Counterfactual worlds: Characterizing the identifying power of incomplete models with conditional and marginal independence restrictions.

Andrew Chesher and Adam M. Rosen
CeMMAP and UCL

December 31, 2013

Abstract

We study a generalization of the treatment effect model in which an observed discrete classifier indicates in which one of a set of counterfactual processes a decision maker is observed. The other observed outcomes are delivered by the particular counterfactual process in which the decision maker is found. Models of the counterfactual processes can be incomplete in the sense that even with knowledge of the values of observed exogenous and unobserved variables they may not deliver a unique value of the endogenous outcomes. We study the identifying power of models of this sort that incorporate (i) conditional independence restrictions under which unobserved variables and the classifier variable are stochastically independent conditional on some of the observed exogenous variables and (ii) marginal independence restrictions under which unobservable variables and a subset of the exogenous variables are independently distributed. We use random set theory methods to characterize the identifying power of these models for fundamental structural relationships and probability distributions and for interesting functionals of these objects, some of which may be point identified. In one example of an application, we observe the entry decisions of firms that can choose which of a number of markets to enter and we observe various endogenous outcomes delivered in the markets they choose to enter.

1 Introduction

In the classical treatment effect model, pioneered in Rubin (1974) and Rosenbaum and Rubin (1983), a discrete classifier indicates which one of a list of counterfactual outcomes is observed. The counterfactual outcomes and the discrete classifier may not be independently distributed because decision makers with beliefs about the counterfactual outcomes strive to end up in desirable situations. The classical model imposes a conditional independence restriction, namely that counterfactual outcomes and the classifier are independently distributed conditional on some known list of observed variables. Under some additional restrictions the resulting model point identifies the marginal distributions of the counterfactual outcomes and thus Average Treatment Effects and Quantile Treatment Effects, as in for instance Imbens and Newey (2009).

In this paper we extend the scope of the treatment effect model. The counterfactual outcomes of the classical model are replaced by counterfactual unobservable variables. These unobservables produce stochastic variation in counterfactual processes which deliver the values of outcomes that the econometrician observes.

The econometrician observes each decision maker engaging in one and only one of the counterfactual processes and observes only the realizations of the endogenous outcomes delivered by that process. Some exogenous variables are also observed. Wary of basing inference on highly restrictive models, the econometrician may come to data with incomplete models of the counterfactual processes. It is this case that is center stage in this paper.

We consider the following types of covariation restriction placed on unobservable variables.

1. Conditional independence restrictions. The unobservable variables appearing in the counterfactual processes and the classifier are independently distributed conditional on the observed exogenous variables. This is the sort of condition that appears in the classical treatment effect model.

2. Marginal independence restrictions. The unobservable variables appearing in the counterfactual processes and a possibly vector-valued function of the exogenous variables are stochastically independent. In the absence of selection this would be a common restriction in nonlinear incomplete models.

The models we study contain a blend of conditional and marginal independence restrictions. Our analysis brings together strands from structural econometrics and analysis of causal inference. A contribution of the paper is to provide a characterization of the (sharp) identified sets delivered by models which may be incomplete and embody conditional and marginal independence restrictions.

Here are examples of cases in which the results of this paper can be applied

1. Some unemployed workers participate in a training programme, others do not. Subsequently the workers engage in one of two counterfactual labor market processes, corresponding to whether or not training was received, and endogenous outcomes such as unemployment duration and wage on re-employment, job tenure and so forth are observed.

2. In a generalization of the Roy model, individuals decide in which of a number of occupations to work whereupon we observe multiple endogenous outcomes that arise in the chosen occupation.[1]

[1] The Roy Model presumes that each individual chooses the alternative (here the occupation) that delivers the highest value of one of the observed outcomes variables. Our model allows alternative criteria for selection among the alternatives.

3. Firms decide whether or not to operate in markets distinguished by regulatory regimes and various endogenous outcomes that ensue are observed.

The research reported here is a first step on the way to the study of a broad class of incomplete models that involve a blend of conditional and marginal independence restrictions. The models studied in this paper impose few restrictions on the determination of the state in which individuals are found. There is just a conditional independence restriction requiring unobservable variables and the classifier variable to be independently distributed conditional on some observed exogenous variables. The way in which the classifier variable is determined is not specified in the models studied in this paper.

In work in progress we extend our analysis to cover more widely applicable models with some of the following features.

1. Economic restrictions on the determination of the process in which an individual is engaged, for example a model of choice.

2. A continuum of processes rather than the discrete classification considered here.

3. Conditional independence restrictions involving multiple endogenous variables as in control function models.

2 Structures, Models and Data

This Section introduces notation and constructs employed in the rest of the paper.

Throughout Y denotes a list of endogenous variables, Z denotes a list of observed exogenous variables and U denotes a list of unobserved exogenous variables. Each of these variables may be vector-valued and the observable variables may be discrete or continuous. The variables have support $\mathcal{R}_{YZU}$ on a subset of Euclidean space. Lower case y, z and u denote values of these variables. For any random vectors A, B, $\mathcal{R}_{A|b}$ denotes the support of A conditional on $B = b$. For random variables A and B, $A \perp\!\!\!\perp B$ indicates that A and B are independently distributed.

With M counterfactual processes there are M components in U, thus: $U = (U_1, \ldots, U_M)$ with only U_m delivering stochastic variation in the m^{th} counterfactual process.

Some econometric selection models impose the restriction $U_1 = \cdots = U_M$. Examples are given in Heckman and Robb (1985). A number of papers study econometric selection models without this restriction. Such models are described in Heckman, Urzua, and Vytlacil (2008) as models with "essential heterogeneity". Examples can be found in Heckman and Vytlacil (2007) and the references therein. In these econometric selection models it is common to find a discrete choice specification of the determination of the classifier variable and instrumental variable restrictions, see for example Heckman and Vytlacil (2005).

In this paper we study models which have no detailed specification of the determination of the classifier variable. In this respect, like classical treatment effect models, they are incomplete, and as in those models there is a conditional independence condition. Our models also allow incompleteness in the specification of the processes that deliver counterfactual outcomes, and this specification may include instrumental variable restrictions.

2.1 Structural functions

A model specifies M structural functions, $h_m(y, z, u) : \mathcal{R}_{YZU} \to \mathbb{R}$ where $\mathbb{R}$ denotes the real line.[2] The variable $u = (u_1, \ldots, u_M)$ and each function h_m is invariant with respect to changes in u_{-m} where u_{-m} denotes u with the element u_m omitted. This representation of structural functions, used in Chesher and Rosen (2013a), is convenient when models of counterfactual processes are incomplete.

One element of Y, denoted Y_*, is discrete taking values in $\{1, \ldots, M\}$. This classifier variable is the "treatment" or "selection" indicator. A realization of (Y, Z) delivered by the m^{th} counterfactual process is observed if and only if Y_* has the realized value m.

Associated with each of the M structural functions are level sets as follows.

$$\left.\begin{array}{l} \mathcal{Y}(u, z; h_m) \equiv \{y : h_m(y, z, u) = 0\} \\ \\ \mathcal{U}(y, z; h_m) \equiv \{u : h_m(y, z, u) = 0\} \end{array}\right\}, \quad m \in \{1, \ldots, M\}$$

The level set $\mathcal{Y}(u, z; h_m)$ contains the values of Y that arise in the m^{th} counterfactual process when $Z = z$ and $U = u$. Every element $y \in \mathcal{Y}(u, z; h_m)$ has $y_* = m$ and the set $\mathcal{Y}(u, z; h_m)$ is invariant with respect to changes in u_{-m}.

The level set $\mathcal{U}(y, z; h_m)$ gives the values of u that can give rise to the value y of Y when $Z = z$ in the m_{th} counterfactual process. This set comprises all vectors $u \in \mathcal{R}_U$ with m^{th} component u_m such that $h_m(y, z, u) = 0$, each such value coupled with every possible value of u_{-m}.

Without any restriction placed on the selection of the M counterfactual processes, the structural function for the composite process is.

$$h(y, z, u) = \sum_{m=1}^{M} 1[y_* = m] \times h_m(y, z, u).$$

Models that place restrictions on selection among the counterfactual M processes incorporate further information from the particular value of y_* observed. For example, in the Roy Model, the observed value of y_* corresponds to that value of m that achieves the maximum payoff or utility

[2] Note that for any m, $h_m(\cdot, \cdot, \cdot)$ will be in general invariant with respect to various components of its arguments. Thus $h_m(\cdot, \cdot, \cdot)$ defined as a mapping on $\mathcal{R}_{YZU}$ for each m is unrestrictive.

among the M available alternatives. For now we do not employ further conditions on the selection process, but such additional restrictions may be added.

There are associated zero-level sets of the composite structural function, h. Given a value (u,z) any one of the sets $\mathcal{Y}(u,z;h_m)$, $m \in \{1,\ldots,M\}$ may be observed so the y-level set of the composite structural function is the union of the y-level sets of the structural functions of the counterfactual processes:

$$\mathcal{Y}(u,z;h) \equiv \{y : h(y,z,u) = 0\} = \bigcup_{m=1}^{M} \mathcal{Y}(u,z;h_m).$$

Given a value (y,z) one and only one of the sets $\mathcal{U}(y,z;h_m)$ is observed, which one being determined by the value y_* of the treatment indicator variable, so

$$\mathcal{U}(y,z;h) \equiv \{u : h(y,z,u) = 0\} = \mathcal{U}(y,z;h_{y^*}),$$

where y_* is the value of the element of y that is the selection or classifier variable.

