobserved heterogeneity, ruling out for example random coefficients and discrete-valued outcomes. Restrictions on the structural relation h that guarantee the existence of a residual function as in (1.4) include additive separability in U, or strict monotonicity of h in scalar U, so that h has a unique inverse. Important cases in which invertibility fails and where GIV models deliver new results include models with discrete or mixed continuous/discrete outcomes and continuous unobservables, and models with excess heterogeneity where many combinations of values of multiple unobserved variables, such as random coefficients, can deliver the same realizations of outcomes. Even in fully parametric non-linear models invertibility can fail, and there may be only set identification.[1,2]

In this paper we relax this invertibility restriction, thereby substantially increasing the range of problems to which IV models can be applied. In the previous papers Chesher (2010), Chesher and Smolinski (2012), Chesher, Rosen, and Smolinski (2013), Chesher and Rosen (2012, 2013a, 2013b) we have given some results for particular cases in which outcomes are discrete. Here we present a complete development and results for a general class which includes problems in which outcomes may be continuous or discrete. Unlike our previous analyses, we consider conditional

[1] Examples of parametric non-linear models where there is a residual correspondence as in (1.1) but not a residual function as in (1.4), include those of Ciliberto and Tamer (2009) and Section 4 of Chesher, Rosen, and Smolinski (2013).

[2] These issues can sometimes be alleviated by imposing additional structure on the determination of endogenous variables as a function of exogenous variables Z, for example by specifying an equation restricting how variation in Z_2 induces variation in Y_2. Such restrictions are helpful from an identification standpoint, but may not always be credible. We do not require such restrictions, although they can be incorporated into our setup.

mean and conditional quantile restrictions on unobservables given exogenous variables, in addition to stochastic independence restrictions. We also provide a novel result allowing for characterization of the identified set for structural function h when U and Z are independent, but the distribution of U is completely unrestricted.

A simple example of an econometric model where there is invertibility and that can be written in the more restrictive form (1.4) is the linear model with a single endogenous covariate, where $\rho(Y,Z) = Y_1 - Y_2\beta$. The instrument Z is excluded from the structural relation and hence from ρ, and is assumed exogenous, for example through the conditional mean restriction $E[U|Z] = 0$. Given this, the parameter β, and hence the distribution of U, are identified under the classical rank condition set out by Koopmans, Rubin, and Leipnik (1950).

Many model specifications, both linear and non-linear, as well as parametric, semiparametric, and nonparametric, yield residual *functions* of the form (1.4). The recent literature on nonparametric IV models achieves identification from (1.4) in conjunction with *completeness* conditions, as in Newey and Powell (2003), Chernozhukov and Hansen (2005), Hall and Horowitz (2005), Chernozhukov, Imbens, and Newey (2007), Blundell, Chen, and Kristensen (2007), Darolles, Fan, Florens, and Renault (2011), and Chen, Chernozhukov, Lee, and Newey (2011). These completeness conditions on the conditional distribution of endogenous covariates Y_2 given exogenous variables Z can be viewed as nonparametric analogs of the classical rank condition, and have been a central focus in recent papers by D'Haultfoeuille (2011), Andrews (2011), and Canay, Santos, and Shaikh (2013).

The literature on nonparametric IV models has significantly advanced applied researchers' ability to deal with endogeneity beyond the classical linear setup. Although many models produce residual functions of the form (1.4), the requirement of a unique U associated with observed (Y,Z) is not innocuous. When there is no such residual function the usual derivation of identification from completeness conditions does not go through. The requirement that U belong to some residual correspondence as in (1.1) is more widely applicable, but completeness conditions are generally insufficient for point identification, even if they are thought to hold.

A central object in our identification analysis is the random set $\mathcal{U}(Y,Z;h)$ defined in (1.3). We use random set theory methods reviewed in Molchanov (2005) and introduced into econometric identification analysis by Beresteanu, Molchanov, and Molinari (2011), henceforth BMM11, with particular use of a result known as Artstein's (1983) Inequality. We extend our analysis employing the distribution of random sets in the space of unobserved heterogeneity as in for example Chesher, Rosen, and Smolinski (2013) and Chesher and Rosen (2012b) to the much more general class of GIV models considered here. Our characterizations of identified sets are constructive in that they can be expressed as systems of conditional moment inequalities that can used as a basis for estimation and inference, see for example Chesher and Rosen (2013a,b), and Aradillas-Lopez and Rosen (2013) for empirical applications using treatment effect and simultaneous ordered response models. More generally a large number of methods from the recent literature on inference based on conditional and unconditional moment inequalities may be applicable, as we discuss further in the conclusion.

The paper proceeds as follows. In Section 2 we provide an informal overview of our results and present some leading examples of GIV models to which our analysis applies. In Section 3 we lay out the formal restrictions of GIV models and provide identification analysis. This includes a new generalization of the classical notion of observational equivalence, e.g. Koopmans (1949), Koopmans and Reiersøl (1950), Hurwicz (1950), Rothenberg (1971), and Bowden (1973) to models where a structure need not generate a unique distribution of outcomes given other observed variables.[3] We use the notion of selectionability from random set theory, and in particular show how it can be applied in the space of unobserved heterogeneity to incorporate restrictions on unobserved variables of the sort commonly used in econometric models. We demonstrate how this is done for models invoking stochastic independence, conditional mean, and conditional quantile restrictions.

In Section 4 we illustrate the set identifying power of GIV models through application of our Theorems in Section 3 to a continuous outcome model with an interval-censored endogenous explanatory variable. We provide conditional moment inequality characterizations of identified sets under each of the alternative restrictions on unobserved heterogeneity considered in Section 3.4, namely mean independence, quantile independence, and stochastic independence both with and without parametric distributional restrictions. Interval censoring is common in practice, and our treatment of this example demonstrates the application of our results to produce novel set identification characterizations in a setting of practical interest that has not been previously considered in the literature. Prior research invoking random set theory and other methods for characterizing identified sets have not been applied to this continuous outcome model, and do not appear applicable without significant modification.

All proofs are provided in the Appendix. Section 5 concludes.

Notation: We use capital Roman letters A to denote random variables and lower case letters a to denote particular realizations. For probability measure $\mathbb{P}$, $\mathbb{P}(\cdot|a)$ is used to denote the conditional probability measure given $A = a$. We write $\mathcal{R}_A$ to denote the support of random vector A, and $\mathcal{R}_{A_1 \cdots A_m}$ to denote the joint support of random vectors $A_1, ..., A_m$. $\mathcal{R}_{A_1|a_2}$ denotes the support of random vector A_1 conditional on $A_2 = a_2$. $q_{A|B}(\tau|b)$ denotes the τ conditional quantile of A given $B = b$. $A \perp\!\!\!\perp B$ means that random vectors A and B are stochastically independent. $\emptyset$ denotes the empty set. Script font ($\mathcal{S}$) is reserved for sets, and sans serif font (S) is reserved for collections of sets. The sign $\subseteq$ is used to indicate nonstrict inclusion so "$\mathcal{A} \subseteq \mathcal{B}$" includes $\mathcal{A} = \mathcal{B}$, while "$\mathcal{A} \subset \mathcal{B}$" means $\mathcal{A} \subseteq \mathcal{B}$ but $\mathcal{A} \neq \mathcal{B}$. $\mathrm{cl}(\mathcal{A})$ denotes the closure of $\mathcal{A}$. $C_h(\mathcal{S}|z)$ denotes the containment functional of random set $\mathcal{U}(Y, Z; h)$ conditional on $Z = z$, defined in Section 3.2. The notation $F \lesssim \mathcal{A}$ is used to indicate that the distribution F of a random vector is *selectionable* with respect

[3] Hurwicz (1950) does not explicitly use the term *observational equivalence* but employs the concept nonetheless, writing for instance on page 248 that, "In general, more than one structure generating G can be found," where G denotes the distribution of observed variables. Such structures are observationally equivalent.

to the distribution of random set $\mathcal{A}$, as defined in Section 3.1. $1[\mathcal{E}]$ denotes the indicator function, taking the value 1 if the event $\mathcal{E}$ occurs and 0 otherwise. For any real number c, $|c|_+$ and $|c|_-$ denote the positive and negative parts of c, respectively.[4] $\mathbb{R}^m$ denotes m dimensional Euclidean space, and for any vector $v \in \mathbb{R}^m$, $\|v\|$ indicates the Euclidean norm: $\|v\| = \sqrt{v_1^2 + \cdots + v_m^2}$. Finally, in order to deal with sets of measure zero and conditions required to hold almost everywhere, we use the "sup" and "inf" operators to denote "essential supremum" and "essential infimum" with respect to the underlying measure when these operators are applied to functions of random variables (e.g. conditional probabilities, expectations, or quantiles). Thus $\sup_{z\in\mathcal{Z}} f(z)$ denotes the smallest value of $c \in \mathbb{R}$ such that $\mathbb{P}[f(Z) > c] = 0$ and $\inf_{z\in\mathcal{Z}} f(z)$ denotes the largest value of $c \in \mathbb{R}$ such that $\mathbb{P}[f(Z) < c] = 0$.

2 Examples and Informal Overview

The structural function $h : \mathcal{R}_{YZU} \to \mathbb{R}$ such that (1.2) holds with probability one determines the feasible values of endogenous variables Y when $(Z, U) = (z, u)$. We define the zero level sets of h for each $(y, z) \in \mathcal{R}_{YZ}$ and $(z, u) \in \mathcal{R}_{ZU}$, respectively as

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

These level sets are dual to each other in that for all z and all functions h, a value u^* lies in $\mathcal{U}(y^*, z; h)$ if and only if y^* lies in $\mathcal{Y}(u^*, z; h)$. Duality will be exploited to good effect.

In contrast to conventional IV models, GIV models admit structural functions h for which the level sets $\mathcal{U}(y, z; h)$ may have cardinality *exceeding* one. However, in both IV and GIV models $\mathcal{Y}(u, z; h)$ may have cardinality greater than one, in which case the model is incomplete. When this happens there are values y and y' such that the level sets $\mathcal{U}(y, z; h)$ and $\mathcal{U}(y', z; h)$ have non-empty intersection. This occurs in particular when the model does not specify the way in which endogenous explanatory variables Y_2 are determined, even if Y_1 is uniquely determined by (Y_2, Z, U). Some leading examples follow. Each of these may be combined with alternative restrictions on the joint distribution of (U, Z), for example $U \perp\!\!\!\perp Z$, $E[U|Z] = 0$, $q_{U|Z}(\tau|Z) = 0$, and/or parametric restrictions on the distribution of U.

2.1 Examples

Example 1. *A classical linear IV model with an instrument exclusion restriction has structural function*

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

[4] $|c|_- = -\min(c, 0)$, $|c|_+ = \max(c, 0)$.

in which case $\mathcal{Y}(u,z;h) = \{((\alpha+\beta y_2+u), y_2) : y_2 \in \mathcal{R}_{Y_2}\}$. *The level set* $\mathcal{U}(y,z;h)$ *is the singleton set* $\{(y_1-\alpha-\beta y_2)\}$. *In this instance the realization of exogenous variables* z *does not enter into* h *and so* $z = z_2$.

Example 2. *A binary outcome, threshold crossing GIV model with* $Y_1 = 1[g(Y_2,Z_1) < U]$ *and* U *normalized uniformly distributed on* $[0,1]$, *as studied in Chesher (2010) and Chesher and Rosen (2013b), has structural function*

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

where $y_1 \in \{0,1\}$.[5] *The corresponding level sets are pairs of values of* (y_1,y_2),

$$\mathcal{Y}(u,z;h) = \{(y_1,y_2) \in \mathcal{R}_{Y_1Y_2} : \{y_1 = 1 \wedge u \geq g(y_2,z_1)\} \text{ or } \{y_1 = 0 \wedge u \leq g(y_2,z_1)\}\},$$

and intervals

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

Example 3. *Multiple discrete choice with endogenous explanatory variables as studied in Chesher, Rosen, and Smolinski (2013). The structural function is*

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

where $\pi_j(y_2,z_1,u_j)$ *is the utility associated with choice* $j \in \mathcal{J} \equiv \{1,\ldots,M\}$ *and* $u = (u_1,\ldots,u_M)$ *is a vector of unobserved preference heterogeneity.* Y_1 *is the outcome or choice variable and* Y_2 *are endogenous explanatory variables.* Z_1 *are exogenous variables allowed to enter the utility functions* $\pi_1,\ldots,\pi_M$, *while* Z_2 *are excluded exogenous variables, or instruments. The level sets are thus*

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

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 endogeneity has structural function*

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

[5] With U 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. For simplicity of exposition we define our function h such that either value of Y_1 is permitted when $g(Y_2,Z_1) = U$.

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 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

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

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 increasing in its first argument and strictly increasing 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*

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

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*

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

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

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

and

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

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*,*

$$g^{-1}(y_2, z_1, g(y_2, z_1, u)) = u.$$

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* M *bidders making non-negative final bids* $Y = (Y_1, \ldots, Y_M)$*, there are continuously distributed non-negative valuations* $U = (U_1, \ldots, U_M)$ *and* U *has probability distribution* $G_{U|Z}(\cdot|z)$ *conditional on auction characteristics* $Z = z$*. Let* $a_{[m]}$ *denote the* m*th largest element of a vector* $(a_1, \cdots, a_M)$*. The structural function and resulting level sets* *are as follows.*

$$h(y,z,u) = \sum_{m=1}^{M} |y_m - u_m|_+ + \left|y_{[1]} - u_{[2]}\right|_-$$

$$\mathcal{Y}(u,z;h) = \left\{ y \in \mathcal{R}_Y : \bigwedge_{m=1}^{M} (y_m \leq u_m) \wedge \left(y_{[1]} \geq u_{[2]}\right) \right\}$$

$$\mathcal{U}(y,z;h) = \left\{ u \in \mathcal{R}_U : \bigwedge_{m=1}^{M} (y_m \leq u_m) \wedge \left(y_{[1]} \geq u_{[2]}\right) \right\}$$

The structural function h is known and does not depend on z. The unknown object of interest is $\mathcal{G}_{U|Z} = \{G_{U|Z}(\cdot|z) : z \in \mathcal{R}_Z\}$. 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.

GIV models have numerous applications and many other incomplete models that include restrictions on the distribution of unobserved variables and exogenous covariates fall in this class. Examples include models with structural relationships defined by a system of inequalities such as Example 6 above drawing on Haile and Tamer (2003), simultaneous equations models admitting multiple solutions such as the oligopoly entry game of Bresnehan and Reiss (1989, 1991) and Tamer (2003), and models with a nonadditive scalar unobservable and no monotonicity restriction as studied by Hoderlein and Mammen (2007).

2.2 The approach and the main results

A model $\mathcal{M}$ comprises a collection of admissible structures. A structure is a pair $\left(h, \mathcal{G}_{U|Z}\right)$ where h is a structural function previously defined and $\mathcal{G}_{U|Z}$ is shorthand for the collection of conditional distributions,

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

where $G_{U|Z}(\mathcal{S}|z)$ denotes the probability mass placed on any set $\mathcal{S} \subseteq \mathcal{R}_{U|z}$ when $U \sim G_{U|Z}(\cdot|z)$. The model $\mathcal{M}$ can restrict both the structural function h and the distributions $\mathcal{G}_{U|Z}$. For example $\mathcal{M}$ may admit structures $\left(h, \mathcal{G}_{U|Z}\right)$ such that h is nonparametrically specified and sufficiently smooth and collections $G_{U|Z}$ of unknown form but such that U and Z are independently distributed. Alternatively the model may restrict h to be known up to a finite dimensional parameter vector, or require for example that $E[U|Z] = 0$, or that $G_{U|Z}(\cdot|z)$ be parametrically specified.

The sampling process identifies a collection of probability distributions denoted

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

where $F_{Y|Z}(\mathcal{T}|z)$ denotes the probability mass placed on any set $\mathcal{T} \subseteq \mathcal{R}_{Y|z}$. Our identification analysis answers the question: precisely which admissible structures $\left(h, \mathcal{G}_{U|Z}\right)$ are capable of generating the collection of conditional distributions $\mathcal{F}_{Y|Z}$?

We begin by revisiting the notion of observational equivalence. Classical definitions of observational equivalence, e.g. those of Koopmans (1949), Koopmans and Reiersøl (1950), Rothenberg (1971), and Bowden (1973) require that a given structure produce for each $z \in \mathcal{R}_Z$ a unique conditional distribution of endogenous variables $F_{Y|Z}(\cdot|z)$. Thus, to employ the classical definition of observational equivalence of structures admitted by an incomplete model (that is one that can give rise to multiple conditional distributions $F_{Y|Z}(\cdot|z)$), one must augment the model with an additional component in order to complete it, and then apply the classical notion of observational equivalence to the augmented model. As shown by BMM11, the identified set of incomplete structures for the original incomplete model are precisely those structures that can be augmented (with an equilibrium selection mechanism in their analysis) to obtain a complete structure that generates the observed conditional distributions in the data. All such complete structures are thus observationally equivalent.

In our framework a structure $\left(h, \mathcal{G}_{U|Z}\right)$ produces a collection of random sets

$$\left\{\mathcal{Y}(U, z; h) : U \sim G_{U|Z}(\cdot|z), z \in \mathcal{R}_Z\right\}.$$

Conditional on $Z = z$, the random set $\mathcal{Y}(U, z; h)$ can be viewed as a collection of its selections, where the selection of a random set $\mathcal{Y}$ is any random variable Y such that $\Pr[Y \in \mathcal{Y}] = 1$. Thus a structure $\left(h, \mathcal{G}_{U|Z}\right)$ admitted by our model generally produces collections of compatible distributions for $F_{Y|Z}(\cdot|z)$.

We generalize the classical notion of observational equivalence to allow incomplete structures that permit collections of conditional distributions $F_{Y|Z}(\cdot|z)$, each $z \in \mathcal{R}_Z$. We use this more general definition to establish duality between observational equivalence expressed in terms of random sets in the space of outcomes Y and the space of unobserved heterogeneity U, as expressed in Theorem 1 and employed in Theorem 2. Our generalization of observational equivalence accords with the logic underlying previous identification analyses in incomplete models, while introducing some interesting subtleties. For example, conditional on any $Z = z$, the collections of conditional distributions produced by two structures may be different yet overlap, so that whether or not the two structures are observationally distinct depends on the identified conditional distributions $\mathcal{F}_{Y|Z}$. This cannot happen when structures produce unique collections $\mathcal{F}_{Y|Z}$. The identified set of structures $\left(h, \mathcal{G}_{U|Z}\right)$ are then precisely those such that $F_{Y|Z}(\cdot|z)$ is *selectionable* with respect to the distribution of $\mathcal{Y}(U, z; h)$ given $Z = z$, for almost every $z \in \mathcal{R}_Z$.[6] This set may be empty, allowing for the possibility that model $\mathcal{M}$ of Restriction A4 is misspecified.

In Theorem 1 of Section 3, we show that because of the dual relationship between the two level

[6] A distribution F of a random variable is said to be selectionable with respect to the distribution of random set $\mathcal{Y}$ if there exists a random variable, say W, with distribution F, and a random set, say $\tilde{\mathcal{Y}}$, with the same distribution as $\mathcal{Y}$, such that $W \in \tilde{\mathcal{Y}}$ with probability one. See Definition 1, Section 3.1 for the formal definition.

sets of the structural function h, $\mathcal{U}(y, z; h)$ and $\mathcal{Y}(u, z; h)$, the probability distribution $F_{Y|Z}(\cdot|z)$ is selectionable with respect to the distribution of $\mathcal{Y}(U, z; h)$ when $U \sim G_{U|Z}(\cdot|z)$ if and only if $G_{U|Z}(\cdot|z)$ is selectionable with respect to the distribution of the random set $\mathcal{U}(Y, z; h)$ when $Y \sim F_{Y|Z}(\cdot|z)$. Using this result we show that we can completely characterize observational equivalence, and hence identified sets of model structures, through selectionability of $G_{U|Z}(\cdot|z)$ with respect to the conditional distribution of $\mathcal{U}(Y, z; h)$. This result, formalized in Theorem 2, facilitates the imposition of restrictions directly on the joint distribution of U and Z, i.e. restrictions on the distribution of unobserved heterogeneity that are common in econometric modeling.

