Set-Valued Control Functions*

Sukjin Han
School of Economics
University of Bristol
vincent.han@bristol.ac.uk

Hiroaki Kaido
Department of Economics
Boston University
hkaido@bu.edu

March 18, 2024

Abstract

The *control function* approach allows the researcher to identify various causal effects of interest. While powerful, it requires a strong invertibility assumption, which limits its applicability. This paper expands the scope of the nonparametric control function approach by allowing the control function to be *set-valued* and derive sharp bounds on structural parameters. The proposed generalization accommodates a wide range of selection processes involving discrete endogenous variables, random coefficients, treatment selections with interference, and dynamic treatment selections.

Keywords: Control Function, Control Variable, Partial Identification, Spillover Effects, Dynamic Treatment Effects.

*We gratefully acknowledge financial support from NSF grant SES-2018498.

1 Introduction

Endogeneity is the main challenge in conducting causal inference with observational data. The control function (CF) approach has been a valuable tool in addressing endogeneity and recovering various causal parameters. Although this approach originated in linear models, it has been proven to be a powerful tool for identification and estimation in nonparametric models that allow causal effect heterogeneity. The CF approach constructs control variables V, which define a *latent type* conditional on which endogenous explanatory variables D can be viewed as unconfounded. In observational settings, such V is typically constructed by inverting treatment selection processes so that it is written as a function of observables—thus a *control function*. Many empirical studies build on this insight to construct and utilize control variables (Kline and Walters, 2016; Card et al., 2019; Abdulkadiroğlu et al., 2020; Bishop et al., 2022). While powerful, this approach relies on the *invertibility* of selection models. For example, in nonparametric triangular models, invertibility requires D to be continuously distributed and the selection equation for D to be strictly monotone in a scalar unobservable variable. This type of restriction is viewed as the most important limitation of the CF approach (Blundell and Powell, 2003).

This paper expands the scope of the CF approach by dropping the invertibility assumption. We allow the control function to be set-valued. That is, one only needs to know the set of values V takes for each value of observable variables. This adaptation accommodates a wide range of selection processes. Observational data are often generated through complex selection processes, which may exhibit rich heterogeneity with continuous or discrete decisions, dynamically optimizing behavior, and strategic interaction of multiple agents. Such processes typically violate the invertibility assumption, as the mapping from observables to V is only a correspondence. We show that the CF approach can still be used with these selection processes to partially identify structural (i.e., causal) parameters, such as average and quantile structural functions for outcomes.

Formally, a *set-valued control function* $\boldsymbol{V}$ is a random closed set, constructed from observable variables, that contains the true control variable V inside it. This set can be used to construct a random set that collects all outcome values compatible with model structure and the *containment functional* or *Aumann expectation* associated with the set (Molchanov, 2017). The latter quantities generate identifying restrictions, which then yield the (sharp) identified set for structural parameters. The random set $\boldsymbol{V}$ can be constructed in a variety of settings. For example, D can be a binary variable generated by a generalized Roy model (Eisenhauer et al., 2015). One can also consider a selection model with binary D that violates the local average treatment effect (LATE) monotonicity (Imbens and Angrist, 1994) or, analogously, a model with continuous D with vector unobservables (e.g., selection with random coefficients). Other examples are the cases where D is determined through interaction of multiple agents (Tamer, 2003; Balat and Han, 2022) (e.g., due to violation of the stable-unit treatment value assumption (SUTVA) in forming outcomes); and where D and outcomes are dynamically determined over time (Han, 2021). We demonstrate how to derive identifying restrictions in these models.