Example 1. **Treatment effects**. The binary treatment effect model of Rosenbaum and Rubin (1983) has counterfactual outcomes U_1 and U_2 and a binary indicator Y_2 equal to 1 if U_1 is observed and equal to 2 if U_2 is observed so that

$$Y_1 = 1[Y_2 = 1] \times U_1 + 1[Y_2 = 2] \times U_2$$

is the observed outcome. This simple (binary) treatment effect model has classifier variable $Y_* = Y_2$ and

$$h_m(y,z,u) = y_1 - u_m, \quad m \in \{1,2\}$$

with singleton y-level sets:

$$\begin{aligned}
\mathcal{Y}(u,z;h_1) &= \{(u_1,1)\}, \\
\mathcal{Y}(u,z;h_2) &= \{(u_2,2)\},
\end{aligned}$$

and non-singleton u-level sets:

$$\begin{aligned}
\mathcal{U}(y,z;h_1) &= \{(y_1,u_2) : u_2 \in \mathcal{R}_{U_2}\}, \\
\mathcal{U}(y,z;h_2) &= \{(u_1,y_1) : u_1 \in \mathcal{R}_{U_1}\}.
\end{aligned}$$

Exogenous variables are excluded from the counterfactual structural functions which involve neither unknown parameters nor unknown functions. There is the following composite structural function:

$$h(y,z,u) = 1[y_2 = 1] \times (y_1 - u_1) + 1[y_2 = 2] \times (y_1 - u_2).$$

□

Example 2. Supermarket choice and demand. A household is observed to shop in one of M supermarkets. In a household's supermarket of choice the endogenous variables: share of total expenditure on food, Y_1, and log total expenditure, Y_2, are observed. For each supermarket, indexed by $Y_3 \in \{1, \ldots, M\}$, there is an incomplete linear model with structural functions as follows.

$$h_m(y, z, u) = y_1 - \alpha_m - \beta_m y_2 - \gamma_m z_1 - u_m, \quad m \in \{1, \ldots, M\}$$

Define $U = (U_1, \ldots, U_M)$. There may be exogenous variables Z_2 and a restriction $U \perp\!\!\!\perp (Z_1, Z_2)$ and a conditional independence restriction $U \perp\!\!\!\perp Y_3|Z$ where $Z \equiv (Z_1, Z_2, Z_3)$. There are level sets as follows for each $m \in \{1, \ldots, M\}$:

$$\mathcal{Y}(u, z; h_m) = \{(\alpha_m + \beta_m y_2 + \gamma_m z_1 + u_m, y_2, m) : y_2 \in \mathcal{R}_{Y_2}\},$$

$$\mathcal{U}(y, z; h_m) = \{u \in \mathcal{R}_U : u_m = y_1 - \alpha_m - \beta_m y_2 - \gamma_m z_1\}.$$

The classifier variable $Y_* = Y_3$ and there is the following composite structural function:

$$h(y, z, u) = \sum_{m \in \{1, \ldots, M\}} 1[y_3 = m] \times (y_1 - \alpha_m - \beta_m y_2 - \gamma_m z_1 - u_m).$$

□

Example 3. **Training and labor market processes**. An unemployed worker either does ($Y_3 = 1$), or does not ($Y_3 = 2$), take part in a training program. A binary outcome Y_1 is observed, equal to one if employment is found within one year and zero otherwise. For each state there are incomplete threshold crossing-type models for this binary outcome with structural functions:

$$h_m(y, z, u) = y_1 \, |u_m - g_m(y_2, z_1)|_- + (1 - y_1) \, |u_m - g_m(y_2, z_1)|_+, \quad m \in \{1, 2\}$$

where $|c|_-$ and $|c|_+$ are respectively the negative and positive part of c.[3] Here y_2 is an possibly endogenous binary variable, for example an indicator of receipt of unemployment benefit, and z_1 is a component of a vector z whose elements are values of observed exogenous variables.[4] There are y-level sets:

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

[4] State-specific threshold crossing models such as this can arise using mixed proportionate hazard models of unemployment duration (see Example 1 in Chesher (2009)) with state-specific heterogeneity and baseline hazards.

$$\mathcal{Y}(u,z;h_m) = \{(y_1,y_2) \in \mathcal{R}_{Y_1Y_2} : (y_1 = 1 \wedge u_m \geq g_m(y_2,z_1)) \\ \vee (y_1 = 0 \wedge u_m \leq g_m(y_2,z_1))\}\,, \quad m \in \{1,2\}$$

where $\mathcal{R}_{Y_1Y_2}$ denotes the support of (Y_1, Y_2). There are u-level sets:

$$\mathcal{U}(y,z;h_m) = \left\{ \begin{array}{lcl} \{(u \in \mathbb{R}^2 : u_m \in (-\infty, g_m(y_2,z_1)]\} & , & y_1 = 0 \\ \{(u \in \mathbb{R}^2 : u_m \in [g_m(y_2,z_1), \infty)\} & , & y_1 = 1 \end{array} \right\}, \quad m \in \{1,2\}.$$

The classifier variable is $Y_* = Y_3$ and the structural function for the composite process is

$$h(y,z,u) = 1[y_3 = 1] \times \left(y_1 \left|u_1 - g_1(y_2,z_1)\right|_- + (1-y_1)\left|u_1 - g_1(y_2,z_1)\right|_+\right) \\ + 1[y_3 = 2] \times \left(y_1 \left|u_2 - g_2(y_2,z_1)\right|_- + (1-y_1)\left|u_2 - g_2(y_2,z_1)\right|_+\right).$$

□

2.2 Distributions of unobservables

Conditional on $Z = z$ the unobserved random variables $U \equiv (U_1, \ldots, U_M)$ have joint probability distribution $G_{U|Z}(\cdot|z)$ and marginal distributions $G_{U_m|z}(\cdot|z)$, $m \in \{1, \ldots, M\}$. There are collections of conditional probability distributions as follows:

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

and

$$\mathcal{G}_{U_m|Z} \equiv \{G_{U_m|Z}(\cdot|z) : z \in \mathcal{R}_Z\}, \quad m \in \{1, \ldots, M\}.$$

Here $\mathcal{R}_Z$ denotes the support of the observed exogenous variables and for any set $\mathcal{S} \subset \mathcal{R}_{U|z}$, $G_{U|Z}(\mathcal{S}|z)$ denotes the probability mass placed on the set $\mathcal{S}$ by the conditional probability distribution $G_{U|Z}(\cdot|z)$.

Each counterfactual process is characterized by a counterfactual structure $\left(h_m, \mathcal{G}_{U_m|Z}\right)$ and the complete process is characterized by a composite structure $(h, \mathcal{G}_{U|Z})$.

Models comprise restrictions which limit the set of admissible structures. In the models of counterfactual processes studied here there are restrictions on structural functions and two types of restrictions on the probability distribution of unobservable variables. Recall Y_* is the element of Y which has the role of selection or classifier variable. This is Y_2 in Example 1 and Y_3 in Examples 2 and 3.

1. **Conditional independence restrictions**. $U \perp\!\!\!\perp Y_*|Z$.

2. **Marginal independence restrictions**. There is a function $e(\cdot)$ such that $U \perp\!\!\!\perp e(Z)$.

The function $e(Z)$ is brought into play because one will typically want conditional independence to hold conditional on one set of exogenous variables and marginal independence to involve a different set of exogenous variables. One reason why this is likely to be desirable is that restricting $U \perp\!\!\!\perp Y_*|Z$ *and* $U \perp\!\!\!\perp Z$ (that is $e(Z) = Z$) implies $Y_* \perp\!\!\!\perp U$ which, in many cases, will not capture essential features of a problem. Specifying $e(Z) = Z_1$, a selection of the elements of Z, may be a common choice.[5]

In Example 1 it is common to impose $U \perp\!\!\!\perp Y_2|Z$. In Example 2 one might have reason to impose the conditional independence restriction $U \perp\!\!\!\perp Y_3|Z$ and the marginal independence restriction $U \perp\!\!\!\perp (Z_1, Z_2)$ where $Z = (Z_1, Z_2, Z_3)$.

2.3 Data

We consider cases in which realizations of (Y, Z) are obtained *via* an observation process such that the joint distribution of these variables, F_{YZ}, is identified. Of particular importance will be the conditional distributions of Y given Z and Y given (Y_*, Z). For any set $\mathcal{T} \subset \mathcal{R}_{Y|z}$, $F_{Y|Z}(\mathcal{T}|z)$ denotes the probability mass placed on the set $\mathcal{T}$ by the conditional probability distribution $F_{Y|Z}(\cdot|z)$ and $F_{Y|Y_*Z}(\mathcal{T}|y_*, z)$ denotes the probability mass placed on the set $\mathcal{T}$ by the conditional probability distribution $F_{Y|Y_*Z}(\cdot|y_*, z)$. The cumulative distribution function of Y given $Z = z$ evaluated at a point t is

$$\mathbb{P}[Y \leq t|Z = z] = F_{Y|Z}(\{y : y \leq t\}|z).$$

Likewise

$$\mathbb{P}[Y \leq t|Y_* = y_* \wedge Z = z] = F_{Y|Y_*Z}(\{y : y \leq t\}|y_*, z).$$

3 Identification

We ask: what characterizes the set of structures $(h, \mathcal{G}_{U|Z})$ admitted by a model, $\mathcal{M}$, that can deliver the joint distribution of F_{YZ}? This set, denoted $\mathcal{M}^*(F_{YZ})$, is the *identified set* delivered by the model when presented with F_{YZ}. We obtain characterizations of identified sets under conditional and marginal independence restrictions building on the results in Chesher and Rosen (2013a), henceforth CR2013.[6] Our analysis employs random set theory, also used for partial identification analysis in Beresteanu, Molchanov, and Molinari (2011, 2012), Chesher, Rosen, and Smolinski

[5] There is the possibility that conditional independence could be conditional on some function of Z, $d(Z)$, but that is not considered here.