The selectionability criteria provide a widely applicable characterization of the identified set of structures $\left(h, \mathcal{G}_{U|Z}\right)$. Selectionability itself can be characterized in a variety of ways, and depending on which restrictions are imposed on the conditional distributions of unobserved heterogeneity $\mathcal{G}_{U|Z}$, different characterizations may prove more or less convenient. One such way is through the use of the containment functional of the random set $\mathcal{U}(Y, Z; h)$ defined in (1.3), as we show at the end of Section 3.2 in Corollary 1. This produces a generally applicable characterization comprising a collection of conditional moment inequalities. The collection of implied moment inequalities is potentially extremely large. In Section 3.3 we employ the notion of core-determining test sets to exploit the underlying geometry of possible realizations of random sets $\mathcal{U}(Y, Z; h)$ and thereby produce a reduction in the number of inequality restrictions necessary to characterize the identified set. We refine the containment functional characterization of the identified set by reducing the *number* of conditional moment restrictions required, and also by providing conditions under which some of these inequalities must in fact hold with *equality.* An implication is that models that admit only singleton residual sets $\mathcal{U}(Y, Z; h)$ or only singleton outcome sets $\mathcal{Y}(U, z; h)$ have identified sets characterized entirely by conditional moment *equalities.*

With sufficient variation in exogenous variables Z, as reflected in familiar rank and large support conditions in the econometrics literature, conditional moment equalities and inequalities may result in singleton identified sets, equivalently point identification.[7] We allow for set identification and therefore do not require such conditions. Nonetheless, our identified sets reduce to singleton points if there is sufficient variation in the distribution of observed variables for point identification to hold.

Up to this point, our identification analysis holds no matter what restrictions are imposed on $\mathcal{G}_{U|Z}$. In Section 3.4 we consider three types of restrictions on the joint distribution of U and Z common to the econometrics literature, namely stochastic independence, mean independence, and quantile independence restrictions. Under stochastic independence we show how the containment functional characterization of Corollary 1 simplifies. We provide characterizations allowing for the distribution of unobserved heterogeneity to be either restricted to some known family, or completely

[7]Some examples of sufficient conditions for point identification stemming from conditional moment inequalities can be found in for instance Manski (1985), Tamer (2003), Khan and Tamer (2009), and Rosen (2012).

unrestricted. With only mean independence imposed rather than stochastic independence, we employ the Aumann expectation to characterize the identified set, and show how, when applicable, the support function of random set $\mathcal{U}(Y, Z; h)$ can be used for computational gains. The ideas here follow closely those of BMM11, although unlike their analysis we continue to consider characterizations through random sets in $\mathcal{R}_U$. We then provide a characterization of identified sets in models with conditional quantile restrictions and interval-valued random sets $\mathcal{U}(Y, Z; h)$, such as those used in models with censored variables. In Section 4 we illustrate the identifying power of each of these restrictions on $\mathcal{G}_{U|Z}$ in the context of Example 5 above, featuring a censored endogenous explanatory variable.

3 Identified sets for GIV models

We impose the following restrictions throughout.
Restriction A1: (Y, Z, U) are random vectors defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, endowed with the Borel sets on Ω. The support of (Y, Z, U) is a subset of Euclidean space. $\square$
Restriction A2: A collection of conditional distributions

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

is identified by the sampling process, where for all $\mathcal{T} \subseteq \mathcal{R}_{Y|z}$, $F_{Y|Z}(\mathcal{T}|z) \equiv \mathbb{P}[Y \in \mathcal{T}|z]$. $\square$
Restriction A3: There is an $\mathcal{F}$-measurable function $h(\cdot, \cdot, \cdot) : \mathcal{R}_{YZU} \to \mathbb{R}$ such that

$$\mathbb{P}[h(Y, Z, U) = 0] = 1,$$

and there is a collection of conditional distributions

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

where for all $\mathcal{S} \subseteq \mathcal{R}_{U|z}$, $G_{U|Z}(\mathcal{S}|z) \equiv \mathbb{P}[U \in \mathcal{S}|z]$ denotes a conditional distribution of U given $Z = z$. $\square$
Restriction A4: The pair $\left(h, \mathcal{G}_{U|Z}\right)$ belongs to a known set of admissible structures $\mathcal{M}$ such that $h(\cdot, \cdot, \cdot) : \mathcal{R}_{YZU} \to \mathbb{R}$ is continuous in its first and third arguments. $\square$

Restriction A1 defines the probability space on which (Y, Z, U) reside and restricts their support to Euclidean space.[8] Restriction A2 requires that, for each $z \in \mathcal{R}_Z$, the conditional distribution of Y given $Z = z$ is identified. Random sampling of observations from $\mathbb{P}$ is sufficient, but not required.

[8]The restriction of the support of unobserved heterogeneity to a subset of Euclidean space is convenient, but not required for our identification analysis. What is essential for our use of random set theory is that the support of U is a locally compact Hausdorff second countable topological space. Euclidean space fulfills this requirement. For further details we refer to Molchanov (2005).

Restriction A3 posits the existence of structural relation h and provides notation for the collection of conditional distributions $\mathcal{G}_{U|Z}$ of U given Z induced by population measure $\mathbb{P}$. Restrictions A1-A3 place restrictions on neither the properties of h nor $\mathcal{G}_{U|Z}$. These restrictions are maintained throughout.

Restriction A4 imposes model $\mathcal{M}$, the collection of admissible structures $\left(h, \mathcal{G}_{U|Z}\right)$. Unlike restrictions A1-A3, restriction A4 is refutable based on knowledge of $\mathcal{F}_{Y|Z}$. Our 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 possibly deliver the identified conditional distributions $\mathcal{F}_{Y|Z}$. It is possible of course 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 $\mathcal{G}_{U|Z}$. We allow for this possibility, noting that in such cases the identified set of structures delivered by our results is empty, which would indicate that the model is misspecified. Continuity of $h\left(y,z,u\right)$ in y and u guarantees that $\mathcal{Y}\left(U,Z;h\right)$ and $\mathcal{U}\left(Y,Z;h\right)$ are random closed sets, but can be relaxed.[9]

In places we will find it convenient to refer separately to collections of admissible structural functions and distributions $\mathcal{G}_{U|Z}$. Notationally these are defined as the following projections of $\mathcal{M}$.

$$
\begin{aligned}
\mathcal{H} &\equiv \left\{h : \left(h, \mathcal{G}_{U|Z}\right) \in \mathcal{M} \text{ for some } \mathcal{G}_{U|Z}\right\}, \\
\mathsf{G}_{U|Z} &\equiv \left\{\mathcal{G}_{U|Z} : \left(h, \mathcal{G}_{U|Z}\right) \in \mathcal{M} \text{ for some } h\right\}.
\end{aligned}
$$

The model $\mathcal{M}$ could, but does not necessarily, consist of the full product space $\mathcal{H} \times \mathsf{G}_{U|Z}$.

Restrictions A1-A4 are noticeably weak. This is intentional, and not without consequence. The identification analysis built up in Sections 3.1 and 3.2 below is set out with only these restrictions in place, and is thus extremely general. The analysis characterizes the identified set of model structures, and the level of generality allows for the possibility that these sets are either large or small, for example the entire admissible space at one extreme, or a singleton point at the other.

The identifying power of any particular model manifests through three different mechanisms: (i) restrictions on the class of functions h; (ii) restrictions on the joint distribution of (U,Z), embodied through the admitted collections $\mathcal{G}_{U|Z}$, and (iii) the joint distribution of (Y,Z). The first two mechanisms are part of the model specification, with $\mathcal{M}$ constituting the set of structures $\left(h, \mathcal{G}_{U|Z}\right)$ deemed admissible for the generation of (Y,Z). A researcher may restrict these to belong to more or less restrictive classes, parametric, semiparametric, or nonparametric. For example, h could be allowed to be any function satisfying some particular smoothness restrictions, or it could be restricted to a parametric family as in a linear index model. Likewise $\mathcal{G}_{U|Z}$ could be collections of conditional distributions such that $E\left[U|Z\right] = 0$, or $q_{U|Z}\left(\tau|z\right) = 0$, or $U \perp\!\!\!\perp Z$, or $U \perp\!\!\!\perp Z$ with $G_{U|Z}\left(\cdot|z\right)$ restricted to a parametric family. We consider various types of restrictions on $\mathcal{G}_{U|Z}$ in

[9]Note that continuity of the function h does not rule out either Y or U being discrete-valued. Continuity itself is not essential, and may be relaxed as long as the relevant random sets can be shown to be closed in some topology.

Section 3.4 below.

The third source of identifying power, the joint distribution of (Y, Z), is identified and hence left completely unrestricted. This may appear at odds with results requiring rank or completeness conditions, but in fact is not. Such conditions are used to ensure point identification. We allow for set identification, and so these are not required. Rather, we provide characterizations of identified sets in the class of models we study. If there is "sufficient variation" in the distribution of (Y, Z) to achieve point identification, such as the usual rank condition in a linear IV model, our characterization reduces to a singleton set. Thus, such conditions are not required in our set identification analysis, but are still very much of interest in consideration of which qualities of observed data may result in identified sets that are singleton points, or more generally affect the size of the identified set.

More broadly, the generality of our identification analysis offers a formal framework for the consideration of which models $\mathcal{M}$ and qualities of the joint distribution of (Y, Z) may be usefully applied. Given a joint distribution for (Y, Z), less restrictive models will necessarily result in larger identified sets than will more restrictive models. A sufficiently unrestrictive model may have so little identifying power as to be uninformative when used in practice, or even have no identifying power all. This is useful for researchers to know when considering which models to employ in practice, and can be used to motivate the incorporation of further restrictions that may be thought credible.

3.1 Observational Equivalence and Selectionability in Outcome Space

The standard definition of observational equivalence found in the econometrics literature applies in contexts in which each structure, m, delivers a single collection of conditional distributions, which we write as:

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

where $P_{Y|Z}(\cdot|z; m)$ is the conditional distribution of Y given $Z = z$ delivered by structure m.[10]

Under this definition of observational equivalence, structures m and m' are observationally equivalent if they deliver the same collection of conditional distributions, that is if $\mathcal{P}_{Y|Z}(m) = \mathcal{P}_{Y|Z}(m')$. The particular identified set of conditional distributions, $\mathcal{F}_{Y|Z}$, that is under consideration in identification analysis plays no role in determining whether structures are observationally equivalent.

In the incomplete models which arise in this paper, a particular structure m may generate more than one collection of conditional distributions. We denote the set of collections of conditional

[10]For identification results employing classical treatments of observational equivalence, that is where a structure produces a unique $\mathcal{P}_{Y|Z}(m)$, equivalently a unique $P_{Y|Z}(\cdot|z; m)$ for each $z \in \mathcal{R}_Z$, see for example Koopmans and Reiersøl (1950), Hurwicz (1950), Rothenberg (1971), and Bowden (1973). For a nonparametric treatment see e.g. Matzkin (2007, 2008).

distributions that can be generated by a structure m by $\mathsf{P}_{Y|Z}(m)$. Since sets $\mathsf{P}_{Y|Z}(m)$ and $\mathsf{P}_{Y|Z}(m')$ generated by distinct structures, m and m', may have collections of conditional distributions in common, the property of observational equivalence of two structures, m and m', may depend on the particular collection of conditional distributions $\mathcal{F}_{Y|Z}$ under consideration in identification analysis. This is so because there may be a collection, say $\mathcal{F}^*_{Y|Z}$, which lies in $\mathsf{P}_{Y|Z}(m)$ and in $\mathsf{P}_{Y|Z}(m')$ and a collection $\mathcal{F}^{**}_{Y|Z}$ which lies in $\mathsf{P}_{Y|Z}(m)$ but not in $\mathsf{P}_{Y|Z}(m')$. Structures m and m' are observationally equivalent in identification analysis employing $\mathcal{F}^*_{Y|Z}$ but not in identification analysis employing $\mathcal{F}^{**}_{Y|Z}$.

Consequently, in the following development, we define observational equivalence with respect to the (identified) collection of distributions $\mathcal{F}_{Y|Z}$, and a corresponding notion of *potential* observational equivalence, which is a property which two structures can possess irrespective of the collection of conditional distributions $\mathcal{F}_{Y|Z}$ under consideration in identification analysis.[11]

To provide formal definitions of these properties we begin by providing the definition of a selection from a random set and the definition of selectionability as given by Molchanov (2005, Definition 2.2, p. 26 and Definition 2.19, p. 34).

Definition 1 *Let W and $\mathcal{W}$ denote a random vector and random set defined on the same probability space. W is a* ***selection*** *of $\mathcal{W}$, denoted $W \in \mathrm{Sel}(\mathcal{W})$, 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 random a 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 $\mathcal{Y}(U,Z;h)$ conditional on $Z = z$, for all $z \in \mathcal{R}_Z$. If $\mathcal{Y}(U,Z;h)$ is a singleton set with probability one, then the model is complete, and the conditional distribution of $\mathcal{Y}(U,Z;h)$ given $Z = z$ is simply that of $\{Y\}$ given $Z = z$ for each $z \in \mathcal{R}_Z$. In this case, again for each $z \in \mathcal{R}_Z$, $F_{Y|Z}(\cdot|z)$ is the *only* conditional distribution of Y given $Z = z$ that is selectionable with respect to the conditional distribution of $\mathcal{Y}(U,Z;h)$, and our 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

[11] In our formulation of observational equivalence and characterizations of identified sets, we continue to work with conditional distributions of endogenous and latent variables, $F_Y(\cdot|z)$ and $G_U(\cdot|z)$, respectively, for almost every $z \in \mathcal{R}_Z$. Knowledge of the distribution of Z, F_Z, combined with $F_Y(\cdot|z)$ or $G_U(\cdot|z)$ a.e. $z \in \mathcal{R}_Z$ is equivalent to knowledge of the joint distribution of (Y,Z) denoted F_{YZ}, or that of (U,Z), denoted G_{UZ}, respectively. We show formally in Appendix B that our characterizations using selectionability conditional on $Z = z$, a.e. $z \in \mathcal{R}_Z$, are indeed equivalent to using analogous selectionability criteria for the joint distributions F_{YZ} or G_{UZ}.

precisely those distributions for which $h(Y,Z,U)=0$ can hold with probability one for the given $\left(h,\mathcal{G}_{U|Z}\right)$. 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 $\left(h,\mathcal{G}_{U|Z}\right)$.

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

Definition 2 *Under Restrictions A1-A3, two structures $\left(h,\mathcal{G}_{U|Z}\right)$ 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 $\left(h,\mathcal{G}_{U|Z}\right)$ and $\left(h',\mathcal{G}'_{U|Z}\right)$ are **observationally equivalent** with respect to $\mathcal{F}_{Y|Z}=\left\{F_{Y|Z}(\cdot|z):z\in\mathcal{R}_Z\right\}$ 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 $\left(h,\mathcal{G}_{U|Z}\right)$ is as follows.[12]

Definition 3 *Under Restrictions A1-A4, the **identified set** of structures $\left(h,\mathcal{G}_{U|Z}\right)$ 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. $z\in\mathcal{R}_Z$:*

$$\mathcal{M}^*\equiv\left\{\left(h,\mathcal{G}_{U|Z}\right)\in\mathcal{M}:F_{Y|Z}(\cdot|z)\preccurlyeq\mathcal{Y}(U,z;h)\ \textit{when } U\sim G_{U|Z}(\cdot|z)\,,\ \textit{a.e. } z\in\mathcal{R}_Z\right\}. \tag{3.1}$$

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. A task that remains is to characterize, in any particular model, the set of structures for which the given selectionability criteria holds. Any characterization of selectionability will suffice. For example, Beresteanu, Molchanov, and Molinari (2011) show 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 broad class of econometric models.

Given Definition 3 of the identified set of structures admitted by a model, we can now define set identification of structural features. As is standard, we define a structural feature $\psi(\cdot,\cdot)$ as any functional of a structure $\left(h,\mathcal{G}_{U|Z}\right)$. Examples include the structural function h itself, $\psi\left(h,\mathcal{G}_{U|Z}\right)=h$, the distributions of unobserved heterogeneity, $\psi\left(h,\mathcal{G}_{U|Z}\right)=\mathcal{G}_{U|Z}$, and counterfactual probabilities such as the probability that a component of Y exceeds a given threshold conditional on $Z=z$.

[12] The identified set $\mathcal{M}^*$ depends upon the collection of conditional distributions $\mathcal{F}_{Y|Z}$, although we do not make this dependence explicit in our notation.

Definition 4 *The* ***identified set of structural features*** $\psi(\cdot,\cdot)$ *under Restrictions A1-A4 is*

$$\Psi \equiv \left\{\psi\left(h, \mathcal{G}_{U|Z}\right) : \left(h, \mathcal{G}_{U|Z}\right) \in \mathcal{M}^*\right\}.$$

Depending on the context, a variety of different features may be of interest. The identified set of structures $\mathcal{M}^*$ can be used to ascertain the identified set of any such feature. We thus take the identified set of structures $\mathcal{M}^*$ as the focus of our analysis, and unless we specify a particular feature of interest, reference to only the "identified set" without qualification refers to $\mathcal{M}^*$.[13]

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 by Beresteanu, Molchanov, and Molinari (2012). In models with $G_{U|Z}(\cdot|z)$ parametrically specified, this approach or a capacity functional characterization of selectionability can be used, see e.g. BMM11 or the related characterization of Galichon and Henry (2011). Proposed estimation and inference strategies based on these approaches entail simulation of admissible distributions of $G_{U|Z}(\cdot|z)$, which are then plugged into the outcome correspondence $\mathcal{Y}(\cdot,\cdot;h)$ to generate simulated conditional distributions of random set $\mathcal{Y}(U, Z; h)$, see also Henry, Meango, and Queyranne (2011). In this class of models, that is with the same independence restriction and $G_{U|Z}(\cdot|z)$ parametrically specified, our characterization using selectionability from random sets in U-space provides characterizations of $\mathcal{M}^*$ comprising inequalities that can be checked via either numerical computation or simulation.

In the following Section we prove the equivalence of the characterizations of observational equivalence and the identified set $\mathcal{M}^*$ given in Definitions 2 and 3 and characterizations expressed in terms of selectionability of $G_{U|Z}(\cdot|z)$ from the random residual set $\mathcal{U}(Y, Z; h)$. These characterizations in terms of sets on the support of unobserved heterogeneity enable direct and immediate consideration of any conceivable alternative restrictions on the joint distribution of $G_{U|Z}(\cdot|z)$.

For example, we show in Section 3.4 that when U and Z are restricted to be independently distributed but the form of $G_{U|Z}(\cdot|z)$ is left unspecified, characterization of $\mathcal{H}^*$, the identified set of structural functions h, can be reduced to a collection of inequality restrictions from which $G_{U|Z}(\cdot|z)$ is absent. By contrast, characterizations of identified sets using the selectionability criteria of $F_{Y|Z}(\cdot|z)$ from the random outcome set $\mathcal{Y}(U, z; h)$, do involve $G_{U|Z}(\cdot|z)$. Following this approach, to check whether any given h can be a component of an element $\left(h, \mathcal{G}_{U|Z}\right)$ of the identified set $\mathcal{M}^*$, one must devise a way to check the selectionability criteria over a nonparametric infinite dimensional class of functions $G_{U|Z}(\cdot|z)$ for each $z \in \mathcal{R}_Z$.

[13]The identified set of structural features Ψ depends on both $\mathcal{M}$ and the conditional distributions $\mathcal{F}_{Y|Z}$, but for ease of notation we suppress this dependence.

3.2 Set Identification via Selectionability in U-Space

In this section a dual relation is derived between random outcome sets $\mathcal{Y}(U, Z; h)$ and random residual sets $\mathcal{U}(Y, Z; h)$. This is then used to relate selectionability of $F_{Y|Z}(\cdot|z)$ with respect to $\mathcal{Y}(U, Z; h)$ and selectionability of $G_{U|Z}(\cdot|z)$ with respect to $\mathcal{U}(Y, Z; h)$.

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 $Y \sim F_{Y|Z}(\cdot|z)$.*

With Theorem 1 established, we now characterize the identified set in terms of random variables and sets in the space of unobserved heterogeneity. As previously expressed, a key advantage to doing this is the ability to impose restrictions directly on $G_{U|Z}$ 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 Section 4.