This paper contributes to the vast literature on identification and estimation in nonparametric models with endogenous explanatory variables. In linear models, the two-stage least squared (TSLS) estimator can have two different interpretations: the instrumental variable (IV) approach and the CF approach (Blundell and Powell, 2003). The noparametric version of the IV approach is considered in Newey and Powell (2003); Hall and Horowitz (2005); Chernozhukov and Hansen (2005); Darolles et al. (2011); D'Haultfoeuille and Février (2015); Torgovitsky (2015); Vuong and Xu (2017); Chen and Christensen (2018). Typically, this approach assumes invertibility in the outcome equation and thus relies on a scalar unobservable, so that the IV assumption can be utilized. The CF approach is generalized to nonparametric models by Newey et al. (1999); Chesher (2003); Das et al. (2003); Blundell and Powell (2004); Imbens and Newey (2009); D'Haultfœuille et al. (2021); Newey and Stouli (2021), following the adaptation to nonlinear parametric models in Newey (1987); Rivers and Vuong (1988); Smith and Blundell (1986); Blundell and Smith (1989). The nonparametric CF literature typically assumes a model for endogenous explanatory variables and its invertibility in a scalar unobservable. Then, this approach generates control variables and combines it with the CF assumption (non-nested to the IV assumption) to identify structural parameters. Although this approach restricts selection behavior and is not applicable to discrete treatments, its advantage is the freedom from restricting heterogeneity directly relevant in generating causal effects. Another important strand of the causal inference literature concerns a binary or discrete treatment with a monotonicity assumption (Imbens and Angrist, 1994; Abadie et al., 2002) or equivalently (Vytlacil, 2002) a threshold-crossing model (Heckman and Vytlacil, 2005).

Chesher and Rosen (2017) generalize the IV approach to a framework where one can partially identify structural parameters in a range of complete and incomple models (see also Chesher and Smolinski (2012); Chesher and Rosen (2013)). Under the IV assumption, they define a random set of structural unobservables, which is used to construct sharp bounds on structural parameters. Instead, this paper proposes the CF approach to partial identification, filling the gap in the literature and complementing Chesher and Rosen (2017). We construct a random set for selection unobservables and let true control variables be a (measurable) selection of the random set. Sharing the aspect of the CF literature above, we allow for arbitrary causal effect heterogeneity (e.g., multi-dimensional outcome unobservables). Overcoming the aspect of the CF literature, we allow for discrete treatments and heterogeneity and complexity in treatment selection (e.g., multi-dimensional selection unobservables, incompleteness of selection models).

Chesher (2005) also considers partial identification without requiring invertibility in selection processes. He assumes that discrete endogenous variables are generated from ordered structure and focuses on local parameters. This paper in contrast focuses on global parameters, while encompassing a range of selection processes including ordered selection. Shaikh and Vytlacil (2011); Jun et al. (2011); Mourifié (2015); Mogstad et al. (2018); Machado et al. (2019); Han and Yang (2024) consider partial identification in nonparametric models without requiring invertibility in selection processes; they consider either a binary treatment generated from threshold-crossing models (equivalently, under the LATE monotonicity) or a discrete treatment with similar restrictions. These models are nested in the class of models we consider, but our distinct feature is the generality in selection processes and the use of the CF approach.

2 Setup

Let $Y \in \mathcal{Y} \subseteq \mathbb{R}^{d_Y}$ be the outcome of interest generated according to the following *outcome equation*:

$$Y = \mu(D, X, U), \tag{2.1}$$

where $D \in \mathcal{D} \subseteq \mathbb{R}^{d_D}$ is a vector of endogenous treatment variables, $X \in \mathcal{X} \subseteq \mathbb{R}^{d_X}$ is a vector of covariates, and $U \in \mathcal{U} \subseteq \mathbb{R}^{d_U}$ is a vector of latent variables. All random variables are defined on a complete probability space $(\Omega, \mathfrak{F}, P)$. The *structural function* μ determines the value of the *potential outcome* $Y_d = \mu(d, X, U)$ that would realize when the endogenous variable is set to $d \in \mathcal{D}$. Many policy-relevant parameters are features of the potential outcome, and hence functionals of μ. Examples are the *average structural function* and the *distributional structural function*: $\text{ASF}(d) \equiv E[\mu(d, X, U)] = E[Y_d]$ and $\text{DSF}(d) \equiv F_{\mu(d,X,U)} = F_{Y_d}$, respectively. Other examples are the *policy-relevant structural function* and the *mediated structural function*, defined later.