[6] We use the term *identified set* to refer to the collection of **all** structures $\left(h, \mathcal{G}_{U|Z}\right) \in \mathcal{M}$ that can generate the joint distribution F_{YZ}. This set is sharp in that there is no structure $\left(h, \mathcal{G}_{U|Z}\right)$ belonging to the identified set that can be distinguished from one generating F_{YZ} on the basis of modeling restrictions and observed data.

(2013), and Chesher and Rosen (2012a, 2012b, 2013b). This is the first paper explicitly applying these tools in models with conditional independence restrictions. Moreover, we are unaware of previous papers featuring the combination of conditional and marginal independence restrictions with regard to the joint distribution of unobserved heterogeneity with observed variables in the class of models considered.

3.1 Restrictions

We impose Restrictions A1 - A3 throughout. These are as in CR2013 where they are presented and discussed in Section 3 of that paper.[7] Restriction A4 below extends Restriction A4 of CR2013 to the particular cases considered in this paper.

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. $\Box$

Restriction A2: The joint distribution of (Y, Z), F_{YZ}, is identified by the sampling process. $\Box$

Restriction A3: There is an $\mathcal{F}$-measurable function $h(\cdot,\cdot,\cdot) : \mathcal{R}_{YZU} \to \mathbb{R}$ such that

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

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

Restriction A4: The pair $\left(h, \mathcal{G}_{U|Z}\right)$ belongs to a known set of admissible structures $\mathcal{M}$. The model $\mathcal{M}$ contains restrictions as follows. One element of Y, denoted Y_*, only takes values in $\{1, \ldots, M\}$ and U has M components, $U = (U_1, \ldots, U_M)$, each of which may be vectors. The structural function has the form

$$h(y, z, u) = \sum_{m=1}^{M} 1[y_* = m] \times h_m(y, z, u)$$

where for $m \in \{1, \ldots, M\}$, $h_m(\cdot,\cdot,\cdot) : \mathcal{R}_{YZU} \to \mathbb{R}$ is continuous in its first and third arguments

[7] Restriction A2 in CR2013 requires that 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. The identification of conditional distributions $F_{Y|Z}(\cdot|z)$ for all $z \in \mathcal{R}_Z$ is equivalent to identification of the joint distribution of Y and Z.

In this paper conditional independence restrictions will require conditioning on components of Y together with Z in places, rather than only conditional on Z. This makes the statement of Restriction A2 as it appears here more natural in the present context.

and invariant with respect to variation in those elements of u not contained in u_m. $\square$

With regard to Restriction A3, the collection of admissible distributions specified may include restrictions on conditional distributions $G_{U|Y_*Z}(\cdot|y,z)$, each $(y_*,z) \in \mathcal{R}_{Y_*Z}$, where for all $\mathcal{S} \subseteq \mathcal{R}_{U|y_*z}$, $G_{U|Y_*Z}(\mathcal{S}|y_*,z) \equiv \mathbb{P}[U \in \mathcal{S}|y_*,z]$. In this case the components of $\mathcal{G}_{U|Z}$ are restricted to be such that there exists for each $z \in \mathcal{R}_Z$ conditional distributions $G_{U|Y_*Z}(\cdot|y_*,z)$ satisfying

$$G_{U|Z}(\cdot|z) = \int_{y_* \in \mathcal{R}_{Y_*}} G_{U|Y_*Z}(\cdot|y_*,z)\, dF_{Y_*|Z}(y_*|z).$$

Notation

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

is used to denote a collection of such conditional distributions where required.

Restriction A4 places restrictions on structural functions $h_m(\cdot,\cdot,\cdot)$ through the specification of admissible pairs $\left(h, \mathcal{G}_{U|Z}\right)$, which may include parametric or shape restrictions. There will in general also be restrictions on the covariation of observable and unobservable exogenous variables embodied in admissible $\mathcal{G}_{U|Z}$. Continuity of the structural functions $h_m(\cdot,\cdot,\cdot)$ in their first and third arguments is a sufficient condition to ensure that the sets $\mathcal{Y}(u,z;h_m)$ and $\mathcal{U}(y,z;h_m)$ are closed.

It should be noted that Restriction A4 places no restriction on the determination of y_* from the M counterfactual processes. For now we leave this selection process completely unspecified, noting that additional restrictions may be added through subsequent restrictions.

3.2 Identification: foundation results from CR2013

This Section extends results proved in CR2013 in order to provide the basis for the identification analysis to follow.[8] The distinguishing features of the results contained here stems from the need to work with conditional independence restrictions of the sort $U \perp\!\!\!\perp Y_*|Z$. This requires results to be stated conditional on realizations of exogenous variables Z as well as the classifier variable Y_*, rather conditional on Z alone as in CR2013. All of these results apply to the class of models considered in this paper when Restrictions A1 - A3 hold.

Our first result, Theorem 1, proven in the Appendix, builds on Theorem 2 of CR2013. This Theorem gives a characterization of identified sets in terms of a selectionability property of the distributions of unobservable variables admitted by a model. The random set $\mathcal{U}(Y,Z;h)$ which

[8] Theorem 1 uses the concept of selectionability. The probability distribution, F_A, of a point valued random variable is selectionable with respect the probability distribution of a random set, $\mathcal{A}$, if (i) there exists a random variable, A, distributed F_A, (ii) there exists a random set $\mathcal{A}^*$ with the same probability distribution as $\mathcal{A}$, such that $\mathbb{P}[A \in \mathcal{A}^*] = 1$.

appears in the theorem is defined as

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

Theorem 1 *Let Restrictions A1-A3 hold. Then the identified set of structures $\mathcal{M}^*(F_{YZ})$ are those $(h,\mathcal{G}_{U|Z})$ admitted by the model $\mathcal{M}$ such that for almost every $z \in \mathcal{R}_Z$ and each $y_* \in \{1,...,M\}$ there exist conditional probabilities $G_{U|Y_*Z}(\cdot|y_*,z)$ defined on measurable subsets of $\mathcal{R}_U$ such that*

1. $G_{U|Z}(\cdot|z) = \int\limits_{y_*\in\mathcal{R}_{Y_*}} G_{U|Y_*Z}(\cdot|y_*,z)\, dF_{Y_*|Z}(y_*|z)$.

2. $G_{U|Y_*Z}(\cdot|y_*,z)$ *is selectionable with respect to the conditional distribution of random set* $\mathcal{U}(Y,Z;h)$ *given* $(Y_*=y_*\wedge Z=z)$ *induced by the distribution of* Y *conditional on* $(Y_*=y_*\wedge Z=z)$ *as given by* F_{YZ}.

The following Corollary gives an alternative characterization of the identified set in terms of moment inequalities. This result follows from using Artstein's (1983) Inequality which gives necessary and sufficient conditions for selectionability in terms of containment functionals of random sets. This result is the analog of Corollary 1 in CR2013, which uses Artstein's Inequality to produce moment inequalities conditional on realizations of Z rather than on realizations of both Y_* and Z. The proof is a straightforward consequence of the selectionability statement in Theorem 1 and Corollary 1 of CR2013 and is omitted.

Corollary 1 *Under Restrictions A1-A3 the identified set can be written*

$$\mathcal{M}^*(F_{YZ}) \equiv \left\{ \begin{array}{c} (h,\mathcal{G}_{U|Z}) \in \mathcal{M} : \exists \mathcal{G}_{U|Y_*Z} \;\; s.t. \;\; \forall \mathcal{S} \in \mathsf{F}(\mathcal{R}_U), \\ C(\mathcal{S},h|y_*,z) \leq G_{U|Y_*,Z}(\mathcal{S}|y_*,z) \;\; a.e. \;\; (y_*,z) \in \mathcal{R}_{Y_*Z}, \\ and\; G_{U|Z}(\mathcal{S}|z) = \int\limits_{y_*\in\mathcal{R}_{Y_*}} G_{U|Y_*Z}(\mathcal{S}|y_*,z)\, dF_{Y_*|Z}(y_*|z) \;\; a.e. \;\; z \in \mathcal{R}_Z \end{array} \right\}, \quad (3.1)$$

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

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

is the conditional containment functional of the random set $\mathcal{U}(Y,Z;h)$ *when the conditional distribution of* Y *given* $(Y_*=y_*\wedge Z=z)$ *is as given by* F_{YZ}.