Theorem 2 *Let Restrictions A1-A3 hold. Then (i) 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. $z \in \mathcal{R}_Z$; and (ii) the identified set of structures $\left(h, \mathcal{G}_{U|Z}\right)$ are those such that $G_{U|Z}(\cdot|z)$ is selectionable with respect to the conditional (on $Z = z$) distribution of random set $\mathcal{U}(Y, Z; h)$.*

Theorem 2 uses duality to express observational equivalence and characterization of the identified set of structures $\left(h, \mathcal{G}_{U|Z}\right)$ in terms of selectionability from 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 $G_{U|Z}$ is selectionable with respect to the conditional distribution of $\mathcal{U}(Y, Z; h)$ will suffice for characterization of the identified set.

One such characterization, used in previous papers allowing only discrete outcomes, e.g. Chesher, Rosen, and Smolinski (2013) and Chesher and Rosen (2013a, 2013b), uses Artstein's Inequality, see e.g. Artstein (1983), Norberg (1992), and Molchanov (2005, Section 1.4.8). This result allows us to characterize the identified set $\mathcal{M}^*$ through the conditional containment functional of random set $\mathcal{U}(Y, Z; h)$, defined as

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

Characterization via the containment functional produces an expression for $\mathcal{M}^*$ in the form of conditional moment inequalities, as given in the following Corollary.

Corollary 1 *Under the restrictions of Theorem 2, the identified set can be written*

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

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

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

3.3 Core Determining Test Sets

We now characterize a smaller collection $\mathsf{Q}\left(h, z\right)$ of *core-determining* test sets $\mathcal{S}$ for any h, and any $z \in \mathcal{R}_Z$, such that if

$$\forall \mathcal{S} \in \mathsf{Q}\left(h, z\right), \; C_h\left(\mathcal{S}|z\right) \leq G_{U|Z}\left(\mathcal{S}|z\right), \tag{3.2}$$

then the same inequality holds for all $\mathcal{S} \in \mathsf{F}\left(\mathcal{R}_U\right)$. Thus, characterization of the identified set is reduced to those structures such that (3.2) holds for all $\mathcal{S} \in \mathsf{Q}\left(h, z\right)$.[14] Galichon and Henry (2011) initially introduced core-determining sets for identification analysis in consideration of sets in outcome space, characterizing such classes of sets for incomplete models that satisfy a certain monotonicity requirement, which is not needed here. We extend their definition of such collections by devising core-determining sets for our characterizations in U-space, and by allowing them to be specific to the structural relation h and covariate value z.[15]

Our construction builds on ideas from Chesher, Rosen, and Smolinski (2013), but is much more widely applicable. We do not require independence of U and Z, allowing for other restrictions on $\mathcal{G}_{U|Z}$. We also establish conditions whereby for some of the core-determining sets belonging to $\mathsf{Q}\left(h, z\right)$, the inequality (3.2) must hold with *equality.* Under these conditions the initial characterization via inequalities is sharp, but through use of the law of total probability some of these inequalities can be strengthened to equalities. Much is known about observable implications of models with conditional moment *equalities*, and recognition that some of the inequalities must hold

[14] Earlier versions of some results in this section appeared in the 2012 version of the CeMMAP working paper Chesher and Rosen (2012b), which concerned models with only discrete endogenous variables. In this Section we provide more general results that cover the broader class of models studied here. In revisions of Chesher and Rosen (2012b) we refer to the more general results here.

[15] Through (3.4) in Theorem 3 below, the core-determining sets may also be dependent upon $G_{U|Z}\left(\cdot|z\right)$, but we suppress this from the notation.

with equality can potentially be helpful for estimation and in consideration of conditions that may feasibly lead to point identification.[16]

For this development we first define the support of the random set $\mathcal{U}(Y,Z;h)$ conditional on $Z=z$, and the collection of sets comprising unions of such sets. These objects both play important roles. In this definition and the subsequent analysis we employ the following notation at the cost of some slight abuse of notation:

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

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

Definition 5 *Under Restrictions A1-A3, the* ***conditional support of random set*** $\mathcal{U}(Y,Z;h)$ *given* $Z=z$ *is*

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

The collections of all sets that are unions of elements of $\mathsf{U}(h,z)$ *is denoted*

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

Lemma 1 below establishes that for any $(h,z) \in \mathcal{H} \times \mathcal{Z}$, in order for (3.2) to hold for all closed $\mathcal{S} \subseteq \mathcal{R}_U$, it suffices to show only that (3.2) holds for those sets $\mathcal{S} \in \mathsf{U}^*(h,z)$. To state the result, we define some additional notation. For any set $\mathcal{S} \subseteq \mathcal{R}_U$ and any $(h,z) \in \mathcal{H} \times \mathcal{R}_Z$, define

$$\mathsf{U}^{\mathcal{S}}(h,z) \equiv \{\mathcal{U} \in \mathsf{U}(h,z) : \mathcal{U} \subseteq \mathcal{S}\}, \qquad \mathsf{U}_{\mathcal{S}}(h,z) \equiv \left\{\mathcal{U} \in \mathsf{U}(h,z) : G_{U|Z}(\mathcal{U} \cap \mathcal{S}|z) = 0\right\},$$

which are the sets $\mathcal{U} \in \mathsf{U}(h,z)$ that are contained in $\mathcal{S}$ and that, up to zero measure $G_{U|Z}(\cdot|z)$, do not hit $\mathcal{S}$, respectively. Define

$$\overline{\mathsf{U}}^{\mathcal{S}}(h,z) \equiv \mathsf{U}(h,z) / \left(\mathsf{U}^{\mathcal{S}}(h,z) \cup \mathsf{U}_{\mathcal{S}}(h,z)\right),$$

which comprises the sets $\mathcal{U} \in \mathsf{U}(h,z)$ that belong to neither $\mathsf{U}^{\mathcal{S}}(h,z)$ nor $\mathsf{U}_{\mathcal{S}}(h,z)$. For ease of reference, Table 1 provides a summary of the collections of sets $\mathsf{U}(h,z)$, $\mathsf{U}^*(h,z)$, $\mathsf{U}^{\mathcal{S}}(h,z)$, $\mathsf{U}_{\mathcal{S}}(h,z)$, and $\overline{\mathsf{U}}^{\mathcal{S}}(h,z)$ used to establish the following Lemma and Theorem 3 below.

[16] For instance, Tamer (2003) shows that observable implications of a simultaneous binary outcome model constitute conditional moment equalities and inequalities, and proves point identification of parameters through use of the equalities. Aradillas-Lopez and Rosen (2013) provide conditions for point identification of a subset of model parameters in a simultaneous ordered entry model that also delivers conditional moment inequalities and equalities through the conditional moment *equalities*. The use of moment equalities for point identification of course additionally requires that observable variables satisfy certain rank and/or support conditions.

Collection	Description
$\mathsf{U}(h,z)$	Support of $\mathcal{U}(Y,Z;h)$ conditional on $Z=z$.
$\mathsf{U}^*(h,z)$	Sets that are unions of sets in $\mathsf{U}(h,z)$.
$\mathsf{U}^{\mathcal{S}}(h,z)$	Sets in $\mathsf{U}(h,z)$ that are contained in $\mathcal{S}$.
$\mathsf{U}_{\mathcal{S}}(h,z)$	Sets in $\mathsf{U}(h,z)$ with $G_{U\|Z}$ measure zero intersection with $\mathcal{S}$.
$\overline{\mathsf{U}}^{\mathcal{S}}(h,z)$	Sets in $\mathsf{U}(h,z)$ contained in neither $\mathsf{U}^{\mathcal{S}}(h,z)$ nor $\mathsf{U}_{\mathcal{S}}(h,z)$.

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 $\mathcal{U}_{\mathcal{S}}(h,z)$ denote the union of all sets in $\mathsf{U}^{\mathcal{S}}(h,z)$,*

$$\mathcal{U}_{\mathcal{S}}(h,z) \equiv \bigcup_{\mathcal{U} \in \mathsf{U}^{\mathcal{S}}(h,z)} \mathcal{U}. \tag{3.3}$$

If

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

then

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

Lemma 1 establishes that if (3.2) holds for all unions of sets in $\mathsf{U}(h,z)$, then it holds for all closed test sets $\mathcal{S} \subseteq \mathcal{R}_U$.

All sets in the collection of core-determining sets are also unions of sets in $\mathsf{U}(h,z)$, but not all such unions lie in the core-determining collection. Theorem 3 below defines a collection of core-determining test sets $\mathsf{Q}(h,z)$, which is a refinement of $\mathsf{U}^*(h,z)$. The elements of $\mathsf{U}^*(h,z)$ that can be excluded from the core-determining collection have the property that each one can be partitioned into two sub-collections of $\mathsf{U}^*(h,z)$ such that (i) each is itself a member of the core-determining collection, and (ii) the unions of all elements in each sub-collection are disjoint relative to the probability measure $G_{U|Z}(\cdot|z)$.

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

$$\mathcal{S}_1 \equiv \bigcup_{\mathcal{T} \in \mathsf{S}_1} \mathcal{T},\ \mathcal{S}_2 \equiv \bigcup_{\mathcal{T} \in \mathsf{S}_2} \mathcal{T},\ \text{and } G_{U|Z}\left(\mathcal{S}_1 \cap \mathcal{S}_2|z\right) = 0, \tag{3.4}$$

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

Note that all sets of the form $\mathcal{U}(y,z;h)$ with $y \in \mathcal{R}_Y$ are contained in $\mathsf{Q}(h,z)$, so that all sets in $\mathsf{U}(h,z)$ are elements of $\mathsf{Q}(h,z)$. Theorem 3 implies that the identified sets of Theorem 2 are characterized by the set of structures $\left(h, \mathcal{G}_{U|Z}\right)$ that satisfy the containment functional 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 (3.4) is implied if the sets $\mathcal{S}_1$ and $\mathcal{S}_2$ have non-overlapping interiors. This implication was used in the construction of core-determining sets in Chesher, Rosen, and Smolinski (2013), in which all elements of $\mathsf{U}(h,z)$ were in fact connected.

To illustrate the results of Theorem 3 in a relatively simple context we refer back to Example 2 of Section 2.1, also studied in Chesher and Rosen (2013b). 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 $y_1 = 0$. 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 $\mathcal{M}^*$ 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

$$\mathcal{S} = [0, g(y_2,z_1)] \cup \left[g\left(y_2',z_1\right), 1\right], \text{ for some } y_2, y_2' \in \mathcal{R}_{Y_2}. \tag{3.5}$$

Now consider test sets $\mathcal{S}$ of the form given in (3.5). 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 (3.2) 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 $\mathcal{R}_U$. 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 $y_2 \in \mathcal{R}_{Y_2}$.

The following Corollary shows that in some models some of the containment functional inequalities for core-determining sets can be replaced by equalities.[17] Then the identified set can be written as a collection of conditional moment inequalities and equalities. We consider models in which $G_{U|Z}(\cdot|z)$ is restricted to be absolutely continuous with respect to Lebesgue measure. The strengthening of inequality (3.2) to an equality occurs for test sets $\mathcal{S} \in \mathsf{Q}(h,z)$ such that any $\mathcal{U}(y,z;h)$ not contained in $\mathcal{S}$ lies in the closure of the complement of $\mathcal{S}$, denoted $\overline{\mathcal{S}^c}$. In this case each set $\mathcal{U}(y,z;h)$ is either contained in $\mathcal{S}$ or contained in $\overline{\mathcal{S}^c}$, and we have that $C_h(\mathcal{S}|z) + C_h\left(\overline{\mathcal{S}^c}|z\right) = 1$. Likewise $G_{U|Z}(\mathcal{S}|z) + G_{U|Z}\left(\overline{\mathcal{S}^c}|z\right) = 1$, and this combined with the inequalities (3.2) for both $\mathcal{S}$ and $\overline{\mathcal{S}^c}$ imply that the weak inequality must hold with equality. The formal statement of this result

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

follows.

Corollary 2 *Define*

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

Then, if $G_{U|Z}(\cdot|z)$ is absolutely continuous with respect to Lebesgue measure, under the conditions of Theorem 3, the collection of equalities and inequalities

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

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

There are two particular kinds of models in which *all* elements of $\mathsf{Q}(h,z)$ belong to $\mathsf{Q}^E(h,z)$, so that the characterization of the identified set delivered by the Corollary comprises a collection of *only* conditional moment *equalities*. First, the conditions defining $\mathsf{Q}^E(h,z)$ in the Corollary apply to *every* test set $\mathcal{S}$ in models where $\mathcal{U}(Y,Z;h)$ is a singleton set with probability one. Models with this property include traditional IV models with additive unobservables such as the classical linear IV model of Example 1 and the nonparametric IV model of Newey and Powell (2003), as well as IV models with structural function strictly monotone in a scalar unobservable, for example the quantile IV model studied by Chernozhukov and Hansen (2005). Second, in complete models, that is models where the set $\mathcal{Y}(U,Z;h)$ is a singleton set with probability one, the identified set is also characterized entirely by conditional moment equalities. In such models for any z and any $y \neq y'$ the sets $\mathcal{U}(y,z;h)$ and $\mathcal{U}(y',z;h)$ have measure zero intersection, as otherwise the model would produce non-singleton outcome sets $\mathcal{Y}(U,Z;h)$ with positive probability. Since $\mathsf{Q}(h,z)$ is a collection of sets that are unions of sets on the support of $\mathcal{U}(Y,Z;h)$ we then have that for all (y,z) and any h, either $\mathcal{U}(y,z;h) \subseteq \mathcal{S}$ or $\mathcal{U}(y,z;h) \subseteq \overline{\mathcal{S}^c}$ for all $\mathcal{S} \in \mathsf{Q}(h,z)$.

In general the collection of core-determining sets from Theorem 2 and Corollary 2 may be infinite. However, in models in which all endogenous variables are discrete and finite, the sets $\mathsf{Q}^E(h,z)$ and $\mathsf{Q}^I(h,z)$ are finite. In Chesher and Rosen (2012b) we provide an algorithm based on the characterization of core-determining sets in Theorem 2 and Corollary 2 to construct the collections $\mathsf{Q}^E(h,z)$ and $\mathsf{Q}^I(h,z)$ in such models.

3.4 Restrictions on the Joint Distribution of (U,Z)

Given a model $\mathcal{M}$, Theorem 2 provides a characterization of the structures $(h, \mathcal{G}_{U|Z})$ that belong to the identified set. A key element of econometric models are restrictions on the conditional distributions of unobserved heterogeneity, i.e. restrictions on the collections $\mathcal{G}_{U|Z}$ that are admitted by $\mathcal{M}$. The generality of Theorem 2 allows for its application whatever the specification of $\mathcal{M}$, though of course the size of the identified set $\mathcal{M}^*$ will depend crucially on the restrictions that $\mathcal{M}$ embodies.

In this section we consider particular restrictions on admissible collections of conditional distributions $\mathcal{G}_{U|Z}$, showing how additional restrictions refine characterization of the identified set. The restrictions we consider are much-used in practice, namely independence, conditional mean, conditional quantile, and parametric restrictions. Theorem 2 can also be applied with other restrictions.

Stochastic Independence

Here we consider the implications of stochastic independence of unobservables and exogenous variables set out in the following restriction.

Restriction SI: For all collections $\mathcal{G}_{U|Z}$ of conditional distributions admitted by $\mathcal{M}$, $U \perp\!\!\!\perp Z$. $\square$

With this restriction in place, the conditional distributions $G_{U|Z}(\cdot|z)$ cannot vary with z, and we can simply write G_U in place of the collection $\mathcal{G}_{U|Z}$, where for each z, $G_{U|Z}(\cdot|z) = G_U(\cdot)$, and $\mathcal{M}$ is then denoted by a collection of structures (h, G_U). Let $\mathcal{G}_U \equiv \{G_U : \exists h \text{ s.t. } (h, G_U) \in \mathcal{M}\}$ denote the collection of distributions of unobserved heterogeneity admitted by model $\mathcal{M}$.

It follows from Theorem 2 that a given structure $(h, G_U) \in \mathcal{M}$ belongs to $\mathcal{M}^*$ if and only if G_U is selectionable with respect to the conditional (on $Z = z$) distribution of the random set $\mathcal{U}(Y, Z; h)$ induced by $F_{Y|Z}(\cdot|z)$ a.e. $z \in \mathcal{R}_Z$. A characterization of such structures is succinctly given through the conditional containment inequality representation, as set out in the following Theorem.

Theorem 4 *Let Restrictions A1-A4 and SI hold. Then*

$$\begin{aligned}
\mathcal{M}^* &= \left\{(h, G_U) \in \mathcal{M} : F_{Y|Z}(\cdot|z) \preccurlyeq \mathcal{Y}(U, z; h) \text{ when } U \sim G_U(\cdot), \text{ a.e. } z \in \mathcal{R}_Z\right\} && (3.6)\\
&= \left\{(h, G_U) \in \mathcal{M} : G_U(\cdot) \preccurlyeq \mathcal{U}(Y, z; h) \text{ when } Y \sim F_{Y|Z}(\cdot|z), \text{ a.e. } z \in \mathcal{R}_Z\right\}, && (3.7)
\end{aligned}$$

equivalently,

$$\begin{aligned}
\mathcal{M}^* &= \left\{ \begin{array}{c} (h, G_U) \in \mathcal{M} : \forall \mathcal{S}_I \in \mathsf{Q}^I(h, z), \forall \mathcal{S}_E \in \mathsf{Q}^E(h, z), \\ C_h(\mathcal{S}_I|z) \le G_U(\mathcal{S}_I), C_h(\mathcal{S}_E|z) = G_U(\mathcal{S}_E), \text{ a.e. } z \in \mathcal{R}_Z \end{array} \right\}. && (3.8)\\
&= \left\{ \begin{array}{l} (h, G_U) \in \mathcal{M} : \forall \mathcal{K} \in \mathsf{K}(\mathcal{Y}), \\ \quad F_{Y|Z}(\mathcal{K}|z) \le G_U\{\mathcal{Y}(U, z; h) \cap \mathcal{K} \neq \emptyset\}, \text{ a.e. } z \in \mathcal{R}_Z \end{array} \right\}, && (3.9)
\end{aligned}$$

where $\mathsf{K}(\mathcal{Y})$ denotes the collection of compact sets in $\mathcal{R}_Y$.

Theorem 4 presents various representations of the identified set under Restriction SI. The first two characterizations, (3.6) and (3.7) are direct applications of this restriction to Definition 3.1 and Theorem 2, respectively. The characterization (3.8) applies Theorem 3 and Corollary 2 to provide a characterization of the identified set through the conditional containment functional of $\mathcal{U}(Y,Z;h)$. The representation makes use of core-determining sets to reduce the required number of moment conditions in the characterization, and to distinguish which must hold as equalities and inequalities.

The last characterization, (3.9), characterizes the identified set through conditional moment inequalities implied by the capacity functional applied to random set $\mathcal{Y}(U,z;h)$. These inequalities coincide with the characterizations provided by Beresteanu, Molchanov, and Molinari (2011, Appendix D.2) and Galichon and Henry (2011) when applied to incomplete models of games. They must hold applied to all compact sets $\mathcal{K} \subseteq \mathcal{R}_Y$, although Galichon and Henry (2011) provide core determining sets for this characterization in $\mathcal{R}_Y$ when a certain monotonicity condition holds. There are many models where the required monotonicity condition does not hold. Nonetheless, the representation (3.8), and in particular the reduction in moment conditions achieved *via* the use of core determining sets on $\mathcal{R}_U$ given by Theorem 3, still holds.

A further difference between characterizations (3.8) and (3.9) is how they incorporate restrictions placed on the distribution of unobserved heterogeneity. Given an admissible distribution G_U, the use of characterization (3.9) computationally requires that one compute for each compact set $\mathcal{K}$ the probability that the random outcome set $\mathcal{Y}(U,z;h)$ hits $\mathcal{K}$. This has typically been achieved by means of simulation from each conjectured distribution G_U, see e.g. Beresteanu, Molchanov, and Molinari (2011, Appendix D.2) and Henry, Meango, and Queyranne (2011). $F_{Y|Z}(\mathcal{K}|z)$ is observed directly. On the other hand, characterization (3.8) requires, for each conjectured distribution G_U and each core-determining set $\mathcal{S}$, computation of $G_U(\mathcal{S})$. This can be done again via simulation, or by means of numerical integration. The term $\mathbb{P}[\mathcal{U}(Y,Z;h) \subseteq \mathcal{S}|z]$ is, for any fixed h, the probability of an event concerning only the variables (Y,Z), which is point-identified and can be computed or estimated directly.