A vector of control variables $V \in \mathcal{V} \subseteq \mathbb{R}^{d_V}$ is such that, the assignment of D becomes independent of U, once we condition on V and the observable covariates X. That is,

$$D \perp U | X, V. \tag{2.2}$$

Such variables allow the researcher to identify various causal parameters without additional parametric assumptions on μ.

For this approach to work, one needs to express V as a function of observable variables. Let Z be the vector of IVs. A commonly used specification is the additive model $D = \Pi(Z) + V$, in which one may express $V = D - \Pi(Z)$ (e.g., Newey et al., 1999). Imbens and Newey (2009) consider a nonseparable system, in which a single endogenous variable is modeled as $D = h(Z, \tilde{V})$ of a vector of instrumental variables Z and a continuously distributed *scalar* latent variable $\tilde{V}$, where h is strictly monotonic in $\tilde{V}$. They show that, under the independence of $(U, \tilde{V})$ and Z, one may use the conditional cumulative distribution function $V = F_{D|Z}(D|Z)$ as a control variable. The key assumption is the invertibility of h in the latent variable, which ensures that there is a one-to-one relationship between $\tilde{V}$ and V.

The argument above requires the selection equation to be invertible in the control variable.[1] This invertibility requirement often restricts the form of the selection equation and the dimension of V. When V is continuously distributed, the invertibility also requires D to be continuous, which limits the scope of the control function assumption. Moreover, having vector V is important in allowing for heterogeneity in the selection mechanism. We, therefore, aim to remove these restrictions.

2.1 Motivating Examples

We introduce motivating examples below, starting from a single agent's self-selection model to more complex ones. They share a few key features. First, they involve a vector of control variables, conditional on which the treatment decision can be viewed as random. Second, they do not allow the researcher to uniquely recover the control variables. Nonetheless, one may construct a set that contains the true control variable V. We will formally define set-valued control functions in the next section. Finally, the above features are related to the fact that the control variable V may be interpreted as structural unobservables in these examples.

These features of selection mechanism can be summarized in the following generalized selection equation:

$$D = \pi(Z, X, V). \tag{2.3}$$

Note that $(D, X, Z) \mapsto V$ is in general a correspondence, for example, because either $\pi(Z, X, \cdot)$ is not strictly monotonic or V is not scalar. Therefore, the selection process (2.3) restricts V

[1] As discussed in the introduction, Chesher (2005) is an exception even though he employs the CF approach.

to the following set almost surely:

$$\boldsymbol{V}(D,Z,X;\pi)=\{v:D=\pi(Z,X,v)\}\subseteq\mathbb{R}^{d_V}. \tag{2.4}$$

Each example illustrates specific forms of (2.3) and (2.4).

We start with a generalized Roy model of selection (Eisenhauer et al., 2015).

Example 1 (Generalized Roy Model): Let D be a binary treatment that is determined by the selection equation

$$D=1\{\pi(Z,X)\geq V\}, \tag{2.5}$$

where we normalize $V|X$ to the uniform distribution without loss of generality. The selection equation can be motivated by the generalized Roy model. Suppose $Y=DY_1+(1-D)Y_0$ where Y_d follows

$$Y_d=\mu(d,X)+U_d \quad \text{for } d=0,1. \tag{2.6}$$

We allow the unobservables U_d to be treatment-specific, which ensures unobserved heterogeneity necessary to formulate the generalized Roy model. This makes $U=(U_1,U_0)$ a vector. Let $C=\mu_c(Z,X)+U_c$ be the cost of choosing one alternative over the other, where Z is a vector of variables that shifts the cost but not the outcome.[2] The treatment decision is based on the net surplus S from the treatment:

$$D=1\{S\geq 0\}=1\{Y_1-Y_0-C\geq 0\}. \tag{2.7}$$

We may write the surplus as

$$S=\pi(Z,X)-V, \tag{2.8}$$

where $\pi(Z,X)=\mu(1,X)-\mu(0,X)-\mu_c(Z,X)$ is the observable part of the surplus, and $V=(U_c-U_1+U_0)$ is the unobserved part of the surplus. Then, we can express the treatment decision as (2.5). Clearly, V depends on (U_0,U_1).