The collection of sets $\mathsf{F}(\mathcal{R}_U)$ is too large to inspect in practice. Theorem 2 below provides a smaller collection of *core-determining sets*, a concept introduced in Galichon and Henry (2011). Again where CR2013 provided results conditional on exogenous variables Z, we provide results conditional on Z and the discrete classifier Y_*, as required for consideration of core-determining sets under conditional independence restrictions involving Y_* and Z. This turns out to be a simple generalization of Theorem 3 of CR2013, with a formal statement given in Theorem 2. The proof of this Theorem and its Corollary are identical to those of CR2013 Theorem 3 and its Corollary upon substituting "y_*, z" for z in that paper and are therefore omitted.

First to state the results it is necessary to define two collections of sets, $\mathsf{U}(h, y_*, z)$: the conditional support of the random set $\mathcal{U}(Y, Z; h)$ given $(Y_* = y_* \wedge Z = z)$ and $\mathsf{U}^*(h, y_*, z)$: the collection of the unions of these sets.

Definition 1 *Under Restrictions A1-A3, the* ***conditional support of random set*** $\mathcal{U}(Y, Z; h)$ *given* $(Y_* = y_* \wedge Z = z)$ *is*

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

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

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

In the definition of $\mathsf{U}^*(h, y_*, z)$ we employ the notation

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

In the statement of Theorem 2 we use the notation.

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

We also define for any set $\mathcal{S} \subseteq \mathcal{R}_U$ and any $(h, y_*, z) \in \mathcal{H} \times \mathcal{R}_{Y_*} \times \mathcal{R}_Z$,

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

which are those sets on the support of $\mathcal{U}(\mathcal{Y}, Z; h)$ given $(Y_* = y_* \wedge Z = z)$ that are contained in $\mathcal{S}$.

Theorem 2 *Let Restrictions A1-A3 hold. Fix* $(h, y_*, z) \in \mathcal{H} \times \mathcal{R}_{Y_*} \times \mathcal{R}_Z$ *and a distribution function* $G_{U|Y_* Z}(\cdot|y_*, z)$. *Let* $\mathsf{Q}(h, y_*, z) \subseteq \mathsf{U}^*(h, y_*, z)$, *such that for any* $\mathcal{S} \in \mathsf{U}^*(h, y_*, z)$ *with* $\mathcal{S} \notin \mathsf{Q}(h, y_*, z)$, *there exist nonempty collections* $\mathsf{S}_1, \mathsf{S}_2 \in \mathsf{U}^{\mathcal{S}}(h, y_*, z)$ *with* $\mathsf{S}_1 \cup \mathsf{S}_2 = \mathsf{U}^{\mathcal{S}}(h, y_*, z)$ *such that*

$$\mathcal{S}_1 \equiv \bigcup_{\mathcal{T} \in \mathsf{S}_1} \mathcal{T},\ \mathcal{S}_2 \equiv \bigcup_{\mathcal{T} \in \mathsf{S}_2} \mathcal{T},\ \text{and } G_{U|Y_* Z}(\mathcal{S}_1 \cap \mathcal{S}_2 | y_*, z) = 0, \tag{3.2}$$

with $\mathcal{S}_1, \mathcal{S}_2 \in \mathsf{Q}(h, y_*, z)$. *Then* $C(\mathcal{S}, h|y_*, z) \leq G_{U|Y_* Z}(\mathcal{S}|y_*, z)$ *for all* $\mathcal{S} \in \mathsf{Q}(h, y_*, z)$ *implies that* $C(\mathcal{S}, h|y_*, z) \leq G_{U|Y_* Z}(\mathcal{S}|y_*, z)$ *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* $\mathsf{Q}(h, y_*, z)$ *is core-determining.*

Finally, Corollary 2 gives conditions under which a core determining set delivers a moment equality rather than a moment inequality.

Corollary 2 *Define*

$$\mathsf{Q}^{E}\left(h,y_{*},z\right)\equiv\left\{\mathcal{S}\in\mathsf{Q}\left(h,y_{*},z\right):\forall y\in\mathcal{R}_{Y|y_{*}z}\text{ either }\mathcal{U}\left(y,z;h\right)\subseteq\mathcal{S}\text{ or }\mathcal{U}\left(y,z;h\right)\cap\mathcal{S}=\emptyset\right\}.$$

Then, under the conditions of Theorem 2, the collection of equalities and inequalities

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

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

A consequence of Corollary 2 is that *all* members of a collection $\mathsf{Q}(h,y_{*},z)$ deliver equalities when the structural function h is such that either (i) every set on the conditional support of $\mathcal{Y}(U,Z;h)$ is singleton and/or (ii) every set on the conditional support of $\mathcal{U}(Y,Z;h)$ is singleton.

3.3 Moment inequalities absent restrictions on selection of Y_{*}

A further simplification of the core determining sets obtains when, in addition to Restrictions A1-A3, Restriction A4 is also imposed, absent further restrictions on the determination of Y_{*}. Without such restrictions, all sets $\mathcal{U}$ of the form $\mathcal{U}\left(y,z;h\right)$ for some $(y,z)\in\mathcal{R}_{YZ}$ are such that for all components $m\in\{1,...,M\}$ with $m\neq y_{*}$, $\mathcal{U}_{m}=\mathcal{R}_{U_{m}}$. To state this formally, we define

$$\mathcal{U}_{m}\left(y,z;h\right)\equiv\left\{u_{m}^{*}\in\mathcal{R}_{U_{m}|z}:\exists u\text{ s.t. }u_{m}=u_{m}^{*}\wedge h\left(y,z,u\right)=0\right\}$$

as the projection of $\mathcal{U}\left(y,z;h\right)$ onto its m^{th} component. Then we have the simplification that

$$\forall m\neq y_{*},\quad \mathcal{U}_{m}\left(y,z;h\right)=\mathcal{R}_{U_{m}}. \tag{3.3}$$

The conditional support of the random set $\mathcal{U}_{m}\left(Y,Z;h\right)$ conditional on $\left(Y_{*}=m\wedge Z=z\right)$ is

$$\mathsf{U}_{m}\left(h,z\right)\equiv\left\{\mathcal{U}_{m}\left(y,z;h\right):y_{*}=m\wedge y\in\mathcal{R}_{Y|y_{*}z}\right\}.$$

The projection of any set $\mathcal{S}$ onto its m^{th} component is

$$\mathcal{S}_{m}\equiv\left\{u_{m}^{*}\in\mathcal{R}_{U_{m}}:\exists u\in\mathcal{S}\text{ s.t. }u_{m}=u_{m}^{*}\right\}.$$

From Theorem 2 we have that all core determining sets, $\mathcal{S}\in\mathsf{Q}\left(h,y_{*},z\right)$ are unions of sets on the support of $\mathcal{U}(y, z; h)$. Thus from (3.3) all core-determining sets $\mathcal{S} \in \mathsf{Q}(h, y_*, z)$ satisfy

$$\forall m \neq y_*, \quad \mathcal{S}_m = \mathcal{R}_{U_m}. \tag{3.4}$$

Consideration of the conditional containment functional applied to such sets then gives

$$C(\mathcal{S}, h|m, z) \equiv \mathbb{P}[\mathcal{U}(Y, Z; h) \subseteq \mathcal{S}|Y_* = m, z] = \mathbb{P}[\mathcal{U}_m(Y, Z; h) \subseteq \mathcal{S}_m|Y_* = m, z], \tag{3.5}$$

which is the probability, conditional on $(Y_* = m \wedge Z = z)$, that the projection of $\mathcal{U}(Y, Z; h)$ onto its m^{th} component is contained in the projection of $\mathcal{S}$ onto its m^{th} component. Consequently, the identified set $\mathcal{M}^*(F_{YZ})$ can be succinctly characterized through inequalities involving only containment functionals for projection level sets $\mathcal{U}_m(Y, Z; h)$ applied to projections of test sets $\mathcal{S}$. We thus define containment functionals for projections of level sets for any test set $\mathcal{S}_m \subseteq \mathcal{R}_{U_m}$ as

$$C_m(\mathcal{S}_m, h|y_*, z) \equiv \mathbb{P}[\mathcal{U}_m(Y, Z; h) \subseteq \mathcal{S}_m|y_*, z]. \tag{3.6}$$

Likewise we have from (3.4) that

$$\forall \mathcal{S} \in \mathsf{Q}(h, y_*, z), \ G_{U|Y_*Z}(\mathcal{S}|m, z) = G_{U_m|Y_*Z}(\mathcal{S}_m|m, z). \tag{3.7}$$

Implications (3.5) and (3.7) together enable us to work in a lower dimensional space, namely that of $\mathcal{R}_{U_m}$ in the construction of core-determining sets, rather than $\mathcal{R}_U$. Specifically, we have that for any $(y_*, z) \in \mathcal{R}_{Y_*Z}$ and any test set $\mathcal{S} \in \mathsf{Q}(h, y_*, z)$, the containment functional inequality

$$C(\mathcal{S}, h|m, z) \leq G_{U|Y_*Z}(\mathcal{S}|m, z), \tag{3.8}$$

appearing in Corollary 1 holds if and only if[9]

$$C_m(\mathcal{S}_m, h|m, z) \leq G_{U_m|Y_*Z}(\mathcal{S}_m|m, z). \tag{3.9}$$

Lemma 1 characterizes a collection of core-determining sets on the lower dimensional space $\mathcal{R}_{U_m}$ sufficient to guarantee (3.9) holds for all closed $\mathcal{S}_m \subseteq \mathcal{R}_{U_m}$. Before stating the lemma we require the following definitions for any $(h, m, z) \in \mathcal{H} \times \mathcal{R}_{Y_*} \times \mathcal{R}_Z$.

$$\mathsf{U}^*_m(h, z) \equiv \{\mathcal{U}_m \subseteq \mathcal{R}_{U_m} : \mathcal{U}_m \text{ is a union of elements of } \mathsf{U}_m(h, z)\},$$

[9] From $\mathcal{S}_m = \mathcal{R}_{U_m}$ for all $m \neq y_*$, (3.8) $\Rightarrow$(3.9) is immediate. The reverse implication is formally proven in the proof of Theorem 3.

and for any set $\mathcal{S}_m \subseteq \mathcal{R}_{U_m}$,

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

which are those sets on the conditional support of $\mathcal{U}_m(Y,Z;h)$ conditional on $(Y_* = m \wedge Z = z)$ that are contained in $\mathcal{S}_m$. With this notation in hand, the proof of the following lemma is a straightforward extension of Theorem 2 and is omitted.