It is important to understand that with Restriction SI imposed, Theorem 4 applies with the admissible distributions $\mathcal{G}_U$ either parametrically or non-parametrically specified. For example, if admissible G_U were parameterized by $\lambda \in \Lambda \subseteq \mathbb{R}^d$ we could write $\mathcal{G}_U = \{G_U(\mathcal{S};\lambda) : \lambda \in \Lambda\}$, and $\mathcal{M}^*$ could be represented by a collection of (h,λ) satisfying the equalities and inequalities of (3.8), with $G_U(\mathcal{S};\lambda)$ replacing $G_U(\mathcal{S})$. Less stringent conditions on the class $\mathcal{G}_U$, all else equal, will result in larger identified sets.

If G_U is left completely unrestricted, then the task of checking the containment functional inequality for all admissible G_U is more difficult than with G_U parametrically specified. Indeed, it is not clear how to do this for all such G_U.

Fortunately, manipulation of the conditional containment functional inequality representation in U-space affords a representation of the identified set of structural functions h that does not explicitly involve the distribution G_U. The ability to characterize distribution-free identified sets using random set theory with statistical independence restrictions is new to the literature. The formal result follows.

Corollary 3 *Let Restrictions A1-A4 and SI hold, and let $\mathcal{G}_{U|Z}$ be otherwise unrestricted. Then the identified set of structural functions h is*

$$\mathcal{H}^* = \left\{ h \in \mathcal{H} : \forall \mathcal{S} \in \mathsf{Q}^*(h), \quad \sup_{z \in \mathcal{R}_Z} C_h(\mathcal{S}|z) \leq \inf_{z \in \mathcal{R}_Z} (1 - C_h(\mathcal{S}^c|z)), \right\}, \tag{3.10}$$

where $\mathsf{Q}^*(h)$ is any collection of sets $\mathcal{S} \subseteq \mathcal{R}_U$ such that for all $z \in \mathcal{R}_Z$, $\mathsf{Q}(h,z) \subseteq \mathsf{Q}^*(h)$.

Corollary 3 shows that with a distribution-free specification, the probability $G_U(\mathcal{S})$ can be profiled out of the containment functional inequality. This holds because for any set $\mathcal{S}$ we have

$$C_h(\mathcal{S}|z) \leq G_U(\mathcal{S}), \text{ and } C_h(\mathcal{S}^c|z) \leq G_U(\mathcal{S}^c). \tag{3.11}$$

The second equality is equivalent to

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

where $1 - C_h(\mathcal{S}^c|z)$ is the *capacity* functional of $\mathcal{U}(Y,Z;h)$ applied to argument $\mathcal{S}$ conditional on $Z = z$.[18] Rearranging and combining with (3.11) we have for almost every $z \in \mathcal{R}_Z$.

$$C_h(\mathcal{S}|z) \leq G_U(\mathcal{S}) \leq 1 - C_h(\mathcal{S}^c|z), \tag{3.12}$$

which in combination with Corollary 2 produces the characterizations of Corollary 3. Going in the opposite direction, if the inequalities $C_h(\mathcal{S}|z) \leq (1 - C_h(\mathcal{S}^c|z))$ hold for all sets $\mathcal{S} \in \mathsf{Q}^*(h)$ and almost every $z \in \mathcal{R}_Z$, then they hold for all core-determining sets. Monotonicity of $C_h(\mathcal{S}|z)$ and $1 - C_h(\mathcal{S}^c|z)$ in $\mathcal{S}$ and the fact that these expressions are bounded between zero and one then guarantees that there exists at least one distribution $G_U(\mathcal{S})$ such that (3.12) holds a.e. $z \in \mathcal{R}_Z$.

Mean Independence

The following restriction limits the collection $\mathcal{G}_{U|Z}$ to those such that $E[U|z]$ does not vary with z, i.e. that $E[U|z] = c$ for some fixed c belonging to a known set $\mathcal{C}$. This nests cases where components of c are either known to be zero or normalized to zero, as well as cases where some components of c are completely unknown. For instance, in a model with bivariate U, Restriction MI with $\mathcal{C} = \{0\} \times \mathbb{R}$ entails the restrictions $E[U_1|z] = 0$, e.g. due to normalization, and $E[U_2|z]$ invariant with respect to z.

[18] In other words

$$1 - C_h(\mathcal{S}^c|z) = \mathbb{P}[\mathcal{U}(Y,Z;h) \cap \mathcal{S} \neq \emptyset | z].$$

We use $1 - C_h(\mathcal{S}^c|z)$ for the capacity functional rather than introduce further notation or explicitly write out the longer expression $\mathbb{P}[\mathcal{U}(Y,Z;h) \cap \mathcal{S} \neq \emptyset|z]$ repeatedly.

Restriction MI: $\mathsf{G}_{U|Z}$ comprises all collections $\mathcal{G}_{U|Z}$ of conditional distributions for U given Z satisfying $E[U|z] = c$, a.e. $z \in \mathcal{R}_Z$, for some fixed, finite c belonging to a known set $\mathcal{C} \subseteq \mathcal{R}_U$. $\square$

With this restriction imposed on the conditional distributions of unobserved heterogeneity, it is convenient to characterize the selectionability criterion of Theorem 2 by making use of the Aumann expectation. Also referred to as the selection expectation, the Aumann expectation of a random set $\mathcal{A}$ is the set of values that are the expectation of some random variable A that is a selection of $\mathcal{A}$. For clarity we provide the formal definition, repeated from Molchanov (2005, p. 151).

Definition 6 *The **Aumann expectation** of $\mathcal{A}$ is*

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

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

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

The resulting characterization of the identified sets for structural function h and for the structure $(h, \mathcal{G}_{U|Z})$ employing the Aumann expectation with Restriction MI is given in the following Theorem.

Theorem 5 *Let Restrictions A1-A4 and MI hold and suppose that $(\Omega, \mathcal{F}, \mathbb{P})$ is non-atomic. Then the identified set for structural function h are those functions h such that some $c \in \mathcal{C}$ is an element of the Aumann expectation of $\mathcal{U}(Y, Z; h)$ conditional on $Z = z$, a.e. $z \in \mathcal{R}_Z$:*

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

The identified set for $(h, \mathcal{G}_{U|Z})$ is:

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

where by virtue of Restriction MI all structures $(h, \mathcal{G}_{U|Z}) \in \mathcal{M}^* \subseteq \mathcal{M}$ are such that for some $c \in \mathcal{C}$, $E[U|z] = c$ a.e. $z \in \mathcal{R}_Z$.

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

Properties of the random set $\mathcal{U}(Y, Z; h)$ can be helpful in characterizing its Aumann expectation, and consequently in determining whether any particular h is in $\mathcal{H}^*$. For example, if $\mathcal{U}(Y, Z; h)$ is integrably bounded, that is if

$$E\sup\{\|U\|:U\in\mathcal{U}(Y,Z;h)\}<\infty, \tag{3.13}$$

then from Molchanov (2005, Theorem 2.1.47-iv, p. 171), $c\in\mathbb{E}[\mathcal{U}(Y,Z;h)|z]$ if and only if

$$\inf_{v\in\mathcal{R}_Z:\|v\|=1}\left\{E[m(v,\mathcal{U}(Y,Z;h))|z]-v'c\right\}\geq 0, \tag{3.14}$$

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

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

denotes the support function of $\mathcal{S}$ evaluated at v. BMM11 previously employed Molchanov (2005, Theorem 2.1.47-iv, p. 171) in consideration of the conditional Aumann expectation of random outcome set $\mathcal{Y}(Z,U;h)$ in characterizing its selections for identification analysis, and it can likewise lead to simplifications in determining whether $c\in\mathbb{E}[\mathcal{U}(Y,Z;h)|z]$ for some $c\in\mathcal{C}$.[19] Indeed, if structural function h is additively separable in Y, the two representations are equivalent, differing only by a trivial location shift.

More generally, Theorem 5 does not require the sets $\mathcal{U}(Y,Z;h)$ to be integrably bounded, but only integrable, which is important for dealing with cases where the support of unobserved heterogeneity $\mathcal{R}_U$ is unbounded, e.g. when $\mathcal{R}_U$ is some finite dimensional Euclidean space.

In some commonly occurring models, including all those of Examples 1-5 in Section 2.1, the random set $\mathcal{U}(Y,Z;h)$ is convex with probability one. In such cases the characterization of $\mathcal{H}^*$ can be simplified further as in the following Corollary. Unlike the simplification afforded by the support function characterization (3.14), it does not require that $\mathcal{U}(Y,Z;h)$ be integrably bounded.

Corollary 4 *Let the restrictions of Theorem 5 hold and suppose $\mathcal{U}(Y,Z;h)$ is convex with probability one. Then*

$$\mathcal{H}^*=\left\{\begin{array}{l} h\in\mathcal{H}:\exists c\in\mathcal{C}\ s.t.\ E[u(Y,Z)|z]=c\ a.e.\ z\in\mathcal{R}_Z, \\ for\ some\ function\ u:\mathcal{R}_{YZ}\to\mathcal{R}_U\ with\ \mathbb{P}[u(Y,Z)\in\mathcal{U}(Y,Z;h)]=1 \end{array}\right\}.$$

Finally, Theorem 5 can be generalized to characterize $\mathcal{H}^*$ under more general forms of conditional mean restrictions than Restriction MI, as expressed in Restriction MI*.

Restriction MI*: $\mathsf{G}_{U|Z}$ comprises all collections $\mathcal{G}_{U|Z}$ of conditional distributions for U given Z such that for some known function $d(\cdot,\cdot):\mathcal{R}_U\times\mathcal{R}_Z\to\mathbb{R}^{k_d}$, $E[d(U,Z)|z]=c$ a.e. $z\in\mathcal{R}_Z$, for

[19]For instance, when $\mathcal{C}=0$, equivalently when $E[U|z]=0$ a.e. $z\in\mathcal{R}_Z$ is imposed, the support function inequality (3.14) implies that the identified set for h are those $h\in\mathcal{H}$ such that

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

some fixed c belonging to a known set $\mathcal{C} \subseteq \mathcal{R}_U$, where $d(u,z)$ is continuous in u for all $z \in \mathcal{R}_Z$. $\square$

Restriction **MI*** requires that the conditional mean of some function taking values in $\mathbb{R}^{k_d}$, namely $d(U,Z)$, the conditional mean $E[d(U,Z)|z]$ does not vary with respect to z. This restriction can accommodate models that impose conditional mean restrictions on functions of unobservables U, and nests Restriction MI upon setting $d(U,Z) = U$. To express the identified set delivered under such a restriction, we define

$$\mathcal{D}(y,z;h) \equiv \{d(u,z) : u \in \mathcal{U}(y,z;h)\}.$$

Thus the random set $\mathcal{D}(Y,Z;h)$ is the set of feasible values for $d(U,Z)$ given observed (Y,Z). Given the requirement of Restriction MI* that $d(\cdot,z)$ is continuous for each z, the set $\mathcal{D}(Y,Z;h)$ is a random *closed* set. The same logic as that used for Theorem 5 then yields the following result.

Corollary 5 *Let Restrictions A1-A4 and MI* hold and suppose that* $(\Omega,\mathcal{F},\mathbb{P})$ *is non-atomic. Then the identified set for structural function h are those such that there exists at least one* $c \in \mathcal{C}$ *that is an element of the Aumann expectation of* $\mathcal{D}(Y,Z;h)$ *conditional on* $Z = z$*, a.e.* $z \in \mathcal{R}_Z$*:*

$$\mathcal{H}^* = \{h \in \mathcal{H} : \exists c \in \mathcal{C} \text{ s.t. } c \in \mathbb{E}[\mathcal{D}(Y,Z;h)|z],\ a.e.\ z \in \mathcal{R}_Z\}.$$

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

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

where by Restriction MI, all structures* $\left(h,\mathcal{G}_{U|Z}\right) \in \mathcal{M}^* \subseteq \mathcal{M}$ *are such that for some* $c \in \mathcal{C}$, $E[d(U,Z)|z] = c$*, a.e.* $z \in \mathcal{R}_Z$*.*

Quantile Independence

Our analysis can also accommodate conditional quantile restrictions on unobserved heterogeneity. To illustrate how in a relatively simple yet useful framework we restrict the analysis of this section to models with univariate unobserved heterogeneity $U \in \mathbb{R}$, where the sets $\mathcal{U}(y,z;h)$ are closed intervals with lower bound $\underline{u}(y,z;h)$ and upper bound $\overline{u}(y,z;h)$, formalized in Restriction IS (interval support) below. The lower and upper bounds may be $-\infty$ and $+\infty$, respectively. This restriction can be fruitfully applied to GIV models with censored endogenous or exogenous variables, as we show in the model with interval censored endogenous variables in Section 4.1.3 below.[20]

[20] The use of conditional quantile restrictions with non-interval $\mathcal{U}(y,z;h)$ and more generally multivariate unobserved heterogeneity is an interesting line of research. This is so even in models with exogenous covariates absent instrumental variable restrictions, and as such is logically distinct from the study of generalized instrumental variable models that is our focus here.

Restriction IS: $\forall (y,z) \in \mathcal{R}_{YZ}$,

$$\mathcal{U}(y,z;h) = [\underline{u}(y,z;h), \overline{u}(y,z;h)], \tag{3.15}$$

where possibly $\underline{u}(y,z;h) = -\infty$ or $\overline{u}(y,z;h) = +\infty$, in which case the corresponding endpoint of the interval (3.15) above is open. $\square$

The conditional quantile restriction is formalized below. In similar manner as done for Restriction MI, which imposed conditional mean independence, we allow for the possibility that the precise value of the conditional quantile is not known, but require that it be invariant with respect to the value of the conditioning variable Z. This restriction is more general than - but indeed nests - the common case where the conditional quantile of unobserved heterogeneity is restricted to be zero, in which the set $\mathcal{C}$ can be set to $\mathcal{C} = \{0\}$.

Restriction QI: For some known $\tau \in (0,1)$ and some known set $\mathcal{C} \subseteq \mathbb{R}$, $\mathsf{G}_{U|Z}$ comprises all collections $\mathcal{G}_{U|Z}$ of conditional distributions for U given Z that are continuous in a neighborhood of their τ-quantile and satisfy the conditional quantile restriction $q_{U|Z}(\tau|z) = c$, a.e. $z \in \mathcal{R}_Z$ for some $c \in \mathcal{C}$. $\square$

For random sets $\mathcal{U}(Y,Z;h)$ with interval-valued realizations, it is easy to check whether there exists a random variable that is selectionable with respect to the distribution of $\mathcal{U}(Y,Z;h)$ conditional on $Z = z$. It is well known that the quantile of a generic random variable W distributed F_W is a parameter that respects stochastic dominance. That is, if $\tilde{W} \sim F_{\tilde{W}}$, and $F_{\tilde{W}}$ stochastically dominates F_W, then $q_W(\tau) \leq q_{\tilde{W}}(\tau)$ for any $\tau \in [0,1]$. The smallest and largest selections of random set $\mathcal{U}(Y,Z;h)$, in terms of stochastic dominance, are those distributions that place all mass on $\underline{u}(Y,Z;h)$ and $\overline{u}(Y,Z;h)$, respectively. Thus, intuitively, the conditional quantiles of $\underline{u}(Y,Z;h)$ and $\overline{u}(Y,Z;h)$ provide sharp bounds on the conditional quantiles of all selections of $\mathcal{U}(Y,Z;h)$. This is formalized with the following result.

Theorem 6 *Let Restrictions A1-A4, IS, and QI hold. Then (i) the identified set for structural function h is*

$$\mathcal{H}^* = \left\{ h \in \mathcal{H} : \exists c \in \mathcal{C} \ s.t. \ \sup_{z \in \mathcal{R}_Z} F_{Y|Z}[\overline{u}(Y,Z;h) \leq c|z] \leq \tau \leq \inf_{z \in \mathcal{R}_Z} F_{Y|Z}[\underline{u}(Y,Z;h) \leq c|z] \right\}. \tag{3.16}$$

(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 an equivalent formulation of $\mathcal{H}^*$ is given by

$$\mathcal{H}^* = \left\{ h \in \mathcal{H} : \exists c \in \mathcal{C} \ s.t. \ \sup_{z \in \mathcal{R}_Z} \underline{q}(\tau,z;h) \leq c \leq \inf_{z \in \mathcal{R}_Z} \overline{q}(\tau,z;h) \right\}, \tag{3.17}$$

where

$$\underline{q}(\tau, z; h) \equiv \tau\text{-quantile of } \underline{u}(Y, Z; h), \quad \overline{q}(\tau, z; h) \equiv \tau\text{-quantile of } \overline{u}(Y, Z; h).$$

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

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

where following from Restriction QI, all structures $\left(h, \mathcal{G}_{U|Z}\right) \in \mathcal{M}^* \subseteq \mathcal{M}$ *are such that for some* $c \in \mathcal{C}$, $q_{U|Z}(\tau|z) = c$, *a.e.* $z \in \mathcal{R}_Z$.

Under Restriction QI, the conditional distributions belonging to $\mathcal{G}_{U|Z}$ are continuous in a neighborhood of zero. Indeed, it is common for econometric models to impose that unobserved heterogeneity is continuously distributed on its entire support. Given the continuity condition, the equality $q_{U|Z}(\tau|z) = c$ is equivalent to

$$G_{U|Z}((-\infty, t]\,|z) = \tau \quad \Leftrightarrow \quad t = c.$$

The inequalities comprising (3.16) then follow from $\underline{u}(Y, Z; h) \leq U \leq \overline{u}(Y, Z; h)$. These inequalities also comprise the containment functional inequality $C_h(\mathcal{S}|z) \leq G_{U|Z}(\mathcal{S}|z)$ applied to test sets $\mathcal{S} = (-\infty, c]$ and $\mathcal{S} = [c, \infty)$ for any $c \in \mathcal{C}$. In the proof of Theorem 6 we show that for any h and any c, if the containment functional inequalities hold for these two test sets, then we can find an admissible collection of conditional distributions $\mathcal{G}_{U|Z}$ such that it holds for all closed test sets in $\mathcal{R}_U$. From Corollary 1 it follows that the characterization (3.16) is sharp.

This result helps to illustrate the relative identifying power under Restriction QI as compared to Restrictions SI. Under Restriction SI, $U \perp\!\!\!\perp Z$, to characterize $\mathcal{H}^*$ it suffices to consider $C_h(\mathcal{S}|z) \leq G_U(\mathcal{S})$ where $G_{U|Z}(\mathcal{S}|z) = G_U(\mathcal{S})$ for all core-determining test sets delivered by Theorem 3, which imply that it will hold for all closed subsets of $\mathcal{R}_U$. Under restriction QI, for each $c \in \mathcal{C}$, it is enough to consider $C_h(\mathcal{S}|z) \leq G_{U|Z}(\mathcal{S}|z)$ for only two test sets, namely $\mathcal{S} = (-\infty, c]$ and $\mathcal{S} = [c, \infty)$. In the important special case where $\mathcal{C}$ is the singleton set $\{0\}$, the only test sets required are $\mathcal{S} = (-\infty, 0]$ and $\mathcal{S} = [0, \infty)$

The second part of Theorem 6 follows because when $\underline{u}(Y, Z; h)$ and $\overline{u}(Y, Z; h)$ are continuous, the inequalities in (3.16) which involve cumulative distributions $F_{Y|Z}[\cdot|z]$ may be inverted. Then $\mathcal{H}^*$ may be equivalently expressed as inequalities involving the lower and upper envelopes, $\underline{q}(\tau, z; h)$ and $\overline{q}(\tau, z; h)$, respectively, of conditional quantile functions for selections of $\mathcal{U}(Y, Z; h)$. Finally, as was the case for identified sets $\mathcal{M}^*$ using conditional mean restrictions given in Theorem 5, the third part of Theorem 6 states that the identified set of structures $\left(h, \mathcal{G}_{U|Z}\right)$ are elements of $\mathcal{H}^*$ paired with distributions $G_{U|Z}(\cdot|z)$ that are selectionable with respect to the conditional distribution of $\mathcal{U}(Y, Z; h)$ given $Z = z$, a.e. $z \in \mathcal{R}_Z$.