Suppose we are interested in the causal effect of D on Y. Suppose Z is independent of U given (X,V). Then, (X,V) are valid control variables because D's remaining variation is independent of U conditional on them. What prevents us from applying the existing approach is that we cannot recover V by inverting (2.5) because D is binary. Nonetheless, the model

[2] The generalized Roy model above nests the classical Roy model where C is degenerate (Heckman and Honoré, 1990) and the extended Roy model where U_c is degenerate (Heckman and Vytlacil, 2007).

restricts V to the following set almost surely:

$$\boldsymbol{V}(D, Z, X; \pi) = \begin{cases} [0, \pi(Z, X)] & \text{if } D = 1 \\ [\pi(Z, X), 1] & \text{if } D = 0, \end{cases} \tag{2.9}$$

which is a set-valued analog of the control function we may condition on. □

The previous example satisfies the local average treatment effect (LATE) monotonicity, eliminating either compliers or defiers (Imbens and Angrist, 1994; Vytlacil, 2002). Next, we consider a selection model that allows richer compliance types.

Example 2 (Non-Monotonic Treatment Decisions): Suppose the value of the instrument is set to z. Let the potential treatment be

$$D_z = 1\{\pi(z, X) \geq V_z\} \quad \text{for } z \in \mathcal{Z}. \tag{2.10}$$

The observed treatment is $D = \sum_{z \in \mathcal{Z}} D_z 1\{Z = z\}$. Suppose Z is binary below. Given (2.10), both compliers and defiers can have nonzero shares:

$$\begin{aligned} \{D_0 = 0, D_1 = 1\} &= \{V_0 > \pi(0, X), V_1 \leq \pi(1, X)\}, \\ \{D_0 = 1, D_1 = 0\} &= \{V_0 \leq \pi(0, X), V_1 > \pi(1, X)\}. \end{aligned}$$

The observed treatment D satisfies

$$\begin{aligned} D &= 1\{\pi(0, X) - V_0 + (\pi(1, X) - \pi(0, X) - V_1 + V_0)Z \geq 0\} \\ &= 1\{\pi(0, X) + Z(\pi(1, X) - \pi(0, X)) + ZV_1 + (1 - Z)V_0 \geq 0\} \\ &\equiv 1\{\tilde{\pi}(Z, X) + (V_1 - V_0)Z + V_0 \geq 0\}, \end{aligned} \tag{2.11}$$

where $\tilde{\pi}(Z, X) = \pi(0, X) + Z(\pi(1, X) - \pi(0, X))$.[3] One may view the last expression as a *random-coefficient* model, in which the individuals respond heterogeneously to interventions to Z (Gautier and Hoderlein, 2011; Kline and Walters, 2019). Suppose the outcome Y is generated according to (2.1) and Z is independent of U conditional on (X, V_0, V_1). Then, (X, V_0, V_1) are valid control variables. By (2.11), $V = (V_0, V_1)$ belongs to the following set

[3] Note that $D = ZD_1 + (1 - Z)D_0 = 1\{Z(\pi(1, X) - V_1) + (1 - Z)(\pi(0, X) - V_0) \geq 0\}$.

almost surely:

$$\boldsymbol{V}(D, Z, X; \pi) = \begin{cases} \{(v_0, v_1) : \tilde{\pi}(Z, X) + (1 - Z)v_0 + Zv_1 \geq 0\} & \text{if } D = 1 \\ \{(v_0, v_1) : \tilde{\pi}(Z, X) + (1 - Z)v_0 + Zv_1 \leq 0\} & \text{if } D = 0. \end{cases} \tag{2.12}$$

□

The next example involves multiple individuals. Importantly, other individuals' treatment status affects one's outcome through spillover or equilibrium effects.

Example 3 (Treatment Responses with Social Interactions): Consider individuals $j = 1, \ldots, J$. Let $D = (D_1, \ldots, D_J)$ be a vector of treatment decisions across individuals. We may be interested in the effect of the entire profile D on some outcome Y. The observed treatments D satisfy

$$D_j = 1\{\pi_j(D_{-j}, Z_j, X) \geq V_j\},\ j = 1, \ldots, J \tag{2.13}$$

where D_{-j} is defined as the vector D without the element D_j and $V_j|X = x$ is normalized to $U[0, 1]$ without loss of generality.[4] In this specification, each player's treatment decision is affected by others' decisions.