Lemma 1 *Let Restrictions A1-A4 hold. Fix $(h,m,z) \in \mathcal{H} \times \mathcal{R}_{Y_*} \times \mathcal{R}_Z$ and a distribution function $G_{U|Y_*Z}(\cdot|y_*,z)$. Let $\mathsf{Q}_m(h,z) \subseteq \mathsf{U}^*_m(h,z)$, such that for any $\mathcal{S}_m \in \mathsf{U}^*_m(h,z)$ with $\mathcal{S}_m \notin \mathsf{Q}_m(h,z)$, there exist nonempty collections $\mathsf{S}_{m1}, \mathsf{S}_{m2} \in \mathsf{U}^{\mathcal{S}_m}(h,z)$ with $\mathsf{S}_{m1} \cup \mathsf{S}_{m2} = \mathsf{U}^{\mathcal{S}_m}(h,z)$ such that*

$$\mathcal{S}_{m1} \equiv \bigcup_{\mathcal{T} \in \mathsf{S}_{m1}} \mathcal{T},\ \mathcal{S}_{m2} \equiv \bigcup_{\mathcal{T} \in \mathsf{S}_{m2}} \mathcal{T},\ \text{and } G_{U|Y_*Z}(\mathcal{S}_{m1} \cap \mathcal{S}_{m2}|y_*,z) = 0, \tag{3.10}$$

with $\mathcal{S}_{m1}, \mathcal{S}_{m2} \in \mathsf{Q}_m(h,z)$. Then $C_m(\mathcal{S}_m, h|m,z) \leq G_{U_m|Y_*Z}(\mathcal{S}_m|m,z)$ for all $\mathcal{S}_m \in \mathsf{Q}_m(h,z)$ implies that $C_m(\mathcal{S}_m, h|m,z) \leq G_{U_m|Y_*Z}(\mathcal{S}_m|m,z)$ holds for all $\mathcal{S}_m \subseteq \mathcal{R}_{U_m}$, and in particular for $\mathcal{S}_m \in \mathsf{F}(\mathcal{R}_{U_m})$, so that the collection of sets $\mathsf{Q}_m(h,z)$ is core-determining.

The following Theorem, proven in the Appendix, uses this lemma *en route* to characterizing the identified set $\mathcal{M}^*(F_{YZ})$ under Restrictions A1-A4 through conditional containment functional inequalities defined on $\mathcal{R}_{U_m}$, $m \in \{1,...,M\}$.

Theorem 3 *Let Restrictions A1-A4 hold, with no further restrictions imposed on the determination of the classifier Y_*. Given collection of conditional distributions $\mathcal{G}_{U|Y_*Z}$ we have that*

$$\forall \mathcal{S} \in \mathsf{F}(\mathcal{R}_U),\ C(\mathcal{S},h|m,z) \leq G_{U|Y_*Z}(\mathcal{S}|m,z) \ \ a.e.\ \ (y_*,z) \in \mathcal{R}_{Y_*Z}$$

if and only if

$$\forall m \in \{1,...,M\},\ \forall \mathcal{S} \in \mathsf{Q}_m(h,z),\ C_m(\mathcal{S},h|m,z) \leq G_{U_m|Y_*Z}(\mathcal{S}|m,z) \ \ a.e.\ \ (y_*,z) \in \mathcal{R}_{Y_*Z}.$$

Hence

$$\mathcal{M}^*(F_{YZ}) = \left\{ \begin{array}{c} \left(h,\mathcal{G}_{U|Z}\right) \in \mathcal{M} : \exists \mathcal{G}_{U|Y_*Z} \ s.t.\ \forall m \in \{1,...,M\},\ \forall \mathcal{S} \in \mathsf{Q}_m(h,z), \\ C_m(\mathcal{S},h|m,z) \leq G_{U_m|Y_*Z}(\mathcal{S}|m,z) \ \ a.e.\ \ (m,z) \in \mathcal{R}_{Y_*Z},\ and \\ G_{U_m|Z}(\mathcal{S}|z) = \displaystyle\int_{y_* \in \mathcal{R}_{Y_*}} G_{U_m|Y_*Z}(\mathcal{S}|y_*,z)\, dF_{Y_*|Z}(y_*|z) \ \ a.e.\ \ z \in \mathcal{R}_Z \end{array} \right\}.$$

3.4 The identifying power of a conditional independence restriction

The models studied in this paper include a *conditional* independence Restriction CI.

Restriction CI. Let Y_* be the classifier element of Y. Random variables U and Y_* are independently distributed conditional on $Z = z$ for every $z \in \mathcal{R}_Z$.

Restriction CI places restrictions on the collection of distributions $\mathcal{G}_{U|Z}$ in admissible structures, namely that for all sets $\mathcal{S} \subset \mathcal{R}_{U|Z}$, the conditional distribution of U given (Y_*, Z), $G_{U|Y_*Z}(\cdot|y_*, z)$ satisfies $G_{U|Y_*Z}(\mathcal{S}|y_*, z) = G_{U|Z}(\mathcal{S}|z)$ a.e. $(y_*, z) \in \mathcal{R}_{Y_*Z}$. A consequence is equality of the conditional support of unobserved heterogeneity and its components, that is that $\mathcal{R}_{U|y_*z} = \mathcal{R}_{U|z}$ and $\mathcal{R}_{U_m|y_*z} = \mathcal{R}_{U_m|z}$, for all $m \in \{1, \ldots, M\}$.

We build on the characterization of the identified set given in Theorem 3 to develop a characterization of the identified set when there is a conditional independence condition. The result is given in Theorem 4, the proof of which is given in the Appendix.

Theorem 4 *Let Restrictions A1-A3 hold. A model $\mathcal{M}$ which embodies Restriction A4 and the conditional independence restriction CI has an identified set $\mathcal{M}^*(\mathcal{F}_{YZ})$ which can be written as*

$$\mathcal{M}^*(F_{YZ}) \equiv \Big\{ \big(h, \mathcal{G}_{U|Z}\big) \in \mathcal{M} : \forall m \in \{1, \ldots, M\}, \forall \mathcal{S} \in \mathsf{Q}_m(h, z), \\ C_m(\mathcal{S}, h|m, z) \leq G_{U_m|Z}(\mathcal{S}|z), \text{ a.e. } z \in \mathcal{R}_Z \Big\}.$$

Here $\mathcal{S} \subseteq \mathcal{R}_{U_m|z}$, and $\mathsf{Q}_m(h, z)$ is a collection of closed subsets of $\mathcal{R}_{U_m|z}$ comprising unions of sets on the conditional support of $\mathcal{U}_m(Y, Z; h)$ given $Z = z$ and $Y_* = m$ defined in Lemma 1.

Remarks

1. Regarding the collections of distributions $\mathcal{G}_{U|Z}$, the identified set in Theorem 4 only places restrictions on the *marginal* distributions, $G_{U_m|Z}(\cdot|z)$, $m \in \{1, \ldots, M\}$. Data is never informative about the covariation of U_m and $U_{m'}$, for any $m \neq m'$.

2. Applying the unprojected version of the inequality in the definition of the set $\mathcal{M}^*(\mathcal{F}_{YZ})$ in Theorem 4 to the complement, $\mathcal{S}^c$, of a set $\mathcal{S}$ gives an upper bound on $G_{U|Z}(\mathcal{S}|z)$ and thus a two-sided inequality that must hold for almost every $z \in \mathcal{R}_Z$:

$$\forall m, n \in \{1, \ldots, M\}, \; \mathcal{S} \subseteq \mathcal{R}_U\text{: } C(\mathcal{S}, h|m, z) \leq G_{U|Z}(\mathcal{S}|z) \leq 1 - C(\mathcal{S}^c, h|n, z).$$

   This representation leads directly a characterization of bounds on structural function h without direct reference to a distribution of unobserved heterogeneity $G_{U|Z}(\mathcal{S}|z)$.

Example 1 continued. In the simple treatment effect model the projected u-level sets $\mathcal{U}_m(Y, Z; h)$ are singleton and a small modification to the argument that leads to Corollary 2 leads to the conclusion that the inequalities in the definition of $\mathcal{M}^*(\mathcal{F}_{YZ})$ in Theorem 4 reduce to equalities. For any set $\mathcal{S} \subseteq \mathsf{Q}_m(h, z)$,

$$G_{U_m|Z}(\mathcal{S}|z) = F_{Y_1|Y_2,Z}(\mathcal{S}|m, z)$$

and it follows that for $m \in \{1, \dots, M\}$:

1. each conditional distribution function of U_m given $Z = z$ is point identified by the conditional distribution function of Y_1 given $Y_2 = m$ and $Z = z$,
2. each marginal distribution function of U_m is point identified by the expected value with respect to Z of the conditional distribution function of Y_1 given $Y_2 = m$ and $Z = z$,
3. which leads directly to the familiar results on point identification of the Average and Quantile Treatment Effects.