4 Illustration: A Model with an Interval Censored Endogenous Variable

In this Section we return to Example 5 from Section 2.1, a generalization of a single equation model with an interval censored exogenous variable studied by Manski and Tamer (2002). Like Manski and Tamer (2002) we impose no assumption on the censoring process or the realization of the censored variable relative to the observed interval, but we allow the interval censored explanatory variable as well as the endpoints of the censoring interval to be endogenous. We consider the identifying power of various restrictions on the joint distribution of unobservable U and observed exogenous variables Z, and we provide numerical illustrations of identified sets given particular data generating structures. Models allowing censored *outcome* variables with *uncensored* endogenous explanatory variables with sufficient conditions for point identification include those of Hong and Tamer (2003) and Khan and Tamer (2009).

4.1 Identified Sets

The continuously distributed outcome of interest Y_1 is determined by the realizations of endogenous $Y_2^* \in \mathbb{R}$, exogenous $Z = (Z_1, Z_2) \in \mathbb{R}^{k_z}$, and unobservable variable $U \in \mathbb{R}$ with strictly monotone CDF $\Lambda(\cdot)$, such that

$$Y_1 = g(Y_2^*, Z_1, U), \tag{4.1}$$

where the function $g(\cdot,\cdot,\cdot)$ is increasing in its first argument, and strictly increasing in its third argument.[21] The endogenous variable Y_2^* is not observed, but there are observed variables Y_{2l}, Y_{2u} such that

$$Y_2^* = Y_{2l} + W \times (Y_{2u} - Y_{2l}), \tag{4.2}$$

for some unobserved variable $W \in [0,1]$. There is no restriction on the distribution of W on the unit interval, and no restriction on its stochastic relation to observed variables. Together (U, W) comprise a two-dimensional vector of unobserved heterogeneity.

Since nothing is assumed about the censoring process, it is convenient to suppress the unobserved variable W by replacing (4.2) with the equivalent formulation

$$\mathbb{P}[Y_{2l} \le Y_2^* \le Y_{2u}] = 1. \tag{4.3}$$

The researcher observes a random sample of observations of (Y_1, Y_{2l}, Y_{2u}, Z).

[21]It is important to note here that U is *marginally* distributed with CDF $\Lambda(\cdot)$. At this point, we have yet to impose restrictions on the joint distribution of (U, Z), so that for any $z \in \mathcal{R}_U$, The conditional CDF of $U|Z = z$ need not be $\Lambda(\cdot)$. It is straightforward to allow $g(y_2^*, z_1, u)$ increasing or decreasing in y_2^* for all (z_1, u), but we maintain that $g(y_2^*, z_1, u)$ is increasing to simplify the exposition.

The structural function is

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

and $\mathbb{P}[h(Y,Z,U)=0] = 1$ is equivalent to equations (4.1) and (4.3). The level sets in Y-space and U-space, respectively, are

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

and

$$\mathcal{U}(y,z;h) = \left[g^{-1}(y_{2u}, z_1, y_1), g^{-1}(y_{2l}, z_1, y_1)\right] , \tag{4.4}$$

where g^{-1} denotes the inverse of g in its last argument.

In some of the following developments and indeed in our numerical illustrations we further restrict h, employing the commonly used linear index structure with additive unobservable. To do so we let

$$g(y_2^*, z, u) = \beta y_2^* + z_1\gamma + u, \tag{4.5}$$

where the first element of z_1 is one, and g (and hence h) are now parameterized by $(\beta, \gamma')' \in \mathbb{R}^{\dim(z_1)+1}$.

We now provide the identified set delivered by our results when applied to this model under alternative restrictions on the collection of conditional distributions $\mathcal{G}_{U|Z}$. For each restriction considered, we show how this representation can be characterized by a conditional moment inequality representation that can be used as a basis for estimation and inference.

4.1.1 Stochastic Independence

Consider the restriction $U \perp\!\!\!\perp Z$. Each set $\mathcal{U}(y,z;h)$ is a closed interval on $\mathbb{R}$ and hence connected. Using Theorem 3 we can express the identified set for h as

$$\mathbb{P}[\mathcal{U}(Y,Z;h) \subseteq \mathcal{S}|z] \leq G_U(\mathcal{S}) \tag{4.6}$$

for all $\mathcal{S} \in \mathsf{Q}(h,z)$, where $\mathsf{Q}(h,z)$ is the collection of 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 $\mathbb{R}$.[22] Unless g has very restricted structure, the conditions for (4.6) to hold with equality will in general not be

[22]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.

satisfied for any test set $\mathcal{S}$, and hence $\mathsf{Q}^E(h,z) = \emptyset$ and $\mathsf{Q}^I(h,z) = \mathsf{Q}(h,z)$ is the collection of all intervals on $\mathbb{R}$, which we henceforth denote

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

Thus we have from Theorem 4 that the identified set is

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

Given structural function h, the probability $\mathbb{P}[\mathcal{U}(Y,Z;h) \subseteq [u_*, u^*] | z]$ is an event concerning only observed variables, and is thus identified. Specifically, the containment functional inequality in the definition of $\mathcal{H}^*$ can be equivalently written

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

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

$$\mathbb{P}[g(Y_{2u}, Z_1, u_*) \leq Y_1 \leq g(Y_{2l}, Z_1, u^*) | z] \leq \Lambda(u^*) - \Lambda(u_*). \tag{4.7}$$

With the added linear index restriction from (4.5) this produces the following representation for the identified set, where the model $\mathcal{M}$ stipulates a collection of admissible parameters β, γ and CDFs $\Lambda(\cdot)$.

$$\mathcal{M}^* = \left\{ \begin{array}{l} m \in \mathcal{M} : \forall [u_*, u^*] \in \mathsf{Q}, \\ \quad \mathbb{P}[u_* + \beta Y_{2u} \leq Y_1 - Z_1\gamma \leq u^* + \beta Y_{2l} | z] \leq \Lambda(u^*) - \Lambda(u_*),\ \text{a.e. } z \in \mathcal{R}_Z \end{array} \right\}. \tag{4.8}$$

We now specialize this result for a model incorporating a parametric restriction for Λ, and for a model leaving Λ completely unspecified.

Gaussian Unobservables

Suppose in addition to the linear index restriction (4.5) we further restrict $\Lambda(\cdot)$ to be a Gaussian CDF with variance $\sigma > 0$ so that $\Lambda(u) = \Phi(\sigma^{-1}u)$, where $\Phi(\cdot)$ is the standard normal CDF. In this case the model is fully parameterized by $\theta \equiv (\beta, \gamma', \sigma)'$, and $\mathcal{M}$ can be represented as the parameter space Θ for admissible θ. Using (4.8) the identified set is now

$$\mathcal{M}^* = \left\{ \begin{array}{l} \theta \in \Theta : \forall [u_*, u^*] \in \mathsf{Q}, \\ \quad \mathbb{P}[u_* + \beta Y_{2u} \leq Y_1 - Z_1\gamma \leq u^* + \beta Y_{2l} | z] \leq \Phi(\sigma^{-1}u^*) - \Phi(\sigma^{-1}u_*),\ \text{a.e. } z \in \mathcal{R}_Z \end{array} \right\}. \tag{4.9}$$

Equivalently, we can employ the change of variables $t^* = \Phi(\sigma^{-1}u^*)$ and $t_* \equiv \Phi(\sigma^{-1}u_*)$ to produce the following.[23]

$$\mathcal{M}^* = \left\{ \begin{array}{l} \theta \in \Theta : \forall \left[t_*, t^*\right] \subseteq [0,1], \\ \quad \mathbb{P}\left[t_* \leq \Phi\left(\frac{Y_1 - \beta Y_{2u} - Z_1\gamma}{\sigma}\right) \wedge \Phi\left(\frac{Y_1 - \beta Y_{2l} - Z_1\gamma}{\sigma}\right) \leq t^* | z\right] \leq t^* - t_*, \text{ a.e. } z \in \mathcal{R}_Z \end{array} \right\}. \quad (4.10)$$

Using (4.9), 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

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

Distribution-Free Unobservables

Suppose now that we impose the independence restriction $U \perp\!\!\!\perp Z$ and the same additive index structure for g, but without imposing a parametric restriction on unobserved heterogeneity. If Y_2^* were observed, we would require a location normalization for identification of the first component of γ, the intercept. Thus it will be prudent to incorporate a location normalization in our model with Y_2^* censored as well, for example that the median of $U|Z = z$ is zero. Since Y_1 is continuously distributed, there is no scale normalization to be made.

We apply Corollary 3 to obtain the identified set for h, equivalently that for parameters $\theta \equiv (\beta, \gamma')'$. To do so, we start with

$$\mathbb{P}\left[u_* + \beta Y_{2u} \leq Y_1 - Z_1\gamma \leq u^* + \beta Y_{2l}|z\right] \leq \Lambda\left(u^*\right) - \Lambda\left(u_*\right), \text{ a.e. } z \in \mathcal{R}_Z, \quad (4.11)$$

for all $[u_*, u^*] \in \mathsf{Q}$ from (4.8) above. Noting that $G_U(\mathcal{S}) = \Lambda(u^*) - \Lambda(u_*)$ for any set $\mathcal{S} = [u_*, u^*]$ and following Corollary 3 we also have for all $-\infty < u_* \leq u^* < \infty$ and a.e. $z \in \mathcal{R}_Z$,

$$\begin{aligned} \Lambda\left(u^*\right) - \Lambda\left(u_*\right) &\leq 1 - C_h\left(\mathcal{S}^c|z\right) \quad (4.12) \\ &= 1 - \mathbb{P}\left[Y_1 - \beta Y_{2l} - Z_1\gamma < u_* \vee Y_1 - \beta Y_{2u} - Z_1\gamma > u^*|z\right] \\ &= \mathbb{P}\left[u_* + \beta Y_{2l} \leq Y_1 - Z_1\gamma \leq u^* + \beta Y_{2u}|z\right]. \end{aligned}$$

[23]This can also be derived by normalizing the distribution of unobserved heterogeneity U to be uniform on the unit interval and defining $g(y_2, z_1, u) = \beta y_2 + z_1\gamma + \sigma\Phi^{-1}(u)$.

Define now

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

each of which are identified for any parameter vector $\theta = (\beta, \gamma')'$ from knowledge of $\mathcal{F}_{Y|Z}$ under Restriction A2. Combining (4.11) and (4.12) as in Corollary 3, the identified set for parameters θ, where Θ denotes values admitted by model $\mathcal{M}$, is given by

$$\Theta^* = \left\{\theta \in \Theta : \forall\left[u_*, u^*\right] \in \mathsf{Q},\ \underline{G}(\theta, u_*, u^*) \leq \overline{G}(\theta, u_*, u^*)\right\}. \tag{4.13}$$

The identified set for $(\theta, \Lambda(\cdot))$ is

$$\mathcal{M}^* = \left\{(\theta, \Lambda(\cdot)) \in \mathcal{M} : \forall\left[u_*, u^*\right] \in \mathsf{Q},\ \underline{G}(\theta, u_*, u^*) \leq \Lambda(u^*) - \Lambda(u_*) \leq \overline{G}(\theta, u_*, u^*)\right\}.$$

Equivalent to (4.13), the identified set for θ are those $\theta \in \Theta$ satisfying the moment inequality representation:

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

where

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

4.1.2 Mean Independence

Now suppose we continue to assume the linear index structure (4.5) but replace the restriction $U \perp\!\!\!\perp Z$ with the conditional mean restriction $E[U|Z=z] = 0$ a.e. $z \in \mathcal{R}_Z$, equivalently Restriction MI from Section 3.4 with $\mathcal{C} = \{0\}$.

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

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

rendering application of Theorem 5 and Corollary 4 particularly simple. This is because there exists a function $u(\cdot,\cdot)$ satisfying the conditions of Corollary 4, namely that (i) $E[u(Y, Z)|z] = 0$

a.e. $z \in \mathcal{R}_Z$, and (ii) $\mathbb{P}[u(Y,Z) \in \mathcal{U}(Y,Z;h)] = 1$ if and only if

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

Thus, applying Corollary 4, the identified set for $\theta \equiv (\beta, \gamma')'$, where again Θ denotes values admitted by model $\mathcal{M}$, is

$$\Theta^* = \left\{\theta \in \Theta : \underline{\mathbb{E}}(\theta) \leq 0 \leq \overline{\mathbb{E}}(\theta)\right\},$$

where

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

4.1.3 Quantile Independence

Finally, we consider the linear index structure (4.5) coupled with Restriction QI with $\mathcal{C} = \{0\}$. That is, we now assert $q_{U|Z}(\tau|z) = 0$, a.e $z \in \mathcal{R}_Z$.

Again we have under (4.5) that

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

and the identified set for h is isomorphic to that of $\theta \equiv (\beta, \gamma')'$. As in Section 4.1.2 we again denote the parameter space and identified set for θ as Θ and Θ^*, respectively. Applying Theorem 6 the identified set of structural functions h is

$$\Theta^* = \left\{\theta \in \Theta : \sup_{z \in \mathcal{R}_Z} F_{Y|Z}[Y_1 \leq Z_1\gamma + \beta Y_{2l}|z] \leq \tau \leq \inf_{z \in \mathcal{R}_Z} F_{Y|Z}[Y_1 \leq Z_1\gamma + \beta Y_{2u}|z]\right\}, \qquad (4.14)$$

equivalently,

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

where $\underline{V_\theta} \equiv Y_1 - \beta Y_{2u}$ and $\overline{V_\theta} \equiv Y_1 - \beta Y_{2l}$. The identified set of structures $\mathcal{M}^*$ is then pairs of structural functions h parameterized by $\theta \in \Theta^*$ coupled with collections of conditional distributions $\mathcal{G}_{U|Z}$ satisfying the required conditional quantile restriction, and such that $G_{U|Z}(\cdot|z)$ is selectionable with respect to $\mathcal{U}(Y,Z;h)$ conditional on $Z = z$, a.e. $z \in \mathcal{R}_Z$.

Using (4.14) we can represent Θ^* via the moment inequalities

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

where

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

4.2 Numerical Illustrations

In this section we provide illustrations of identified sets obtained for the interval censored endogenous variable model with the linear index restriction of (4.5). We first consider the identified set obtained under the restriction that $U \sim N(0, \sigma)$ and $U \perp\!\!\!\perp Z$, i.e. the Gaussian unobservable case above with identified set given by (4.9).

To generate probability distributions $\mathcal{F}_{Y|Z}$ for observable variables (Y, Z) we employ a triangular Gaussian structure as follows.

$$Y_1 = \gamma_0 + \gamma_1 Y_2^* + U,$$
$$Y_2^* = \delta_0 + \delta_1 Z + V.$$

with $(U, V) \perp\!\!\!\perp Z$, $\mathcal{R}_Z = \{-1, 1\}$, and

$$\begin{bmatrix} U \\ V \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{11} & \sigma_{1v} \\ \sigma_{1v} & \sigma_{vv} \end{bmatrix}\right).$$

In this model there are no exogenous covariates Z_1, equivalently $Z = Z_2$. The case where Z has binary support is easy to analyze for the sake of numerical illustration, but a richer support for the instrument would provide greater identifying power, equivalently smaller identified sets.

We specify a censoring process that in place of Y_2^* reveals only 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 assume a sequence of J 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

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

As above, the researcher observes a random sample of observations of (Y_1, Y_{2l}, Y_{2u}, Z).

In our first set of examples we consider two data generation processes denoted DGP1 and DGP2, each with parameter values

$$\gamma_0 = 0, \quad \gamma_1 = 1, \quad \delta_0 = 0, \quad \delta_1 = 1, \quad \sigma_{11} = 0.5, \quad \sigma_{1v} = 0.25, \quad \sigma_{vv} = 0.5, \tag{4.15}$$

	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}
DGP1	$-\infty$	-1.15	-0.67	-0.32	0.00	0.32	0.67	1.15	$+\infty$	-	-	-	-
DGP2	$-\infty$	-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.

and interval censoring endpoints $c_1, ..., c_J$ listed in Table 2. In DGP1, Y_2^* was 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 DGP2, Y_2^* was censored into 12 such intervals with endpoints given by the normal quantile function evaluated at 13 equally spaced values.

Given these data generation processes, the distribution of $Y \equiv (Y_1, Y_{2l}, Y_{2u})$ conditional on Z is easily obtained as the product of the conditional distribution of (Y_{2l}, Y_{2u}) given Y_1 and Z and the distribution of Y_1 given Z. Combining these probabilities and the inequalities of (4.9), the conditional containment functional for random set $\mathcal{U}(Y, Z; h)$ applied to test set $\mathcal{S} = [u_*, u^*]$ is given by

$$C_\theta([u_*, u^*]|z) = \sum_j \mathbb{P}[g_1 c_{j+1} + u_* \le Y_1 - g_0 \le g_1 c_j + u^* | z, [Y_{2l}, Y_{2u}) = I_j] * \mathbb{P}[[Y_{2l}, Y_{2u}) = I_j | z], \tag{4.16}$$

where $\theta = (g_0, g_1, s)$ is used to denote generic parameter values for $(\gamma_0, \gamma_1, \sigma_1)$. C_θ replaces C_h for the containment functional, since in this model the structural function h is a known function of θ.[24] The identified set of structures $\left(h, \mathcal{G}_{U|Z}\right)$ is completely determined by the identified set for θ, which, following (4.10), is given by

$$\Theta^* = \left\{ \begin{array}{l} \theta \in \Theta : \forall [t_*, t^*] \subseteq [0, 1], \\ C_\theta\left(\left[s\Phi^{-1}(t_*), s\Phi^{-1}(t^*)\right]|z\right) \le t^* - t_*, \text{ a.e. } z \in \mathcal{R}_Z \end{array} \right\}. \tag{4.17}$$

The set Θ^* comprises parameter values (g_0, g_1, s) such that the given conditional containment functional inequality holds for almost every z and *all* intervals $[t_*, t^*] \subseteq [0, 1]$. This collection of test sets is uncountable. For the purpose of illustration we used various combinations of collections Q_M of intervals from the full set of all possible $[t_*, t^*] \subseteq [0, 1]$. Each collection of intervals Q_M comprises the $M(M+1)/2 - 1$ *super-diagonal* elements of the following $(M+1) \times (M+1)$ array

[24]Computational details for the conditional containment probability $C_\theta([u_*, u^*]|z)$ are provided in Appendix C.

of intervals that remain after excluding the interval $[0,1]$. Here $m \equiv 1/M$.

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

The inequalities of (4.17) applied to the intervals of any collections of test sets Q_M defines an outer region for the identified set, with larger collections of test sets providing successively better approximations of the identified set.

Figure 1 shows three dimensional (3D) plots of outer regions for (g_0, g_1, s). Outer regions using $M \in \{5,7,9\}$ are noticeably smaller than those using only $M = 5$.[25] There was a noticeable reduction in the size of the outer region in moving from $M = 5$ to $M = \{5,7\}$, but hardly any change on including also the inequalities obtained with $M = 9$. Thus, only the outer regions obtained using $M = 5$ and $M \in \{5,7,9\}$ are shown. Figure 2 shows two dimensional projections of the outer region using $M \in \{5,7,9\}$ for each pair of the three parameter components. The surfaces of these sets were drawn as convex hulls of those points found to lie inside the outer regions and projections considered.[26] We have no proof of the convexity of the outer regions in general, but careful investigation of points found to lie in the outer regions strongly suggested that in the cases considered the sets are convex.

Figure 3 shows the 3D outer region for DGP2 employing 12 bins for the censoring of Y_2^* and $M \in \{5,7,9\}$. Compared to Figure 1, this outer region is smaller, as expected given the finer granularity of intervals with 12 rather than 8 bins. Figure 4 shows two dimensional projections for this outer region, again projecting onto each pair of parameter components. These projections help to further illustrate the extent of the reduction in the size of the outer region for DGP2 relative to DGP1.