One way to motivate (2.13) is by relaxing the *Stable Unit Treatment Value Assumption (SUTVA)* (Rubin, 1978) (or equivalently relaxing the *Individualistic Treatment Response (ITR)* by Manski (2013)) and introducing Roy-type decisions. Let $Y_{j,d_1,\ldots,d_J}$ be the potential outcome of individual j when D is set to $(d_1, \ldots, d_J)$. The previous examples assume an individual's outcome only depended on their own treatment (i.e., SUTVA or ITR) that $Y_{j,d_1,\ldots,d_J} = Y_{j,d_j}$. We relax such an assumption and allow each individual's outcome to depend on the entire vector of treatments received by the individuals. The generalization is important when treatments are expected to have spillover effects (Graham, 2011; Aronow and Samii, 2017). For simplicity, consider two individuals (i.e., $J = 2$). Each individual may either choose $D_j = 0$ or $D_j = 1$. For individual j, the outcome is generated according to

$$Y_j = \sum_{(d_1,d_2)\in\{0,1\}^2} 1\{D_1 = d_1, D_2 = d_2\} Y_{j,d_1,d_2}, \tag{2.14}$$

where

$$Y_{j,d_1,d_2} = \mu_j(d_1, d_2, X) + U_{j,d_1,d_2}. \tag{2.15}$$

[4]The joint distribution of V is unrestricted.

Suppose the individuals are involved in Roy-type decisions:

$$D_1 = 1\{Y_{1,(1,D_2)} - Y_{1,(0,D_2)} \geq \mu_{c1}(Z_j, X) + U_{c1}\},$$
$$D_2 = 1\{Y_{2,(D_1,1)} - Y_{2,(D_1,0)} \geq \mu_{c2}(Z_j, X) + U_{c2}\}.$$

Then, the selection process is compatible with (2.13) with

$$\pi_j(D_{-j}, Z_j, X) = \mu_j(1, D_{-j}, X) - \mu_j(0, D_{-j}, X) - \mu_{cj}(Z_j, X)$$
$$V_j = U_{cj} - U_{j,1,D_{-j}} - U_{j,0,D_{-j}}.$$

The individuals' social interaction is captured by the impact of the other individual's treatment choice on player j's payoff, which corresponds to $\pi_j(1, z_j, x) - \pi(0, z_j, x)$.

Multiple solutions to the simultaneous system (2.13) may exist, which may make the selection process *incomplete* and thus set-valued (Tamer, 2003; Balat and Han, 2022). For example, suppose $\pi_j(1, z_j, x) - \pi(0, z_j, x) \leq 0$ for $j = 1, 2$. For each (z, x, v_1, v_2), the model's prediction is

$$G(v|z, x; \pi) = \begin{cases} \{(0,0)\} & (v_1, v_2) \in S_{\pi,(0,0)}(z, x) \\ \{(0,1)\} & (v_1, v_2) \in S_{\pi,(0,1)}(z, x) \\ \{(1,0)\} & (v_1, v_2) \in S_{\pi,(1,0)}(z, x) \\ \{(1,1)\} & (v_1, v_2) \in S_{\pi,(1,1)}(z, x) \\ \{(1,0),(0,1)\} & (v_1, v_2) \in S_{\pi,\{(1,0),(0,1)\}}(z, x). \end{cases} \tag{2.16}$$