The analysis applies directly when there are vector counterfactual outcomes, $U_1, \dots, U_M$, in the treatment effect model. $\square$

3.5 The additional identifying power of marginal independence conditions

Theorem 4 provides a characterization of the identified set of structures $(h, \mathcal{G}_{U|Z})$ delivered by a model of counterfactual processes embodying Restriction A4 and the conditional independence restriction CI. In models of processes more complex than found in the treatment effects case there may be additional marginal independence restrictions. We consider Restriction MI.

Restriction MI. Let $e(Z)$ be a vector-valued function of Z. Random variables U_m and $e(Z)$ are independently distributed for each $m \in \mathcal{R}_{Y_*}$.

Restriction MI restricts the set of admissible structures$(h, \mathcal{G}_{U|Z}) \in \mathcal{M}$ to be those with U_m and $e(Z)$ independently distributed for all $m \in \{1, \dots, M\}$. A common choice for a function $e(\cdot)$ will be a function that selects certain elements from Z, for example, with $Z = (Z_1, Z_2)$, $e(Z) = Z_1$.[10]

Theorem 5 provides a characterization of the identified set delivered by a model embodying the conditional *and* marginal independence restrictions CI and MI.

Theorem 5 *Let Restrictions A1-A3 hold. A model $\mathcal{M}$ which embodies Restriction A4 and the independence restrictions CI and MI has an identified set $\mathcal{M}^*(\mathcal{F}_{Y|Z})$ which can be written as follows.*

$$\mathcal{M}^*(F_{YZ}) \equiv \left\{ \begin{array}{c} (h, \mathcal{G}_{U|Z}) \in \mathcal{M} : \forall m \in \{1, \dots, M\}, \forall \mathcal{S} \in \mathsf{Q}_m(h, z), \\ C_m(\mathcal{S}, h|m, z) \leq G_{U_m|Z}(\mathcal{S}|z), \ a.e. \ z \in \mathcal{R}_Z \end{array} \right\},$$

where $\mathsf{Q}(h_m, z)$ is the collection of core determining sets defined in Lemma 1.

This characterization appears the same as that of Theorem 4, but it differs because now admissible structures $(h, \mathcal{G}_{U|Z}) \in \mathcal{M}$ are required to be such that $\mathcal{G}_{U|Z}$ satisfies Restriction MI in addition

[10] It would be easy to relax the marginal independence restriction to $U_m \perp\!\!\!\perp e_m(Z)$, $m \in \{1, \dots, M\}$.

to Restriction CI. Thus the identified set of Theorem 5 is subset of that of Theorem 4 because the conditional containment inequality must hold for some $\left(h, \mathcal{G}_{U|Z}\right)$ in this more restrictive collection of admissible structures.

Sharpness is immediate because for any $\mathcal{S} \in \mathcal{R}_{U_m}$, under Restriction CI

$$C_m\left(\mathcal{S}, h|m, z\right) \leq G_{U_m|Z}\left(\mathcal{S}|z\right) \Rightarrow C_m\left(\mathcal{S}, h|m, z\right) \leq G_{U_m|Y_* Z}\left(\mathcal{S}|m, z\right).$$

This is required to hold for all (m, z) and for all core-determining sets, so the selectionability statement of Theorem 1 is satisfied. Again, the difference with Theorem 4 is that the distributions $G_{U_m|Z}$ are now required to belong to more restrictive collections of conditional distributions, namely we have as a requirement of admissible structures that for each $e \in \mathcal{R}_{e(Z)}$,

$$G_{U_m|Z}\left(\mathcal{S}|Z \in \mathcal{Z}_e\right) = G_{U_m}\left(\mathcal{S}\right), \text{ where } \mathcal{Z}_e \equiv \left\{z : e\left(Z\right) = e\right\}. \tag{3.11}$$

The characterization of $\mathcal{M}^*\left(\mathcal{F}_{YZ}\right)$ in the Theorem 5 produces interesting observable implications that may not appear immediate, but which provide bounds on $\left(h, \mathcal{G}_{U|Z}\right)$, potentially non-sharp in isolation. These implications may prove beneficial in developing sufficient conditions for point identification of $\left(h, \mathcal{G}_{U|Z}\right)$ or features of $\left(h, \mathcal{G}_{U|Z}\right)$ in particular models. Two such implications are as follows.

1. For any $m \in \mathcal{R}_{Y_*}$, $e \in \mathcal{R}_{e(Z)}$, and any $\mathcal{S} \subseteq \mathcal{R}_{U_m}$,

   $$E\left[C_m\left(\mathcal{S}, h|m, Z\right)|e(Z) = e\right] \leq G_{U_m}\left(\mathcal{S}\right). \tag{3.12}$$

   This follows from integrating both sides of the inequality $C_m\left(\mathcal{S}, h|m, z\right) \leq G_{U_m|Z}\left(\mathcal{S}|z\right)$ as follows. First we have from the left hand side,

   $$\begin{aligned} \frac{1}{F_Z\left(\mathcal{Z}_e\right)} \int\limits_{z \in \mathcal{Z}_e} C\left(\mathcal{S}, h_m|y_*, z\right) dF_Z\left(z\right) &= E\left[C\left(\mathcal{S}, h_m|y_*, Z\right)|Z \in \mathcal{Z}_e\right] \\ &= E\left[C\left(\mathcal{S}, h_m|y_*, Z\right)|e\left(Z\right) = e\right]. \end{aligned}$$

   Then multiplying the right hand side by $\frac{1}{F_Z(\mathcal{Z}_e)}$ and integrating we obtain

   $$\frac{1}{F_Z\left(\mathcal{Z}_e\right)} \int\limits_{z \in \mathcal{Z}_e} G_{U_m|Z}\left(\mathcal{S}|z\right) dF_Z\left(z\right) = G_{U_m|Z}\left(\mathcal{S}|Z \in \mathcal{Z}_e\right) = G_{U_m}\left(\mathcal{S}\right),$$

   where the final equality follows from Restriction MI.

   It is interesting to note that the expression $E\left[C_m\left(\mathcal{S}, h|m, Z\right)|e(Z) = e\right]$ is a conditional expectation of the containment functional $C_m\left(\mathcal{S}, h|m, Z\right)$ holding m fixed, which may in general differ from $C_m(\mathcal{S}, h|Y_* = m, e(Z) = e)$.

2. For any $m \in \mathcal{R}_{Y_*}$, $e \in \mathcal{R}_{e(Z)}$, and any $\mathcal{S} \subseteq \mathcal{R}_{U_m}$,

$$C_m(\mathcal{S}, h|Z \in \mathcal{Z}_e) \leq G_{U_m|Z}(\mathcal{S}|Z \in \mathcal{Z}_e) = G_{U_m}(\mathcal{S}),$$

by Restriction MI.

Remarks

1. Since the bounded probabilities, $G_{U_m}(\mathcal{S}) = G_{U_m|Z}(\mathcal{S}|Z \in \mathcal{Z}_e)$, do not depend on the value e of $e(Z)$ for each value m and set $\mathcal{S}$ only the supremum of the lower bounding expression over values $e \in \mathcal{R}_{e(Z)}$ is instrumental in (3.12).

2. In the common case in which $Z = (Z_1, Z_2)$ and $e(Z) = Z_1$ is a selection of the elements in Z,

$$E_Z[\cdot|e(Z) = e] = E_{Z_2}[\cdot|Z_1 = e].$$

3. Arguing as in Remark 2 following Theorem 4, a two-sided inequality is obtained:

$$E_Z[C_m(\mathcal{S}, h|m, z)|e(Z) = e_L] \leq G_{U_m}(\mathcal{S}) \leq 1 - E_Z[C_m(\mathcal{S}^c, h|m, z)|e(Z) = e_U],$$

which must hold for all $(e_L, e_U) \in \mathcal{R}_{e(Z)}$.