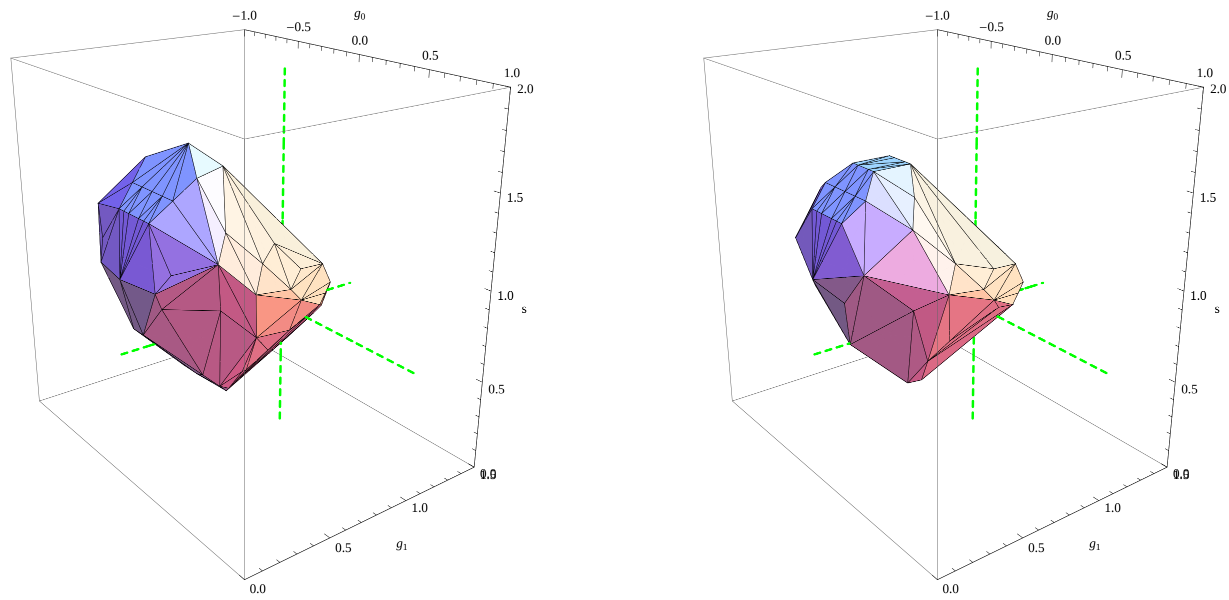
In our second set of numerical illustrations we employ the same triangular Gaussian error structure for our DGPs with parameter values as specified in (4.15). However, we consider two alternative censoring processes, where Y_2^* is again observed only to lie in one of a fixed set of bins, but where now these bins are set to be of a fixed width. We consider fixed bins, first with width 0.4:

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

[25]The notation $M \in \{m_1, m_2, ..., m_R\}$ corresponds to the use of test sets $\mathsf{Q}_{m_1} \cup \mathsf{Q}_{m_2} \cdots \cup \cdots \cup \mathsf{Q}_{m_R}$.

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

Figure 1: Outer regions for parameters (g_0, g_1, s) for DGP1 with 8 bins using the 14 inequalities generated with $M = 5$ (left pane) and the 85 inequalities generated with $M \in \{5, 7, 9\}$ (right pane).



and then of width 0.2:

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

With this censoring structure in place, we now compare the identifying power of alternative restrictions on unobserved heterogeneity, in both cases imposing the linear functional form

$$Y_1 = \gamma_0 + \gamma_1 Y_2^* + U.$$

We again consider the parametric Gaussian restriction on unobserved heterogeneity that $U \sim N(0, \sigma)$ and $U \perp\!\!\!\perp Z$, and compare to the alternative restriction imposing only knowledge that $q_{U|Z}(0.5|z) = 0$, a.e. $z \in \mathcal{R}_Z$. This semiparametric specification has no scale parameter s, so we focus attention on the implied identified set for (γ_0, γ_1).

Figure 5 below illustrates the identified sets obtained for bin widths 0.4 (top panels) and 0.2 (bottom panels), as well as for $\delta_1 = 1$ (left panels) and $\delta_1 = 1.5$ (right panels). In the triangular structure employed to generate the actual distributions $\mathcal{F}_{Y|Z}$ this parameter 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 δ the value of this variable as well as the censoring points is more sensitive with respect to the instrument Z. As we might expect, identified sets when $\delta_1 = 1.5$ are smaller than those for the case $\delta_1 = 1$, as are sets obtained when the bin width is only 0.2 rather than 0.4.

Figure 2: Outer region projections for DGP1 onto the (g_0, g_1), (g_0, s), and (g_1, s) planes, respectively, with 8 bins using inequalities generated with $M \in \{5, 7, 9\}$. The red point marks the data generating value.

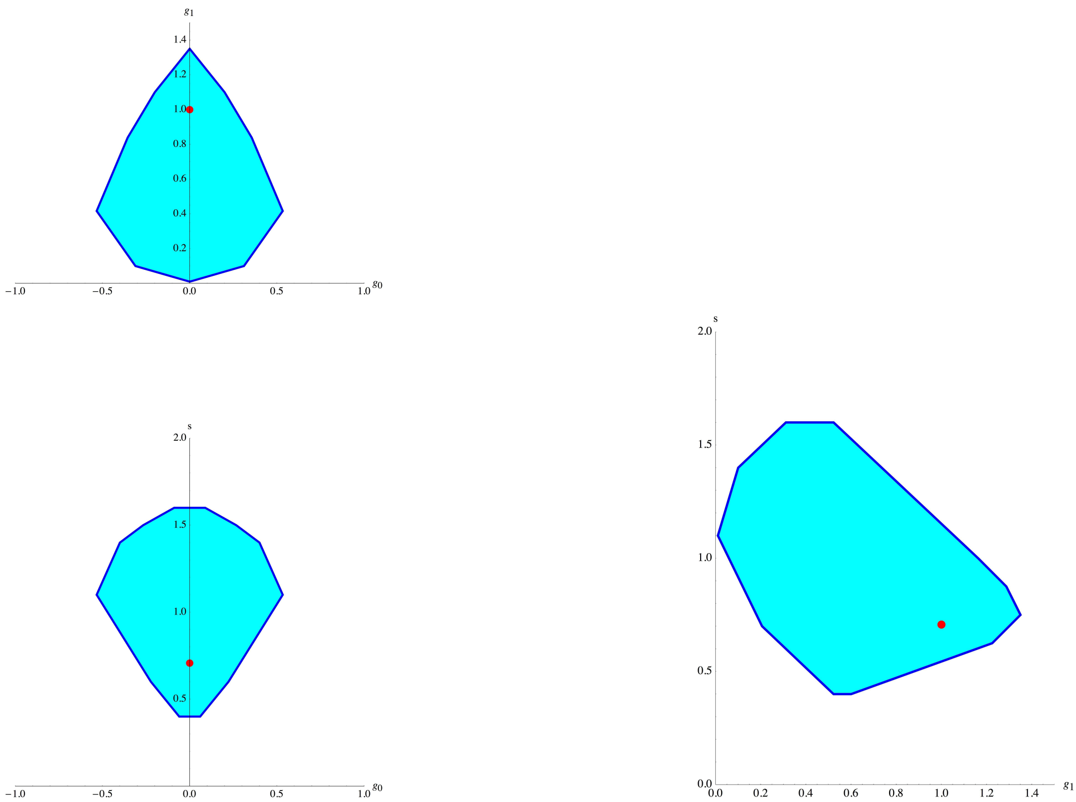


Figure 3: Outer region for DGP2 with 12 bins calculated using inequalities generated with $M \in \{5, 7, 9\}$. Dashed green lines intersect at the data generating value of the parameters.

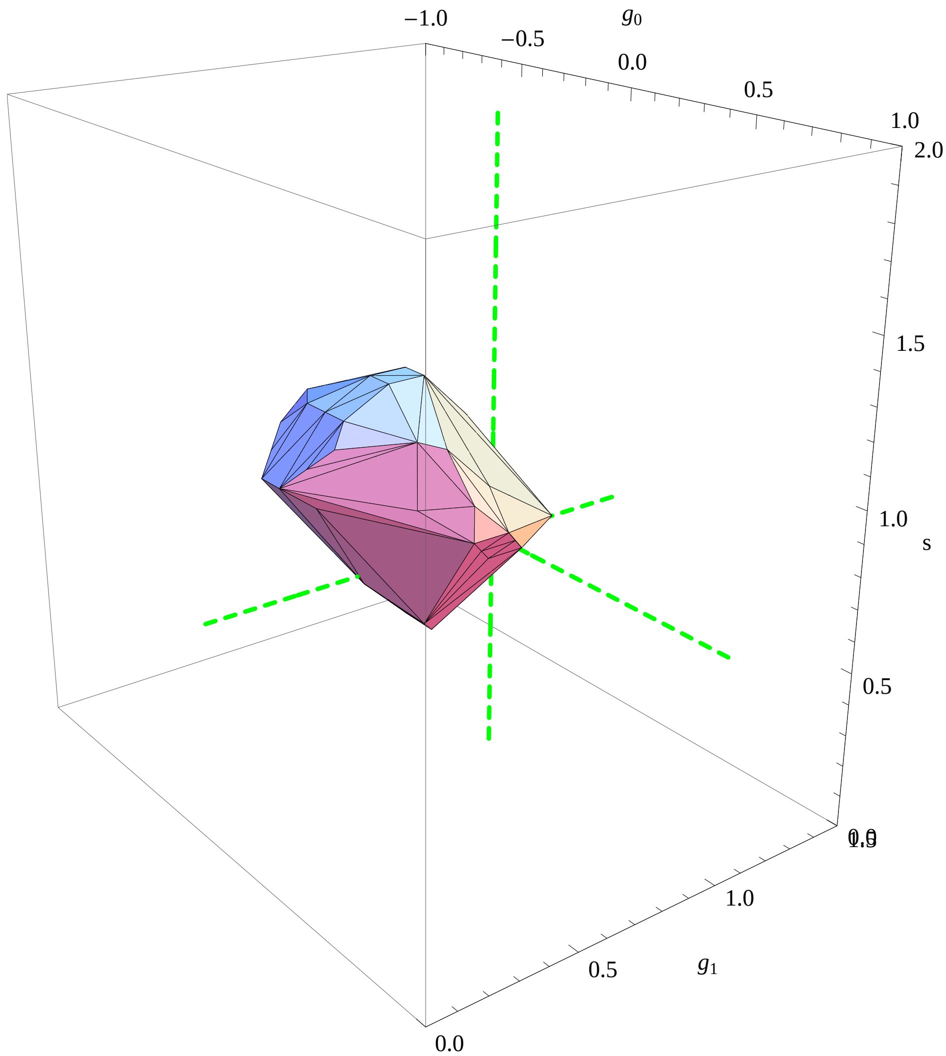
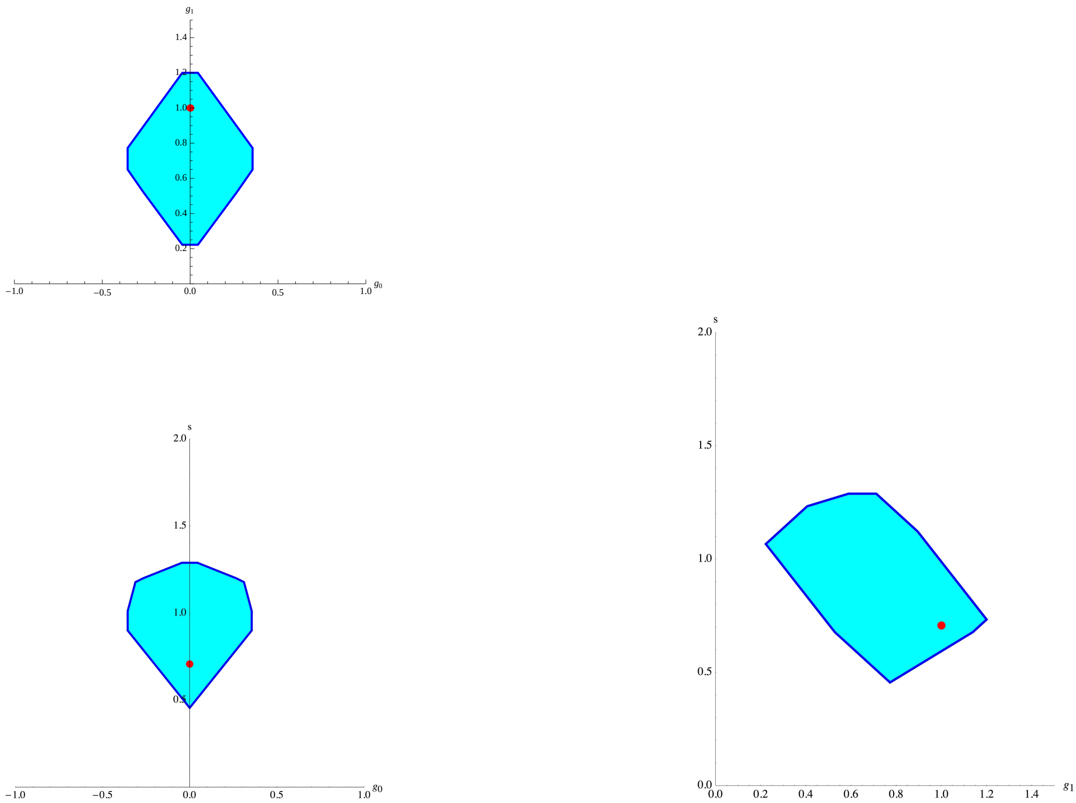


Figure 4: Outer region projections for DGP2 onto the (g_0, g_1), (g_0, s), and (g_1, s) planes, respectively, with 12 bins using inequalities generated with $M \in \{5, 7, 9\}$. The red point marks the data generating value.



Identified sets obtained from a model imposing independent Gaussian unobservable U (in light blue) are of course contained in those obtained from a model only imposing the less restrictive zero conditional quantile restriction (in dark blue). However, the difference between the identified sets obtained under these different restrictions is not so great, at least under the particular data generating structures employed. In these cases, the use of the weaker conditional quantile restriction does not seem to lose much in the way of identifying power relative to the Gaussian distributional restriction.

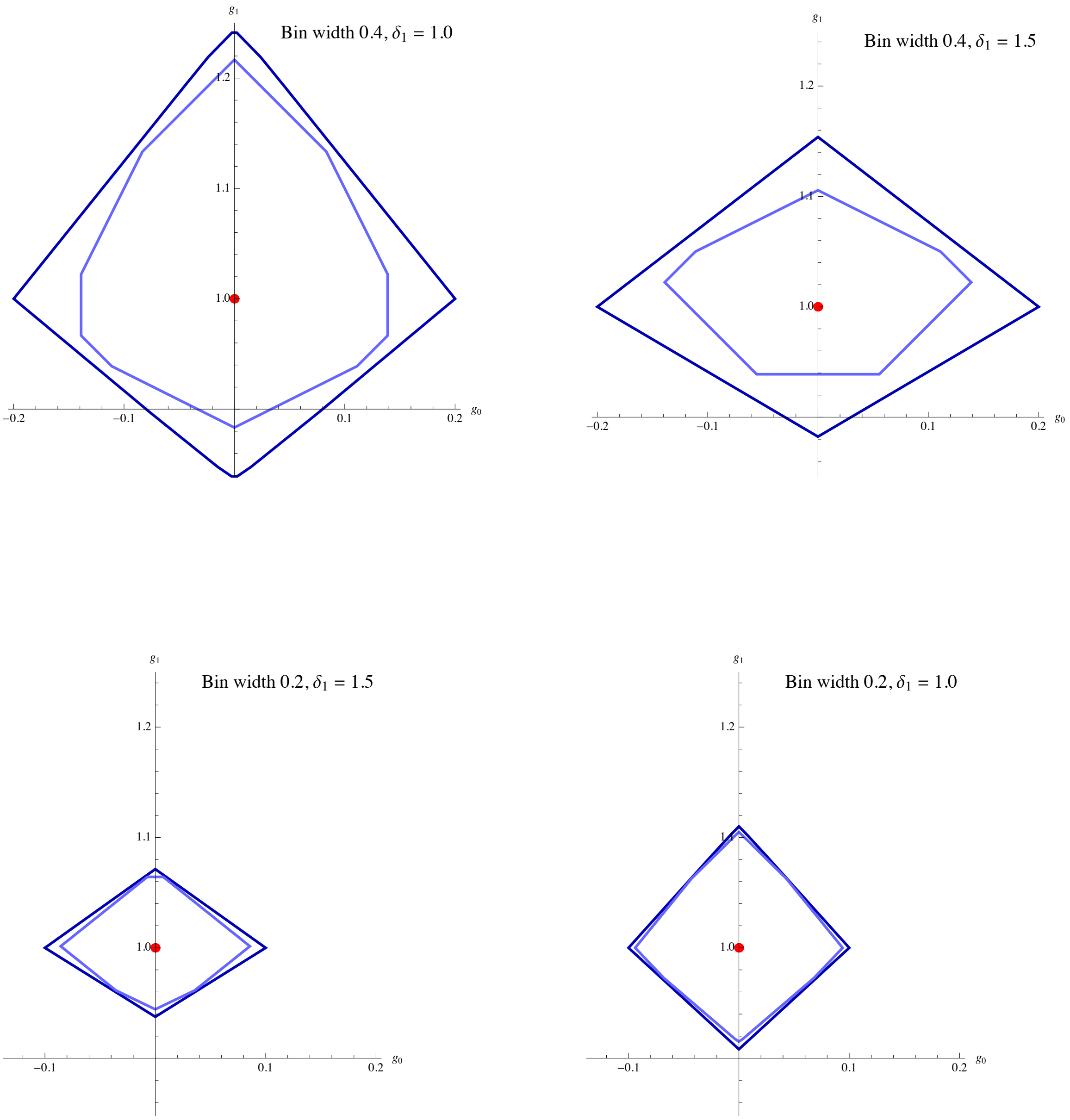
A partial explanation for this observation may be the fixed-width binning setup. Indeed, under this censoring process with the given triangular data-generating structure, it can be shown that a distribution-free independence restriction - that is $U \perp\!\!\!\perp Z$ but with the distribution of U 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 however not a general observation. For other censoring processes (not reported here) the identified set under the distribution-free independence restriction is a strict subset of that obtained under only the conditional quantile restriction, as would be expected.

5 Conclusion

In this paper we have studied a broad class of Generalized Instrumental Variable (GIV) models, extending the use of instrumental variable restrictions to models with more flexible specifications for unobserved heterogeneity than previously allowed. In particular, our analysis allows for the application of instrumental variable restrictions to models in which the structural mapping from unobserved heterogeneity to observed endogenous variables is not invertible. These models permit general forms of multivariate, nonadditive unobserved heterogeneity, as often appears for example in nonlinear models with discrete or mixed discrete/continuous outcomes, random coefficient models, and models with censoring. Without the existence of a unique value of unobserved heterogeneity, or "generalized residual", given values of observed variables, rank or more generally completeness conditions do not guarantee point identification of structural functions. We provided a comprehensive framework for characterizing identified sets for structures and structural features in such contexts.

Using tools from random set theory, relying in particular on the concept of *selectionability*, we formally extended the classical notion of observational equivalence to models whose structures are not required to deliver a unique conditional distribution for endogenous variables given exogenous variables. We showed that the closely related definition of a model's identified set of structures may be equivalently formulated in terms of selectionability criteria in the space of unobserved heterogeneity. This formulation enables direct incorporation of restrictions on conditional distributions of unobserved heterogeneity, of the sort typically employed in econometric models, as we demon-

Figure 5: Identified sets for (γ_0, γ_1). The top panels display sets for bins of width 0.4 and the bottom panels display sets for bins of width 0.2. In the panels on the left $\delta_1 = 1$ and on the right $\delta_1 = 1.5$. The dark blue lines indicate boundaries of identified sets obtained with the conditional quantile restriction $q_{U|Z}(\tau|z) = 0$, while the inner light blue lines indicate boundaries of identified sets when U is restricted to be Gaussian, independent of Z.



strated by characterizing identified sets under stochastic independence, mean independence, and quantile independence restrictions. We specialized these characterizations to a model with interval censored explanatory variables, with a censoring and linear index structure following Manski and Tamer (2002), but without the requirement that censored variables be exogenous. We provided numerical illustrations of identified sets in such models under a stochastic independence restriction between instrumental variables and unobserved heterogeneity.

We showed that all of our characterizations of identified sets can be written as systems of conditional moment inequalities. These can be employed for estimation and inference using a variety of approaches from the recent literature. Which of these approaches are most suitable will depend on the context. With a discrete conditioning variable the identified sets derived in Section 3 can be expressed using unconditional moment inequality representations, for example as in Chernozhukov, Hong, and Tamer (2007), Beresteanu and Molinari (2008), Romano and Shaikh (2008a,b), Rosen (2008), Andrews and Guggenberger (2009), Andrews and Soares (2010), Andrews and Jia-Barwick (2010), Bugni (2010), Canay (2010), and Romano, Shaikh, and Wolf (2013). With a continuous conditioning variable inference using conditional moment inequalities can be performed, see for example Andrews and Shi (2013a,b), Chernozhukov, Lee, and Rosen (2013), Lee, Song, and Whang (2013a,b), Armstrong(2011a,b), and Chetverikov (2011).