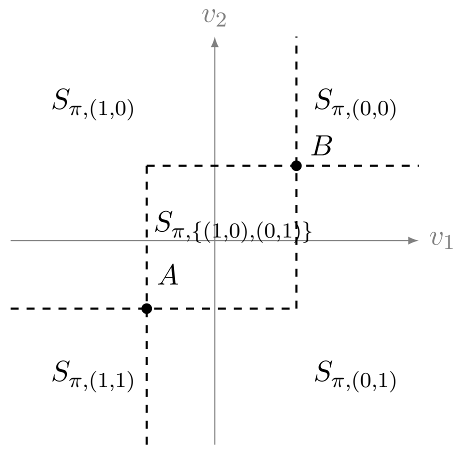
Figure 1 summarizes subsets $S_{\pi,(0,0)}(z, x), \dots, S_{\pi,\{(1,0),(0,1)\}}(z, x)$ of (v_1, v_2) values that correspond to certain model predictions.[5] Let $V_s : \Omega \to \{0, 1\}$ represent an unknown *selection mechanism* that selects a solution when G contains multiple values.[6] If the model predicts multiple equilibria $(V_1, V_2) \in S_{\pi,\{(1,0),(0,1)\}}(Z, X)$, the equilibrium outcome $D = (1, 0)$ is selected when $V_s = 1$, and $D = (0, 1)$ is selected $V_s = 0$. Our model is silent about how

[5] The sets are formally defined as follows.

$$\begin{aligned}
S_{\pi,(0,0)}(z,x) &= \{v : v_1 > \pi_1(0, z_1, x), v_2 > \pi_2(0, z_2, x)\} \\
S_{\pi,(0,1)}(z,x) &= \{v : \pi_1(1, z_1, x) < v_1 \leq \pi_1(0, z_1, x), v_2 \leq \pi_2(1, z, x)\} \cup \{v : \pi_1(0, z_1, x) < v_1, v_2 \leq \pi_2(0, z_2, x)\} \\
S_{\pi,(1,0)}(z,x) &= \{v : v_1 \leq \pi_1(1, z_1, x), v_2 > \pi_2(1, z_2, 0)\} \cup \{v : \pi_1(1, z_1, x) < v_1 \leq \pi_1(0, z_1, x), v_2 > \pi_2(0, z_2, x)\} \\
S_{\pi,(1,1)}(z,x) &= \{v : v_1 \leq \pi_1(1, z_1, x), v_2 \leq \pi_2(1, z_2, x)\} \\
S_{\pi,\{(0,1),(1,0)\}}(z,x) &= \{v : \pi_j(0, z_j, x) < v_j \leq \pi_j(1, z_j, x), j = 1, 2\}.
\end{aligned}$$

[6] Without loss of generality, one may represent the selection mechanism by a latent random variable defining a mixture. See Tamer (2010), Ponomareva and Tamer (2011), and Molinari (2020, p.377).

Figure 1: Level sets of $v \mapsto G(v|z;\pi)$ and set-valued CF



Note: $A = (\pi_1(1, z_1, x), \pi_2(1, z_1, x))$; $B = (\pi_1(0, z_1, x), \pi_2(0, z_2, x))$.

V_s is generated, and hence, we do not restrict the distribution of V_s. It can be correlated with (Z, X, V_1, V_2) and $U = (U_{j,d_1,d_2}, U_{c_j})_{d_1,d_2 \in \{0,1\}, j=1,2}$. Therefore, V_s is another source of possible endogeneity. Suppose Z is independent of U conditional on (X, V_1, V_2, V_s). Then, (X, V_1, V_2, V_s) are valid control variables.

Below, we take V_s as a component of the control variables and let $V = (V_1, V_2, V_s)$. Suppose $D = (1, 1)$ is realized. What does it tell us about V? This outcome occurs if and only if $(V_1, V_2) \in S_{\pi,(1,1)}(Z, X)$ regardless of V_s. Therefore, $V \in S_{\pi,(0,0)}(Z, X) \times \{0, 1\}$ in this case. Now suppose $D = (0, 1)$ realized. This outcome may arise from two scenarios. One is that $(0, 1)$ is the unique equilibirum due to $(V_1, V_2) \in S_{\pi,(0,1)}(Z, X)$, in which case V_s can take any value. The other scenario is that $(0, 1)$ is one of the predicted equilibria $(V_1, V_2) \in S_{\pi,\{(1,0),(0,1)\}}(Z, X)$, and it was selected due to $V_s = 0$. Using this argument, we may define the following set-valued control function:

$$\boldsymbol{V}(D, Z, X; \pi) = \begin{cases} S_{\pi,(0,0)}(Z, X) \times \{0, 1\} & \text{if } D = (0, 0) \\ \left[S_{\pi,(0,1)}(X, Z) \times \{0, 1\}\right] \cup \left[S_{\pi,\{(1,0),(0,1)\}}(Z, X) \times \{0\}\right] & \text{if } D = (0, 1) \\ \left[S_{\pi,(1,0)}(Z, X) \times \{0, 1\}\right] \cup \left[S_{\pi,\{(1,0),(0,1)\}}(Z, X) \times \{1\}\right] & \text{if } D = (1, 0) \\ S_{\pi,(1,1)}(Z, X) \times \{0, 1\} & \text{if } D = (1, 1). \end{cases} \tag{2.17}$$

□

Remark 1: In the example above, we focused on the case in which the treatment is generated through a game of strategic substitution: $\pi_j(1, z_j, x) - \pi(0, z_j, x) \leq 0, j = 1, 2$. The same argument can be applied to games of strategic complementarity and even to models with incoherent predictions.

The next example considers dynamic treatment decisions with imperfect compliance (Robins, 1997; Han, 2021).

EXAMPLE 4 (Dynamic Treatment Effects): In the initial period, the observed treatment and outcome (D_1, Y_1) are generated according to

$$D_1 = 1\{\pi_1(Z_1, X) \geq V_1\}. \tag{2.18}$$

$$Y_1 = 1\{\mu_1(D_1, X) \geq U_1\} \tag{2.19}$$

The observed treatment status in the next period is determined based on (D_1, Y_1) as

$$D_2 = 1\{\pi_2(Y_1, D_1, Z_2, X) \geq V_2\}. \tag{2.20}$$

Finally, the eventual observed outcome Y_2 is determined by

$$Y_2 = 1\{\mu_2(Y_1, D_1, D_2, X) \geq U_2\}. \tag{2.21}$$

Throughout, U_t and V_t are normalized to $U[0, 1]$ conditional on $X = x$.

A researcher may consider various causal effects. For now, consider the effect of the initial outcome and treatment history $D = (Y_1, D_1, D_2)$ on Y_2. Recovering the effect is not straightforward because U_2 may depend on the unobserved determinants (U_1, V_1, V_2) of the treatment. For example, U_1 and U_2 may share a time invariant component. Another possibility is that U_2 may be related to (V_1, V_2) through the agent's dynamic treatment decisions.

Below, we let $U \equiv U_2$ and let $V \equiv (U_1, V_1, V_2)$ be unobserved control variables; also let $Z \equiv (Z_1, Z_2)$. Inspecting the system of selection equations (2.18)-(2.20), one can see that the assignment of $D = (Y_1, D_1, D_2)$ is independent of U_2 conditional on (X, V) as long as the instrumental variables Z are independent of U_2.

Let $\pi \equiv (\mu_1(\cdot), \pi_1(\cdot), \pi_2(\cdot))$. One can construct the following set-valued control function:

$$\boldsymbol{V}(D, Z, X; \pi) = \boldsymbol{V}_{U_1}(D, X; \mu_1) \times \boldsymbol{V}_1(D, Z_1, X; \pi_1) \times \boldsymbol{V}_2(D, Z_2, X; \pi_2), \tag{2.22}$$

where

$$\boldsymbol{V}_{U_1}(D,X;\mu_1)=\begin{cases}[\mu_1(D_1,X),1] & \text{if } Y_1=0\\ [0,\mu_1(D_1,X)] & \text{if } Y_1=1,\end{cases}\quad \boldsymbol{V}_1(D,Z_1,X;\pi_1)=\begin{cases}[\pi_1(Z_1,X),1] & \text{if } D_1=0\\ [0,\pi_1(Z_1,X)] & \text{if } D_1=1,\end{cases}$$

$$\boldsymbol{V}_2(D,Z_2,X;\pi_2)=\begin{cases}[\pi_2(Y_1,D_1,Z_2,X),1] & \text{if } D_2=0\\ [0,\pi_2(Y_2,D_2,Z_2,X)] & \text{if } D_2=1.\end{cases}$$

□

3 Model Predictions

We derive the model's prediction based on our incomplete knowledge that V is a valid control variable but only known to belong to $\boldsymbol{V}$. We first assume that V and observable covariates X form a set of control variables.