Example 3 continued. For simplicity in this illustration exogenous variables z_1 are excluded from the threshold function, so $g_m(y_2, z_1)$ is written $g_m(y_2)$ and since Y_2 is binary the structural function $h(y, z, u)$ is characterized by four parameters: $\theta \equiv (g_1(0), g_1(1), g_2(0), g_2(1))$. Recall $m = 1$ for people who attend a training programme and $m = 2$ for people who do not. Thus, $g_1(0)$ is the threshold parameter for a person who does attend a training programme and is not in receipt of benefit payment. We can normalize the threshold functions so that each U_m is marginally uniformly distributed on the unit interval and then there is the following representation.

$$\text{In state } m\text{:} \quad Y_1 = \begin{cases} 0 \ , & 0 \leq U_m \leq g_m(Y_2) \\ \\ 1 \ , & g_m(Y_2) \leq U_m \leq 1 \end{cases}.$$

The set up here is similar to that in Chesher and Rosen (2013b), henceforth CR2013b. In that paper there was only one state, so there $U_1 = U_2$ (denoted U in that paper) and $g_1(y_2) = g_2(y_2)$ (denoted $p(y_2)$ in that paper). In CR2013b there was no conditional independence restriction but there was a marginal independence restriction $U \perp\!\!\!\perp Z$. For ease of comparison with CR2013b the characterization of the identified set is presented here in terms of $1-g_m(y_2)$, $m \in \{1,2\}$,, $y_2 \in \{0,1\}$ which are counterfactual probabilities of return to work in state m with benefit receipt indicator equal to y_2. Define probabilities which could be estimated using data, as follows.

$$f_{ij}(z,m) \equiv \mathbb{P}[Y_1 = i \wedge Y_2 = j | Y_3 = m, Z = z], \quad (i,j) \in \{0,1\} \times \{0,1\}, \; m \in \{1,2\}$$

Applying Theorem 4, under the conditional independence restriction, $(U_1, U_2) \perp\!\!\!\perp Y_3 | Z$, the identified set of structures $(\theta, \mathcal{G}_{U|Z})$ is characterized by the following inequalities which hold for $m \in \{1,2\}$ and almost every $z \in \mathcal{R}_Z$.

For $g_m(0) \leq g_m(1)$:

$$f_{10}(z,m) + f_{11}(z,m) \quad \leq 1 - G_{U_m|Z}(g_m(0)|z) \leq \quad 1 - f_{00}(z,m)$$

$$f_{11}(z,m) \quad \leq 1 - G_{U_m|Z}(g_m(1)|z) \leq \quad f_{10}(z,m) + f_{11}(z,m)$$

For $g_m(0) \geq g_m(1)$

$$f_{10}(z,m) \quad \leq 1 - G_{U_m|Z}(g_m(0)|z) \leq \quad f_{10}(z,m) + f_{11}(z,m)$$

$$f_{10}(z,m) + f_{11}(z,m) \quad \leq 1 - G_{U_m|Z}(g_m(1)|z) \leq \quad 1 - f_{01}(z,m)$$

We now apply Theorem 5 and impose the marginal independence restriction $(U_1, U_2) \perp\!\!\!\perp Z_1$ jointly with the conditional independence condition $(U_1, U_2) \perp\!\!\!\perp Y_3 | Z$ where $Z = (Z_1, Z_2)$. The inequalities (3.12) deliver the following additional inequalities which hold for $m \in \{1,2\}$.

For $g_m(0)) \leq g_m(1)$:

$$\sup_{z_1 \in \mathcal{R}_{Z_1}} E_{Z_2|Z_1=z_1}[f_{10}(z,m) + f_{11}(z,m)] \quad \leq 1 - g_m(0) \leq \quad \inf_{z_1 \in \mathcal{R}_{Z_1}} \left(1 - E_{Z_2|Z_1=z_1}[f_{00}(z,m)]\right)$$

$$\sup_{z_1 \in \mathcal{R}_{Z_1}} E_{Z_2|Z_1=z_1}[f_{11}(z,m)] \quad \leq 1 - g_m(1) \leq \quad \inf_{z_1 \in \mathcal{R}_{Z_1}} E_{Z_2|Z_1=z_1}[f_{10}(z,m) + f_{11}(z,m)]$$

For $g_m(0) \geq g_m(1)$:

$$\sup_{z_1 \in \mathcal{R}_{Z_1}} E_{Z_2|Z_1=z_1}[f_{10}(z,m)] \quad \leq 1 - g_m(0) \leq \quad \inf_{z_1 \in \mathcal{R}_{Z_1}} E_{Z_2|Z_1=z_1}[f_{10}(z,m) + f_{11}(z,m)]$$

$$\sup_{z_1 \in \mathcal{R}_{Z_1}} E_{Z_2|Z_1=z_1}[f_{10}(z,m) + f_{11}(z,m)] \quad \leq 1 - g_m(1) \leq \quad \inf_{z_1 \in \mathcal{R}_{Z_1}} \left(1 - E_{Z_2|Z_1=z_1}[f_{01}(z,m)]\right)$$

4 Concluding remarks

We have presented an extension of a treatment effect model in which a discrete classifier variable indicates in which one of a number of counterfactual processes an individual engages. The observed process delivers realizations of endogenous variables and values of exogenous variables are available.

We have considered models of counterfactual processes which may be incomplete. Such models can arise when a process involves multiple equilibria and no equilibrium selection mechanism is specified, when a process is defined by inequality restrictions as in some auction models and when only some elements of a simultaneous equations system that determines values of endogenous variables are specified.

We have considered models which place no structure on the determination of the classifier variable but impose a conditional independence restriction requiring the unobservable variables that deliver stochastic variation in the counterfactual processes and the classifier variable to be independently distributed conditional on some observed exogenous variables. Our models may incorporate additionally, marginal independence restrictions requiring unobservable variables and known functions of exogenous variables to be independently distributed.

Using tools from random set theory and in particular the concept of selectionability, we have developed characterizations of the sharp identified sets delivered by these models.

In research in progress we are studying the identifying power of alternative covariation restrictions, for example conditional mean and quantile independence and we are developing characterizations of identified sets in more general cases in which there are combinations of conditional and marginal independence restrictions.

References

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.

Chesher, A. (2009): "Excess Heterogeneity, Endogeneity and Index Restrictions," *Journal of Econometrics*, 152(1), 35–47.

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.

——— (2013a): "Generalized Instrumental Variable Models," CeMMAP working paper CWP43/13.

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

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

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

Heckman, J. J., and R. Robb (1985): "Alternative Methods for Evaluating the Impact of Interventions," in *Longitudinal Analysis of Labor Market Data*, ed. by J. J. Heckman, and B. Singer, pp. 156–245. Cambridge University Press, Econometric Society Monograph Series.

Heckman, J. J., S. Urzua, and E. Vytlacil (2008): "Instrumental Variables in Models with Multiple Outcomes: the General Unordered Case," *Annals of Economics and Statistics*, 91/92, 151–174.

Heckman, J. J., and E. Vytlacil (2005): "Structural Equations, Treatment Effects and Econometric Policy Evaluation," *Econometrica*, 73(3), 669–738.

——— (2007): "Econometric Evaluation of Social Programs Part I: Causal Models, Structural Models, and Econometric Policy Evaluation," in *The Handbook of Econometrics*, ed. by J. J. Heckman, and E. E. Leamer, vol. 6b, pp. 4779–4874. North-Holland.

Imbens, G., and W. Newey (2009): "Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity," *Econometrica*, 77(5), 1481–1512.

Rosenbaum, P. R., and D. B. Rubin (1983): "The central role of the propensity score in observational studies for causal effects," *Biometrika*, 70(1), 41–55.

Rubin, D. B. (1974): "Estimating causal effects of treatments in randomized and nonrandomized studies," *Journal of Educational Psychology*, 66(5), 688–701.

A Appendix: Proofs

Proof of Theorem 1. Theorem 2 of CR2013 states that under Restrictions A1-A3 of that paper, identical to Restrictions A1-A3 here, the identified set of structures $(h, \mathcal{G}_{U|Z})$ are those such that

$$G_{U|Z}(\cdot|z) \precsim \mathcal{U}(Y, z; h) \text{ when } Y \sim F_{Y|Z}(\cdot|z), \text{ a.e. } z \in \mathcal{R}_Z, \tag{A.1}$$

where "$\precsim$" means "is selectionable with respect to the distribution of", as in CR2013. This statement has the following interpretation.

1. There exists a random variable $\tilde{U}$ such that for almost every $z \in \mathcal{R}_Z$, $\tilde{U} \sim G_{U|Z}(\cdot|z)$ conditional on $Z = z$.
2. There exists a random variable $\tilde{Y}$ such that for almost every $z \in \mathcal{R}_Z$, $\tilde{Y} \sim F_{Y|Z}(\cdot|z)$ conditional on $Z = z$.
3. $\tilde{U}$ and $\tilde{Y}$ belong to probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and $\mathbb{P}\left[\tilde{U} \in \mathcal{U}\left(\tilde{Y}, Z; h\right) | Z = z\right] = 1$ a.e. $z \in \mathcal{R}_Z$.