In many models, for example those employing conditional mean or conditional quantile restrictions on unobserved heterogeneity, existing methods such as those above can be directly applicable. In other models, in particular those with stochastic independence restrictions where the number of core determining sets can be extremely large, the number of conditional inequality restrictions characterizing the identified set may be enormous relative to sample size. This raises complications both in terms of computation and the quality of asymptotic approximations in finite samples for inference methods based on inequality restrictions. Different aspects of such problems have been investigated by Menzel (2009) for set estimation and Chernozhukov, Chetverikov, and Kato (2013) for statistical inference, who provide novel inference results in moment inequality models where the number of inequalities can be extremely large. As they note, their results can also be applied to conditional moment inequality models with a relatively small number of continuous conditioning variables. We believe there could potentially be a high payoff to further research on inference based on extremely large numbers of conditional moment inequalities relative to sample size and on inference based on other characterizations of the selectionability property that determine identified sets.

References

Andrews, D. W. K. (2011): "Examples of L2-complete and Boundedly-Complete Distributions," Cowles Foundation Discussion Paper.

Andrews, D. W. K., and P. Guggenberger (2009): "Validity of Subsampling and Plug-In Asymptotic Inference for Parameters Defined by Moment Inequalities," *Econometric Theory*, 25, 669–709.

Andrews, D. W. K., and P. Jia-Barwick (2010): "Inference for Parameters Defined by Moment Inequalities: A Recommended Moment Selection Procedure," *Econometrica*, 80(6), 28052826.

Andrews, D. W. K., and X. Shi (2013a): "Inference Based on Conditional Moment Inequalities," *Econometrica*, 81(2), 609–666.

——— (2013b): "Nonparametric Inference Based on Conditional Moment Inequalities," *Journal of Econometrics*, forthcoming.

Andrews, D. W. K., and G. Soares (2010): "Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection," *Econometrica*, 78(1), 119–157.

Aradillas-Lopez, A., and A. M. Rosen (2013): "Inference in Ordered Response Games with Complete Information," CEMMAP working paper CWP33/13.

Armstrong, T. B. (2011a): "Asypmtotically Exact Inference in Conditional Moment Inequalities," working paper, Yale University.

——— (2011b): "Weighted KS Statistics for Inference on Conditional Moment Inequalities," working paper, Yale University.

Artstein, Z. (1983): "Distributions of Random Sets and Random Selections," *Israel Journal of Mathematics*, 46(4), 313–324.

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.

Beresteanu, A., and F. Molinari (2008): "Asymptotic Properties for a Class of Partially Identified Models," *Econometrica*, 76(4), 763–814.

Blundell, R., X. Chen, and D. Kristensen (2007): "Semi-Nonparametric IV Estimation of Shape-Invariant Engel Curves," *Econometrica*, 75(6), 1613–1669.

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.

BRESNAHAN, T. F., AND P. J. REISS (1990): "Entry in Monopoly Markets," *Review of Economic Studies*, 57, 531–553.

——— (1991): "Empirical Models of Discrete Games," *Journal of Econometrics*, 48(1-2), 57–81.

BUGNI, F. (2010): "Bootstrap Inference for Partially Identified Models Defined by Moment Inequalities: Coverage of the Identified Set," *Econometrica*, 78(2), 735–753.

CANAY, I. (2010): "EL Inference for Partially Identified Models: Large Deviations Optimality and Bootstrap Validity," *Journal of Econometrics*, 156(2), 408–425.

CANAY, I., A. SANTOS, AND A. SHAIKH (2013): "On the Testability of Identification in Some Nonparametric Models with Endogeneity," *Econometrica*, 81(6), 2535–2559.

CHEN, X., V. CHERNOZHUKOV, S. LEE, AND W. NEWEY (2011): "Local Identification of Nonparametric and Semiparametric Models," *Econometrica*, forthcoming. Formerly CeMMAP working paper CWP17/11.

CHERNOZHUKOV, V., D. CHETVERIKOV, AND K. KATO (2013): "Testing Many Moment Inequalities," CeMMAP working paper CWP65/13.

CHERNOZHUKOV, V., AND C. HANSEN (2005): "An IV Model of Quantile Treatment Effects," *Econometrica*, 73, 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., G. IMBENS, AND W. NEWEY (2007): "Instrumental Variable Estimation of Nonseparable Models," *Journal of Econometrics*, 139, 4–14.

CHERNOZHUKOV, V., S. LEE, AND A. 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.

CHESHER, A., AND A. ROSEN (2012a): "An Instrumental Variable Random Coefficients Model for Binary Outcomes," *Econometrics Journal*, forthcoming. Formerly CeMMAP working paper CWP34/12.

——— (2012b): "Simultaneous Equations Models for Discrete Outcomes: Coherence, Completeness, and Identification," CeMMAP working paper CWP21/12.

Chesher, A., A. Rosen, and K. Smolinski (2013): "An Instrumental Variable Model of Multiple Discrete Choice," *Quantitative Economics*, 4(2), 157–196.

Chesher, A., and A. M. Rosen (2013a): "Intersection Bound Inference on Partially Identified Average Treatment Effects," in preparation.

——— (2013b): "What Do Instrumental Variable Models Deliver With Discrete Dependent Variables?," *American Economic Reivew: Papers and Proceedings*, 1(3), 557–562.

Chesher, A., and K. Smolinski (2012): "IV Models of Ordered Choice," *Journal of Econometrics*, 166(1), 33–48.

Chetverikov, D. (2011): "Adaptive Test of Conditional Moment Inequalities," CeMMAP working paper CWP36/12.

Ciliberto, F., and E. Tamer (2009): "Market Structure and Multiple Equilibria in Airline Markets," *Econometrica*, 77(6), 1791–1828.

Darolles, S., Y. Fan, J.-P. Florens, and E. Renault (2011): "Nonparametric Instrumental Regression," *Econometrica*, 79(5), 1541–1565.

D'Haultfoeuille, X. (2011): "On the Completeness Condition in Nonparametric Instrumental Problems," *Econometric Theory*, 27, 460–471.

Galichon, A., and M. Henry (2011): "Set Identification in Models with Multiple Equilibria," *Review of Economic Studies*, 78(4), 1264–1298.

Haile, P. A., and E. Tamer (2003): "Inference with an Incomplete Model of English Auctions," *Journal of Political Economy*, 111(1), 1–51.

Hall, P., and J. Horowitz (2005): "Nonparametric Methods for Inference in the Presence of Instrumental Variables," *Annals of Statistics*, 33(6), 2904–2929.

Henry, M., R. Meango, and M. Queyranne (2011): "Combinatorial Bootstrap Inference in Partially Identified Incomplete Structural Models," working paper, University of Montreal.

Hoderlein, S., and E. Mammen (2007): "Identification of Marginal Effects in Nonseparable Models Without Monotonicity," *Econometrica*, 75(5), 1513–1518.

Hong, H., and E. Tamer (2003): "Inference in Censored Models with Endogenous Regressors," *Econometrica*, 71(3), 905–932.

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., AND W. NEWEY (2009): "Identification and Estimation of Triangular Simultaneous Equations Models with Additivity," *Econometrica*, 77(1481-1512), 5.

KHAN, S., AND E. TAMER (2009): "Inference on Endogenously Censored Regression Models Using Conditional Moment Inequalities," *Journal of Econometrics*, 152.

KOOPMANS, T. (1949): "Identification Problems in Economic Model Construction," *Econometrica*, 17(125-144), 2.

KOOPMANS, T., AND O. REIERSØL (1950): "The Identification of Structural Characteristics," *Annals of Mathematical Statistics*, 21(165-181), 1–27.

KOOPMANS, T., H. RUBIN, AND R. LEIPNIK (1950): "Measuring the equation systems of dynamic economics," in *Statistical Inference in Dynamic Economic Models*, ed. by C. C. for Research in Economics, pp. 53–237. John Wiley and Sons.

LEE, S., K. SONG, AND Y.-J. WHANG (2013a): "Testing for a General Class of Functional Inequalities," working paper, Seoul National University.

——— (2013b): "Testing Functional Inequalities," *Journal of Econometrics*, 172(1), 14–32.

MANSKI, C. F. (1985): "Semiparametric Analysis of Discrete Response: Asymptotic Properties of the Maximum Score Estimator," *Journal of Econometrics*, 27, 313–333.

MANSKI, C. F., AND E. TAMER (2002): "Inference on Regressions with Interval Data on a Regressor or Outcome," *Econometrica*, 70(2), 519–546.

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, 945–978.

MENZEL, K. (2009): "Estimation and Inference with Many Weak Moment Inequalities," working paper, MIT.

MOLCHANOV, I. S. (2005): *Theory of Random Sets*. Springer Verlag, London.

NEWEY, W. K., AND J. L. POWELL (2003): "Instrumental Variable Estimation of Nonparametric Models," *Econometrica*, 71, 1565–1578.

Norberg, T. (1992): "On the Existence of Ordered Couplings of Random Sets – with Applications," *Israel Journal of Mathematics*, 77, 241–264.

Romano, J., and A. Shaikh (2008): "Inference for Identifiable Parameters in Partially Identified Econometric Models," *Journal of Planning and Statistical Inference*, 138, 2786–2807.

Romano, J., A. Shaikh, and M. Wolf (2013): "A practical two-step method for testing moment inequalities with an application to inference in partially identified models.," working paper, Stanford University.

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, 107–117.

——— (2012): "Set Identification via Quantile Restrictions in Short Panels," *Journal of Econometrics*, 166(1), 127–137.

Rothenberg, T. J. (1971): "Identification in Parametric Models," *Econometrica*, 39(3), 577–591.

Tamer, E. (2003): "Incomplete Simultaneous Discrete Response Models with Multiple Equilibria," *Review of Economic Studies*, 70(1), 147–167.

Vytlacil, E., and N. Yildiz (2007): "Dummy Endogenous Variables in Weakly Separable Systems," *Econometrica*, 75(757-779), 3.

A Proofs

Proof of Theorem 1. *Fix $z \in \mathcal{R}_Z$ and suppose that $F_{Y|Z}(\cdot|z)$ is selectionable with respect to the conditional distribution of $\mathcal{Y}(U, Z; h)$ given $Z = z$. By Restriction A3, U is conditionally distributed $G_{U|Z}(\cdot|z)$ given $Z = z$, and thus selectionability implies that there exist random variables $\tilde{Y}$ and $\tilde{U}$ such that*

(i) $\tilde{Y}|Z = z \sim F_{Y|Z}(\cdot|z)$,

(ii) $\tilde{U}|Z = z \sim G_{U|Z}(\cdot|z)$,

(iii) $\mathbb{P}\left[\tilde{Y} \in \mathcal{Y}\left(\tilde{U}, Z; h\right)|Z = z\right] = 1$.

By Restriction A3, $\tilde{Y} \in \mathcal{Y}\left(\tilde{U}, Z; h\right)$ if and only if $h\left(\tilde{Y}, Z, \tilde{U}\right) = 0$, equivalently $\tilde{U} \in \mathcal{U}\left(\tilde{Y}, Z; h\right)$. Condition (iii) is therefore equivalent to

$$\mathbb{P}\left[\tilde{U} \in \mathcal{U}\left(\tilde{Y}, Z; h\right)|Z = z\right] = 1. \tag{A.1}$$

Thus there exist random variables $\tilde{Y}$ and $\tilde{U}$ satisfying (i) and (ii) such that (A.1) holds, equivalently such that $G_{U|Z}(\cdot|z)$ is selectionable with respect to the conditional distribution of $\mathcal{U}(Y,Z;h)$ given $Z=z$. The choice of z was arbitrary, and the argument thus follows for all $z \in \mathcal{R}_Z$. ■

Proof of Theorem 2. *This follows directly from application of Theorem **1** to Definitions 2 and 3, respectively.* ■

Proof of Corollary 1. *From the selectionability characterization of $\mathcal{M}^*$ in U-space in Theorem 2, we have that*

$$\mathcal{M}^* = \left\{(h, G_U) \in \mathcal{M} : G_U(\cdot|z) \preccurlyeq \mathcal{U}(Y,z;h) \text{ when } Y \sim F_{Y|Z}(\cdot|z),\ a.e.\ z \in \mathcal{R}_Z\right\}.$$

Fix $z \in \mathcal{R}_Z$ and suppose $Y \sim F_{Y|Z}(\cdot|z)$. From Artstein's Inequality, see Artstein (1983), Norberg (1992), or Molchanov (2005, Section 1.4.8.), $G_U(\cdot|z) \preccurlyeq \mathcal{U}(Y,z;h)$ if and only if

$$\forall \mathcal{K} \in \mathsf{K}(\mathcal{R}_U),\ G_U(\mathcal{K}|z) \leq F_{Y|Z}[\mathcal{U}(Y,z;h) \cap \mathcal{K} \neq \emptyset | z],$$

where $\mathsf{K}(\mathcal{R}_Z)$ denotes the collection of all compact sets on $\mathcal{R}_U$. By Corollary 1.4.44 of Molchanov (2005) this is equivalent to

$$\forall \mathcal{S} \in \mathsf{G}(\mathcal{R}_U),\ G_U(\mathcal{S}|z) \leq F_{Y|Z}[\mathcal{U}(Y,z;h) \cap \mathcal{S} \neq \emptyset | z],$$

where $\mathsf{G}(\mathcal{R}_U)$ denotes the collection of all open subsets of $\mathcal{R}_U$. Because $G_U(\mathcal{S}|z) = 1 - G_U(\mathcal{S}^c|z)$ and

$$F_{Y|Z}[\mathcal{U}(Y,z;h) \subseteq \mathcal{S}^c | z] = 1 - F_{Y|Z}[\mathcal{U}(Y,z;h) \cap \mathcal{S} \neq \emptyset | z],$$

this is equivalent to

$$\forall \mathcal{S} \in \mathsf{G}(\mathcal{R}_U),\ F_{Y|Z}[\mathcal{U}(Y,z;h) \subseteq \mathcal{S}^c | z] \leq G_U(\mathcal{S}^c|z).$$

The collection of $\mathcal{S}^c$ such that $\mathcal{S} \in \mathsf{G}(\mathcal{R}_U)$ is precisely the collection of closed sets on $\mathcal{R}_U$, $\mathsf{F}(\mathcal{R}_U)$, completing the proof. ■

Proof of Lemma 1. *$\mathcal{U}_{\mathcal{S}}(h,z)$ is a union of sets contained in $\mathcal{S}$, so that $\mathcal{U}_{\mathcal{S}}(h,z) \subseteq \mathcal{S}$ and*

$$G_{U|Z}(\mathcal{U}_{\mathcal{S}}(h,z)|z) \leq G_{U|Z}(\mathcal{S}|z). \tag{A.2}$$

By supposition we have

$$C_h(\mathcal{U}_{\mathcal{S}}(h,z)|z) \leq G_{U|Z}(\mathcal{U}_{\mathcal{S}}(h,z)|z). \tag{A.3}$$

The result then holds because $C_h(\mathcal{S}|z) = C_h(\mathcal{U}_{\mathcal{S}}(h,z)|z)$, *since*

$$\begin{aligned}
C_h(\mathcal{U}_{\mathcal{S}}(h,z)|z) &\equiv \mathbb{P}[\mathcal{U}(Y,Z;h) \subseteq \mathcal{U}_{\mathcal{S}}(h,z)|Z=z] \\
&= \int_{y\in\mathcal{R}_Y} 1[\mathcal{U}(y,z;h) \subseteq \mathcal{U}_{\mathcal{S}}(h,z)]\, dF_{Y|Z}(y|z) \\
&= \int_{y\in\mathcal{R}_Y} 1[\mathcal{U}(y,z;h) \subseteq \mathcal{S}]\, dF_{Y|Z}(y|z) \\
&= C_h(\mathcal{S}|z),
\end{aligned}$$

where the second line follows by the law of total probability, and the third by the definition of $\mathcal{U}_{\mathcal{S}}(h,z)$ *in (3.3). Combining* $C_h(\mathcal{U}_{\mathcal{S}}(h,z)|z) = C_h(\mathcal{S}|z)$ *with (A.2) and (A.3) completes the proof.* ■

Proof of Theorem 3. *Fix* (h,z). *Suppose that*

$$\forall \mathcal{U} \in \mathsf{Q}(h,z),\ C_h(\mathcal{U}|z) \leq G_{U|Z}(\mathcal{U}|z). \tag{A.4}$$

Let $\mathcal{S} \in \mathsf{U}^*(h,z)$ *and* $\mathcal{S} \notin \mathsf{Q}(h,z)$. *Since* $\mathcal{S} \notin \mathsf{Q}(h,z)$ *there exist nonempty collections of sets* $\mathsf{S}_1, \mathsf{S}_2 \in \mathsf{U}^{\mathcal{S}}(h,z)$ *with* $\mathsf{S}_1 \cup \mathsf{S}_2 = \mathsf{U}^{\mathcal{S}}(h,z)$ *such that*

$$\mathcal{S}_1 \equiv \bigcup_{\mathcal{T}\in\mathsf{S}_1} \mathcal{T} \in \mathsf{Q}(h,z),\ \mathcal{S}_2 \equiv \bigcup_{\mathcal{T}\in\mathsf{S}_2} \mathcal{T} \in \mathsf{Q}(h,z),$$

and

$$G_{U|Z}(\mathcal{S}_1 \cap \mathcal{S}_2|z) = 0. \tag{A.5}$$

Since $\mathcal{S}_1, \mathcal{S}_2 \in \mathsf{Q}(h,z)$ *we also have that*

$$C_h(\mathcal{S}_1|z) \leq G_{U|Z}(\mathcal{S}_1|z) \text{ and } C_h(\mathcal{S}_2|z) \leq G_{U|Z}(\mathcal{S}_2|z). \tag{A.6}$$

Because $\mathsf{S}_1 \cup \mathsf{S}_2 = \mathsf{U}^{\mathcal{S}}(h,z)$,

$$\mathcal{U}(Y,z;h) \subseteq \mathcal{S} \Rightarrow \{\mathcal{U}(Y,z;h) \subseteq \mathcal{S}_1 \text{ or } \mathcal{U}(Y,z;h) \subseteq \mathcal{S}_2\}. \tag{A.7}$$

Using (A.7), (A.6), and (A.5) in sequence we then have

$$C_h(\mathcal{S}|z) \leq C_h(\mathcal{S}_1|z) + C_h(\mathcal{S}_2|z) \leq G_{U|Z}(\mathcal{S}_1|z) + G_{U|Z}(\mathcal{S}_2|z) = G_{U|Z}(\mathcal{S}|z).$$

Combined with (A.4) this implies $C_h(\mathcal{S}|z) \leq G_{U|Z}(\mathcal{S}|z)$ *for all* $\mathcal{S} \in \mathsf{U}^*(h,z)$ *and hence all closed* $\mathcal{S} \subseteq \mathcal{R}_U$ *by Lemma 1, completing the proof.* ■

Proof of Corollary 2. *Consider any* $\mathcal{S} \in \mathsf{Q}^E(h,z)$. *Then for all* $y \in \mathcal{Y}$, *either* $\mathcal{U}(y,z;h) \subseteq \mathcal{S}$ *or* $\mathcal{U}(y,z;h) \subseteq \overline{\mathcal{S}^c}$. *Thus*

$$C_h(\mathcal{S}|z) + C_h\left(\overline{\mathcal{S}^c}|z\right) = \mathbb{P}\left[\mathcal{U}(Y,Z;h) \subseteq \mathcal{S}|z\right] + \mathbb{P}\left[\mathcal{U}(Y,Z;h) \subseteq \overline{\mathcal{S}^c}|z\right] = 1. \quad \text{(A.8)}$$

The inequalities of Theorem ***3*** *imply that*

$$G_{U|Z}(\mathcal{S}|z) \geq C_h(\mathcal{S}|z) \text{ and } G_{U|Z}\left(\overline{\mathcal{S}^c}|z\right) \geq C_h\left(\overline{\mathcal{S}^c}|z\right).$$

Then absolute continuity of $G_{U|Z}(\cdot|z)$ *implies that* $G_{U|Z}(\mathcal{S}|z)+G_{U|Z}\left(\overline{\mathcal{S}^c}|z\right) = 1$, *which taken with* (A.8) *implies that both inequalities hold with equality.* ■

Proof of Theorem 4. *Under Restriction SI,* $G_{U|Z}(\cdot|z) = G_U(\cdot)$ *a.e.* $z \in \mathcal{R}_Z$. *(3.6) and (3.7) follow from (3.1) and Theorem 2, respectively, upon substituting* $G_U(\cdot)$ *for* $G_{U|Z}(\cdot|z)$. *(3.8) follows from Corollary 2, again by replacing* $G_{U|Z}(\cdot|z)$ *with* $G_U(\cdot)$. *The equivalence of (3.7) and (3.9) with* $G_{U|Z}(\cdot|z) = G_U(\cdot)$ *holds by Artstein's inequality, see e.g. Molchanov (2005, pp. 69-70, Corollary 4.44).* ■

Proof of Corollary 3. *It follows from the main text that for any structure* $(h, G_U) \in \mathcal{M}^*$, $h \in \mathcal{H}^*$. *Now consider an arbitrary structural function* $h \in \mathcal{H}^*$ *as defined in the statement of the Corollary. Then*

$$\forall \mathcal{S} \in \mathsf{Q}^*(h), \ \sup_{z \in \mathcal{R}_Z} C_h(\mathcal{S}|z) \leq \inf_{z \in \mathcal{R}_Z} (1 - C_h(\mathcal{S}^c|z)).$$

By monotonicity of the containment functional in the argument $\mathcal{S}$, *and the fact that* $C_h(\mathcal{S}|z) \in [0,1]$, *it follows that there exists some probability distribution function* $G_U(\cdot)$ *such that:*

$$\forall \mathcal{S} \in \mathsf{Q}^*(h), \ \sup_{z \in \mathcal{R}_Z} C_h(\mathcal{S}|z) \leq G_U(\mathcal{S}) \leq \inf_{z \in \mathcal{R}_Z} (1 - C_h(\mathcal{S}^c|z)). \quad \text{(A.9)}$$

Since $\mathsf{Q}(h,z) \subseteq \mathsf{Q}^*(h)$ *a.e.* $z \in \mathcal{R}_Z$, *it follows that for almost every* $z \in \mathcal{R}_Z$:

$$\forall \mathcal{S} \in \mathsf{Q}(h,z), \quad C_h(\mathcal{S}|z) \leq G_U(\mathcal{S}).$$

By Theorem 3 this implies that $C_h(\mathcal{S}|z) \leq G_U(\mathcal{S})$ *holds for almost every* $z \in \mathcal{R}_Z$ *and for all closed sets* $\mathcal{S}$ *in* $\mathcal{R}_U$, *so that by Corollary 1,* $(h, G_U) \in \mathcal{M}^*$, *completing the proof.* ■

Proof of Theorem 5. *Restrictions A3 and A4 guarantee that* $\mathcal{U}(Y,Z;h)$ *is integrable and closed. In particular integrability holds because by Restriction A3 first* $G_{U|Z}(\mathcal{S}|z) \equiv \mathbb{P}[U \in \mathcal{S}|z]$ *so that, for some finite* $c \in \mathcal{C}$, $E[U|z] = c$ *a.e.* $z \in \mathcal{R}_Z$, *and second* $\mathbb{P}[h(Y,Z,U) = 0] = 1$ *so that*

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

implying that $\mathcal{U}(Y,Z;h)$ *has an integrable selection, namely* U. *From Definition 6,* $c \in \mathbb{E}[\mathcal{U}(Y,Z;h)|z]$ *a.e.* $z \in \mathcal{R}_Z$ *therefore holds if and only if there exists a random variable* $\tilde{U} \in \mathrm{Sel}(\mathcal{U}(Y,Z;h))$ *such that* $E\left[\tilde{U}|z\right] = c$ *a.e.* $z \in \mathcal{R}_Z$, *and hence* $\mathcal{H}^*$ *is the identified set for* h. *The representation of the identified set of structures* $\mathcal{M}^*$ *then follows directly from Theorem 2.* ■

Proof of Corollary 4. *Fix* $z \in \mathcal{R}_Z$. *The conditional Aumann expectation* $\mathbb{E}[\mathcal{U}(Y,Z;h)|z]$ *is the set of values for*

$$\int_{\mathcal{R}_{Y|z}} \int_{\mathcal{U}(y,z;h)} u dF_{U|YZ}(u|y,z)\, dF_{Y|Z}(y|z),$$

such that there exists for each $y \in \mathcal{R}_{Y|z}$ *a conditional distribution* $F_{U|YZ}(u|y,z)$ *with support on* $\mathcal{U}(y,z;h)$. *Since each* $\mathcal{U}(y,z;h)$ *is convex, the inner integral*

$$\int_{\mathcal{U}(y,z;h)} u dF_{U|YZ}(u|y,z)$$

can take any value in $\mathcal{U}(y,z;h)$, *and hence* $\mathbb{E}[\mathcal{U}(Y,Z;h)|z]$ *is the set of values of the form*

$$\int_{\mathcal{R}_{Y|z}} u(y,z)\, dF_{Y|Z}(y|z)$$

for some $u(y,z) \in \mathcal{U}(y,z;h)$, *each* $y \in \mathcal{R}_{Y|z}$. *Since the choice of* z *was arbitrary, this completes the proof.* ■

Proof of Corollary 5. *Restrictions A3 and A4 and the continuity requirement of Restriction MI* guarantee that* $\mathcal{D}(Y,Z;h)$ *is integrable and closed. From Definition 6, for any* $c \in \mathcal{C}$, $c \in \mathbb{E}[\mathcal{D}(Y,Z;h)|z]$ *a.e.* $z \in \mathcal{R}_Z$ *therefore holds if and only if there exists a random variable* $D \lesssim \mathcal{D}(Y,Z;h)$ *such that* $E[D|z] = c$ *a.e.* $z \in \mathcal{R}_Z$. $D \lesssim \mathcal{D}(Y,Z;h)$ *ensures that*

$$\mathbb{P}[D \in \mathcal{D}(Y,Z;h)|z] = 1, \text{ a.e. } z \in \mathcal{R}_Z.$$

Define

$$\widetilde{\mathcal{U}}(D,Y,Z;h) \equiv \{u \in \mathcal{U}(Y,Z;h) : D = d(u,Z)\}.$$

By the definition of $\mathcal{D}(Y,Z;h)$, $D \in \mathcal{D}(Y,Z;h)$ *implies that* $\widetilde{\mathcal{U}}(D,Y,Z;h)$ *is nonempty. Hence there exists a random variable* $\tilde{U}$ *such that with probability one* $\tilde{U} \in \widetilde{\mathcal{U}}(D,Y,Z;h) \subseteq \mathcal{U}(Y,Z;h)$ *where* $D = d\left(\tilde{U},Z\right)$. *Thus* $\tilde{U}$ *is a selection of* $\mathcal{U}(Y,Z;h)$ *and* $E\left[d\left(\tilde{U},Z\right)|z\right] = d$ *a.e.* $z \in \mathcal{R}_Z$, *and therefore* $\mathcal{H}^*$ *is the identified set for* h, *and the given characterization of* $\mathcal{M}^*$ *follows.* ■

Proof of Theorem 6. *Using Corollary 1 and Definition 4 with* $\psi\left(h, \mathcal{G}_{U|Z}\right) = h$*, the identified set of structural functions* h *is*

$$\mathcal{H}^{**} = \left\{h \in \mathcal{H} : \exists \mathcal{G}_{U|Z} \in \mathsf{G}_{U|Z} \text{ s.t. } \forall \mathcal{S} \in \mathsf{F}\left(\mathcal{R}_U\right),\ C_h\left(\mathcal{S}|z\right) \leq G_{U|Z}\left(\mathcal{S}|z\right) \text{ a.e. } z \in \mathcal{R}_Z\right\}. \quad \text{(A.10)}$$

Consider any $h \in \mathcal{H}^{**}$*. We wish to show first that* h *belongs to the set* $\mathcal{H}^*$ *given in (3.16). Fix* $z \in \mathcal{R}_Z$ *and choose* c *such that* $G_{U|Z}\left(\left(-\infty, c\right]|z\right) = \tau$*, which can be done by virtue of the continuity condition of Restriction QI. Then*

$$C_h\left(\left(-\infty, c\right]|z\right) \leq G_{U|Z}\left(\left(-\infty, c\right]|z\right) = \tau, \quad \text{(A.11)}$$

and because of Restriction IS, $\mathcal{U}\left(Y, Z; h\right) = \left[\underline{u}\left(Y, Z; h\right), \overline{u}\left(Y, Z; h\right)\right]$*,*

$$C_h\left(\left(-\infty, c\right]|z\right) = F_{Y|Z}\left[\overline{u}\left(Y, Z; h\right) \leq c|z\right]. \quad \text{(A.12)}$$

Now consider $\mathcal{S} = [c, \infty)$*. We have by monotonicity of the containment functional* $C_h\left(\cdot|z\right)$ *and from* $C_h\left(\mathcal{S}|z\right) \leq G_{U|Z}\left(\mathcal{S}|z\right)$ *in (A.10) that*

$$C_h\left(\left(c, \infty\right)|z\right) \leq C_h\left(\left[c, \infty\right)|z\right) \leq G_{U|Z}\left(\left[c, \infty\right)|z\right) = 1 - \tau, \quad \text{(A.13)}$$

where the equality holds by continuity of the distribution of $U|Z = z$ *in a neighborhood of its* τ *quantile. Again using Restriction IS,*

$$C_h\left(\left(c, \infty\right)|z\right) = 1 - F_{Y|Z}\left[\underline{u}\left(Y, Z; h\right) \leq c|z\right]. \quad \text{(A.14)}$$

Combining this with (A.13) and also using (A.11) and (A.12) above gives

$$F_{Y|Z}\left[\overline{u}\left(Y, Z; h\right) \leq c|z\right] \leq \tau \leq F_{Y|Z}\left[\underline{u}\left(Y, Z; h\right) \leq c|z\right]. \quad \text{(A.15)}$$

The choice of z *was arbitrary and so we have that the above holds a.e.* $z \in \mathcal{R}_Z$*, implying that* $h \in \mathcal{H}^*$*.*

Now consider any $h \in \mathcal{H}^*$*. We wish to show that* $h \in \mathcal{H}^{**}$*. It suffices to show that for any such* h *under consideration there exists a collection of conditional distributions* $\mathcal{G}_{U|Z}$ *such that for almost every* $z \in \mathcal{R}_Z$ *(1)* $G_{U|Z}\left(\cdot|z\right)$ *has* τ*-quantile equal to* c*, and (2)* $\forall \mathcal{S} \in \mathsf{F}\left(\mathcal{R}_U\right)$*,* $C_h\left(\mathcal{S}|z\right) \leq G_{U|Z}\left(\mathcal{S}|z\right)$*.*

To do so we fix an arbitrary $z \in \mathcal{R}_Z$ *and construct* $G_{U|Z}\left(\cdot|z\right)$ *such that (1) and (2) hold. Namely let* $G_{U|Z}\left(\cdot|z\right)$ *be such that for each* $\mathcal{S} \in \mathsf{F}\left(\mathcal{R}_U\right)$*,*

$$G_{U|Z}\left(\mathcal{S}|z\right) = \lambda\left(z\right) C_h\left(\mathcal{S}|z\right) + \left(1 - \lambda\left(z\right)\right)\left(1 - C_h\left(\mathcal{S}^c|z\right)\right), \quad \text{(A.16)}$$

where $\lambda(z)$ *is chosen to satisfy*

$$\lambda(z) F_{Y|Z}[\overline{u}(Y,Z;h) \leq c|z] + (1-\lambda(z)) F_{Y|Z}[\underline{u}(Y,Z;h) \leq c|z] = \tau. \tag{A.17}$$

The left hand side of equation (A.17) is precisely (A.16) with $\mathcal{S} = (-\infty, c]$*. Because* $h \in \mathcal{H}^*$*, (A.15) holds, which guarantees that* $\lambda(z) \in [0,1]$*. (A.17) and (A.16) deliver*

$$G_{U|Z}((-\infty, c]|z) = \tau,$$

so that (1) holds. Moreover, it easy to verify that for any $\mathcal{S}$*,*

$$C_h(\mathcal{S}|z) \leq 1 - C_h(\mathcal{S}^c|z),$$

since $C_h(\cdot|z)$ *is the conditional containment functional of* $\mathcal{U}(Y,Z;h)$ *and* $1 - C_h(\mathcal{S}^c|z)$ *is the conditional capacity functional of* $\mathcal{U}(Y,Z;h)$*. Hence* $C_h(\mathcal{S}|z) \leq G_{U|Z}(\mathcal{S}|z)$*. Thus (2) holds, and since the choice* z *was arbitrary,* $h \in \mathcal{H}^{**}$ *as desired. This verifies claim (i) of the Theorem.*

Claim (ii) of the Theorem holds because with $\overline{u}(Y,Z;h)$ *and* $\underline{u}(Y,Z;h)$ *continuously distributed given* $Z = z$*, a.e.* $z \in \mathcal{R}_Z$*, their conditional quantile functions are invertible at* τ*. Thus for any* $z \in \mathcal{R}_Z$*,*

$$\underline{q}(\tau, z; h) \leq c \leq \overline{q}(\tau, z; h) \Leftrightarrow F_{Y|Z}[\overline{u}(Y,Z;h) \leq c|z] \leq \tau \leq F_{Y|Z}[\underline{u}(Y,Z;h) \leq c|z].$$

Claim (iii) of the Theorem follows directly from Theorem 2. ■

B Equivalence of Selectionability of Conditional and Joint Distributions

In this section we prove that selectionability statements in the main text required for observational equivalence and characterization of identified sets conditional on $Z = z$ for almost every $z \in \mathcal{R}_Z$ are in fact equivalent to unconditional selectionability statements inclusive of Z. Intuitively this holds because knowledge of a conditional distribution of a random set or random vector given $Z = z$, a.e. $z \in \mathcal{R}_Z$, is logically equivalent to knowledge of the joint distribution of that given random vector or random set and Z.

Proposition 1 *(i)* $F_{Y|Z}(\cdot|z) \preccurlyeq \mathcal{Y}(U,Z;h)|Z = z$ *a.e.* $z \in \mathcal{R}_Z$ *if and only if* $F_{YZ}(\cdot) \preccurlyeq \mathcal{Y}(U,Z;h) \times Z$*. (ii)* $G_{U|Z}(\cdot|z) \preccurlyeq \mathcal{U}(Y,Z;h)|Z = z$ *a.e.* $z \in \mathcal{R}_Z$ *if and only if* $G_{UZ}(\cdot) \preccurlyeq \mathcal{U}(Y,Z;h) \times Z$*.*

Proof of Proposition 1. *Note that since the choice of z in the above Theorem is arbitrary the statement holds because*

$$\begin{aligned}\mathbb{P}\left[\left(\tilde{Y},Z\right)\in\mathcal{Y}\left(U,Z;h\right)\times Z\right] &= \int\limits_{z\in\mathcal{R}_Z}\mathbb{P}\left[\left(\tilde{Y},Z\right)\in\mathcal{Y}\left(\tilde{U},Z;h\right)\times Z|Z=z\right]dF_Z\left(z\right)\\ &= \int\limits_{z\in\mathcal{R}_Z}\mathbb{P}\left[\tilde{Y}\in\mathcal{Y}\left(\tilde{U},Z;h\right)|Z=z\right]dF_Z\left(z\right),\end{aligned}$$

which is equal to one if and only if $\mathbb{P}\left[\tilde{Y}\in\mathcal{Y}\left(\tilde{U},Z;h\right)|Z=z\right]=1$ for almost every $z\in\mathcal{R}_Z$. By identical reasoning, $G_{U|Z}\left(\cdot|z\right)$ is selectionable with respect to the conditional distribution of $\mathcal{U}\left(Y,Z;h\right)$ given $Z=z$ for almost every $z\in\mathcal{R}_Z$ if and only if $G_{UZ}\left(\cdot\right)$ is selectionable with respect to the distribution of $\mathcal{U}\left(Y,Z;h\right)\times Z$. ■

C Computational Details for Numerical Illustrations of Section 4.2

In this Section we describe computation of the conditional containment functional $C_\theta\left(\left[u_*,u^*\right]|z\right)$ in (4.16). Computations were carried out in `Mathematica 9`.

Given the structure specified for DGP1 and DGP2 in Section 4.2, the conditional distribution of Y_2^* given $Y_1=y_1$ and $Z=z$ is for any (y_1,z)

$$\mathcal{N}\left(a\left(z\right)+\frac{\sigma_{1v}+\gamma_1\sigma_{vv}}{\sigma_{11}+2\gamma_1\sigma_{1v}+\gamma_1^2\sigma_{vv}}\left(y_1-\left(\gamma_0+\gamma_1 a\left(z\right)\right)\right),\sigma_{vv}-\frac{\left(\sigma_{1v}+\gamma_1\sigma_{vv}\right)^2}{\sigma_{11}+2\gamma_1\sigma_{1v}+\gamma_1^2\sigma_{vv}}\right),$$

where $a\left(z\right)\equiv\delta_0+\delta_1 z$. From this it follows that the conditional (discrete) distribution of (Y_{2l},Y_{2u}) given Y_1 and Z is:

$$\begin{aligned}\mathbb{P}\left[\left[Y_{2l},Y_{2u}\right)=I_j|y_1,z\right]=\Phi&\left(\frac{c_{j+1}-\left(a(z)+\frac{\sigma_{1v}+\gamma_1\sigma_{vv}}{\sigma_{11}+2\gamma_1\sigma_{1v}+\gamma_1^2\sigma_{vv}}\left(y_1-\left(\gamma_0+\gamma_1 a(z)\right)\right)\right)}{\sqrt{\sigma_{vv}-\frac{\left(\sigma_{1v}+\gamma_1\sigma_{vv}\right)^2}{\sigma_{11}+2\gamma_1\sigma_{1v}+\gamma_1^2\sigma_{vv}}}}\right)\\ &-\Phi\left(\frac{c_{j}-\left(a(z)+\frac{\sigma_{1v}+\gamma_1\sigma_{vv}}{\sigma_{11}+2\gamma_1\sigma_{1v}+\gamma_1^2\sigma_{vv}}\left(y_1-\left(\gamma_0+\gamma_1 a(z)\right)\right)\right)}{\sqrt{\sigma_{vv}-\frac{\left(\sigma_{1v}+\gamma_1\sigma_{vv}\right)^2}{\sigma_{11}+2\gamma_1\sigma_{1v}+\gamma_1^2\sigma_{vv}}}}\right).\end{aligned}$$

The distribution of Y_1 given $Z=z$ is

$$Y_1|Z=z\sim\mathcal{N}\left(\gamma_0+\gamma_1 a(z),\sigma_{11}+2\gamma_1\sigma_{1v}+\gamma_1^2\sigma_{vv}\right). \tag{C.1}$$

The conditional containment functional can thus be written

$$\begin{aligned} C_\theta\left([u_*, u^*]\,|z\right) &= \sum_j \mathbb{P}[(g_0 + g_1 c_{j+1} + u_* \leq Y_1 \leq g_0 + g_1 c_j + u^*) \wedge (Y_2, Y_3) = I_j|z] \\ &= \sum_j \max\left(0, \int_{\gamma_0+\gamma_1 c_{j+1}+u_*}^{\gamma_0+\gamma_1 c_j+u^*} f_{Y_1|Z}\left(y_1|z\right) \times \mathbb{P}\left[[Y_{2l}, Y_{2u}) = I_j|y_1, z\right] dy_1\right). \end{aligned}$$

where $f_{Y_1|Z}\left(\cdot|z\right)$ is the normal pdf with mean and variance given in (C.1).

In the calculations performed in `Mathematica` we used the following equivalent formulation employing a *single* numerical integration for computation of $C_\theta\left([u_*, u^*]\,|z\right)$.

$$C_\theta\left([u_*, u^*]\,|z\right) \equiv$$
$$\int_{-\infty}^{\infty}\left(\sum_j 1[g_0 + g_1 c_{j+1} + u_* < y_1 < g_0 + g_1 c_j + u^*] \times f_{Y_1|Z}\left(y_1|z\right) \times \mathbb{P}\left[[Y_{2l}, Y_{2u}) = I_j|y_1, z\right]\right) dy_1.$$