To prove the theorem it is required to show that (A.1) is equivalent to the existence of a collection of conditional distributions $\mathcal{G}_{U|Y_*Z} \equiv \left\{G_{U|Y_*Z}(\cdot|y_*, z) : (y_*, z) \in \mathcal{R}_{Y_*Z}\right\}$ such that:

1. For almost every $z \in \mathcal{R}_Z$:

$$G_{U|Z}(\cdot|z) = \int\limits_{y_* \in \mathcal{R}_{Y_*}} G_{U|Y_*Z}(\cdot|y_*, z)\, dF_{Y_*|Z}(y_*|z), \text{ and} \tag{A.2}$$

2. For almost every $(y_*, z) \in \mathcal{R}_{Y_*Z}$:

$$G_{U|Y_*Z}(\cdot|y_*, z) \precsim \mathcal{U}(Y, z; h) \text{ when } Y \sim F_{Y|Y_*Z}(\cdot|y_*, z). \tag{A.3}$$

To show this start with (A.1), from which we have, with $\tilde{U}$ and $\tilde{Y}$ as defined in bullet points 1-3,

$$1 = \mathbb{P}\left[\tilde{U} \in \mathcal{U}\left(\tilde{Y}, Z; h\right) | Z = z\right] = \int\limits_{\mathcal{R}_{Y_*|z}} \mathbb{P}\left[\tilde{U} \in \mathcal{U}\left(\tilde{Y}, Z; h\right) | \tilde{Y}_* = y_*, Z = z\right] dF_{Y_*|Z}(y_*|z),$$

where $\tilde{Y} \sim F_{Y|Z}(\cdot|z)$ conditional on $Z = z$. This can hold if and only if

$$\mathbb{P}\left[\tilde{U} \in \mathcal{U}\left(\tilde{Y}, Z; h\right) | \tilde{Y}_* = y_*, Z = z\right] = 1 \text{ a.e. } (y_*, z) \in \mathcal{R}_{Y_*Z},$$

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

Now define $G_{U|Y_*Z}(\cdot|y_*, z)$ such that for any $\mathcal{S} \in \mathcal{R}_U$,

$$G_{U|Y_*Z}(\mathcal{S}|y_*, z) \equiv \mathbb{P}\left[\tilde{U} \in \mathcal{S}|\tilde{Y}_* = y_*, Z = z\right].$$

Consequently, from Restriction A3 and the first consequence of (A.1) above, $G_{U|Z}(\mathcal{S}|z) = \mathbb{P}\left[\tilde{U} \in \mathcal{S}|Z = z\right]$, and then from the law of total probability, (A.2) holds. Then we have that (A.3) holds since

1. There exists a random variable $\tilde{U}$ such that for almost every $z \in \mathcal{R}_Z$, $\tilde{U} \sim G_{U|Z}(\cdot|z)$ conditional on $Z = z$, and such that for almost every $(y_*, z) \in \mathcal{R}_{Y_*Z}$, $\tilde{U} \sim G_{U|Y_*Z}(\cdot|y_*, z)$ conditional on $Z = z$, $Y_* = y_*$.

2. There exists a random variable $\tilde{Y}$ such that for almost every $z \in \mathcal{R}_Z$, $\tilde{Y} \sim F_{Y|Z}(\cdot|z)$ conditional on $Z = z$.

3. $\tilde{U}$ and $\tilde{Y}$ belong to probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and $\mathbb{P}\left[\tilde{U} \in \mathcal{U}\left(\tilde{Y}, Z; h\right)|\tilde{Y}_* = y_*, Z = z\right] = 1$ a.e. $(y_*, z) \in \mathcal{R}_{Y_*Z}$.

That (A.3) implies (A.1) is immediate, and so equivalence is proved. $\square$

Proof of Theorem 3

Fix $(m, z) \in \mathcal{R}_{Y_*Z}$. From Lemma 1 we have that

$$\forall \mathcal{S} \in \mathsf{Q}_m(h, z), C_m(\mathcal{S}, h|m, z) \leq G_{U_m|Y_*Z}(\mathcal{S}|m, z)$$

implies that

$$\forall \mathcal{S} \in \mathsf{F}(\mathcal{R}_{U_m}), C_m(\mathcal{S}, h|m, z) \leq G_{U_m|Y_*Z}(\mathcal{S}|m, z).$$

We need to show that (3.9),

$$C_m(\mathcal{S}, h|m, z) \leq G_{U_m|Y_*Z}(\mathcal{S}|m, z), \tag{A.4}$$

for all $\mathcal{S} \in \mathsf{F}(\mathcal{R}_{U_m})$ implies that (3.8),

$$C(\mathcal{S}, h|m, z) \leq G_{U|Y_*Z}(\mathcal{S}|m, z).$$

for all $\mathcal{S} \in \mathsf{F}(\mathcal{R}_U)$.

To show this, start with

$$C(\mathcal{S}, h|m, z) \leq G_{U|Y_*Z}(\mathcal{S}|m, z). \tag{A.5}$$

for an arbitrary $\mathcal{S} \in \mathsf{F}(\mathcal{R}_U)$.

First suppose that it does not hold that the projection of $\mathcal{S}$ onto its n^{th} projection $\mathcal{S}_n$, $n \neq m$, is equal to $\mathcal{R}_{U_n}$. All elements of the support of $\mathcal{U}(Y, Z; h)$ conditional on $(Y_*, Z) = (m, z)$ have $\mathcal{U}_n(Y, Z; h) = \mathcal{R}_{U_n}$, implying that $C(\mathcal{S}, h|m, z) = 0$ and (A.5) is trivially satisfied.

We now turn to sets $\mathcal{S}$ with n^{th} projection $\mathcal{S}_n$, $n \neq m$, equal to $\mathcal{R}_{U_n}$. In this case

$$C(\mathcal{S}, h|m, z) = C_m(\mathcal{S}_m, h|m, z),$$

and

$$G_{U|Y_*Z}(\mathcal{S}|m, z) = G_{U_m|Y_*Z}(\mathcal{S}_m|m, z),$$

so that (A.5) is in fact equivalent to (A.4), completing the proof. $\square$

Proof of Theorem 4

We start with the characterization of the identified set given in Theorem 3:

$$\mathcal{M}^*(F_{YZ}) = \left\{ \begin{array}{c} \left(h, \mathcal{G}_{U|Z}\right) \in \mathcal{M} : \exists \mathcal{G}_{U|Y_*Z} \text{ s.t. } \forall m \in \{1, ..., M\}, \forall \mathcal{S} \in \mathsf{Q}_m(h, z), \\ C_m(\mathcal{S}, h|m, z) \leq G_{U_m|Y_*Z}(\mathcal{S}|m, z) \text{ a.e. } (m, z) \in \mathcal{R}_{Y_*Z}, \text{ and} \\ G_{U_m|Z}(\mathcal{S}|z) = \displaystyle\int_{y_* \in \mathcal{R}_{Y_*}} G_{U_m|Y_*Z}(\mathcal{S}|y_*, z)\, dF_{Y_*|Z}(y_*|z) \text{ a.e. } z \in \mathcal{R}_Z \end{array} \right\}.$$

Using Restriction CI $G_{U_m|Y_*Z}(\mathcal{S}_m|m, z) = G_{U_m|Z}(\mathcal{S}_m|z)$ so we obtain

$$\mathcal{M}^*(F_{YZ}) = \left\{ \begin{array}{c} \left(h, \mathcal{G}_{U|Z}\right) \in \mathcal{M} : \forall m \in \{1, ..., M\}, \forall \mathcal{S} \in \mathsf{Q}_m(h, z), \\ C_m(\mathcal{S}, h|m, z) \leq G_{U_m|Z}(\mathcal{S}|z) \text{ a.e. } (m, z) \in \mathcal{R}_{Y_*Z} \end{array} \right\},$$

equivalently

$$\mathcal{M}^*(F_{YZ}) = \left\{ \left(h, \mathcal{G}_{U|Z}\right) \in \mathcal{M} : \sup_{(m,z) \in \mathcal{R}_{Y_*Z}} \sup_{\mathcal{S} \in \mathsf{Q}_m(h,z)} C_m(\mathcal{S}, h|m, z) - G_{U_m|Y_*Z}(\mathcal{S}|z) \leq 0 \right\}.$$

$\square$

Proof of Theorem 5

The Theorem is proved using the same argument as in the proof of Theorem 4 but now with structures $\left(h, \mathcal{G}_{U|Z}\right)$ required to belong to a more restrictive set such that Restriction CI and Restriction MI both hold. Thus the set of structures $\left(h, \mathcal{G}_{U|Z}\right)$ satisfying these restrictions (i.e. those such that $\left(h, \mathcal{G}_{U|Z}\right) \in \mathcal{M}$) and also satisfying the condition stated in the Theorem, namely

$$\forall m \in \{1, \ldots, M\}, \forall \mathcal{S} \in \mathsf{Q}_m(h, z), C_m(\mathcal{S}, h|m, z) \leq G_{U_m|Z}(\mathcal{S}|z), \text{ a.e. } z \in \mathcal{R}_Z$$

are by Theorem 3 precisely those $\left(h, \mathcal{G}_{U|Z}\right) \in \mathcal{M}$ satisfying

$$\forall \mathcal{S} \in \mathsf{F}\left(\mathcal{R}_U\right), \, C\left(\mathcal{S}, h|m, z\right) \leq G_{U|Y_*Z}\left(\mathcal{S}|m, z\right) \text{ a.e. } (y_*, z) \in \mathcal{R}_{Y_*Z},$$

where the conditional distribution of U given (Y_*, Z) satisfies the conditional independence restriction

$$G_{U|Y_*Z}\left(\mathcal{S}|m, z\right) = G_{U|Z}\left(\mathcal{S}|z\right).$$

Application of Artstein's Inequality as in Corollary 1 then gives that this collection of $\left(h, \mathcal{G}_{U|Z}\right) \in \mathcal{M}$ satisfies the selectionability criteria of Theorem 1, namely that $G_{U|Y_*Z}\left(\cdot|m, z\right)$ is selectionable with respect to the conditional distribution of random set $\mathcal{U}\left(Y, Z; h\right)$ given $(Y_* = m \wedge Z = z)$ induced by the distribution of Y conditional on $(Y_* = m \wedge Z = z)$ as given by F_{YZ}, a.e. $(m, z) \in \mathcal{R}_{Y_*Z}$. Thus $\mathcal{M}^*\left(F_{YZ}\right)$ is the identified set of structures $\left(h, \mathcal{G}_{U|Z}\right)$. $\square$