Assumption 1: *$U|D,X,V \sim U|X,V$.*

By Assumption 1, the treatment decision is independent of U once we condition on the control variables (X,V). Next, we introduce *random closed sets* and their *measurable selections.*

Definition 1 (Random Closed Set): *A map $\boldsymbol{X}$ from a probability space $(\Omega,\mathfrak{F},P)$ to the family $\mathcal{F}(\mathbb{E})$ of closed subsets of a Euclidean space $\mathbb{E}$ is called a* random closed set *if*

$$\boldsymbol{X}^-(K)\equiv\{\omega\in\Omega:\boldsymbol{X}(\omega)\cap K\neq\emptyset\} \tag{3.1}$$

is in $\mathfrak{F}$ for each compact set $K\subseteq\mathbb{E}$.

Definition 2 (Measurable Selections): *For any random set $\boldsymbol{X}$, a measurable selection of $\boldsymbol{X}$ is a random element X with values in $\mathbb{E}$ such that $X(\omega)\in\boldsymbol{X}(\omega)$ almost surely. We denote by $\mathrm{Sel}(\boldsymbol{X})$ the set of all selections from $\boldsymbol{X}$.*

We assume one can construct a set-valued control function.

Assumption 2: *(i) There is a random closed set $\boldsymbol{V}:\Omega\to\mathcal{F}(\mathbb{R}^{d_V})$ such that $V\in\boldsymbol{V}$ with probability 1; (ii) $\boldsymbol{V}$ is a measurable function of the observable variables and an infinite-dimensional parameter π.*

The set-valued control function $\boldsymbol{V}$ is a random closed-set constructed from the observables.[7] A leading case would be $\boldsymbol{V}$ generated by a selection equation $D = \pi(Z, X, V)$ (e.g., Section 2.1) where Z is a vector of instrumental variables excluded from μ. However, $\boldsymbol{V}$ can also be generated from other sources (e.g., Auerbach, 2022). Assumption 2 is agnostic about the genesis of a set-valued control function. We write $\boldsymbol{V}(D, X, Z; \pi)$ whenever it is useful to show its dependence on (D, X, Z) and π.

Let us discuss Assumptions 1-2 further. In the conventional CF approach, we use the control variable for two purposes. The first is to account for the effects of confounders, which we refer to as "controlling" (Assumption 1), and we continue to use V for this purpose. The second purpose is to condition on the subpopulation for which the conditional independence assumption holds. If V is observable or can be uniquely recovered from other observables, we may use these properties simultaneously. However, in the current setting, this is not the case. Therefore, while we continue to use V to control for the latent confounders, we use $\boldsymbol{V}$ (recovered from other observables ensured by Assumption 2) to condition on a "coarser" subpopulation. Being unable to condition on V can lead to a loss of point identification. Nonetheless, our framework allows the researcher to derive sharp bounds on the parameters of interest.

Finally, we assume U is continuously distributed.

Assumption 3: *$U|D, X, V$ has a strictly positive density with respect to Lebesgue measure on $\mathbb{R}^{d_U}$ almost surely.*

By Assumption 3, one may represent the latent variables in the outcome equation as $U = Q(\eta; D, X, V)$ for some measurable function $Q : [0,1]^{d_U} \times \mathcal{D} \times \mathcal{X} \times \mathcal{V} \to \mathcal{U} \subseteq \mathbb{R}^{d_U}$ and a random vector $\eta \in \mathbb{R}^{d_U}$, which is independent of (D, X, V) and is uniformly distributed over $[0,1]^{d_U}$. This representation holds generally. For example, suppose $U = (U_0, U_1)$ is two dimensional as in Example 1. One can apply the Knothe-Rosenblatt transform (see, e.g., Villani, 2008; Carlier et al., 2010; Joe, 2014) to represent $(U_0, U_1) \sim F_{U|D,X,V}$ sequentially

$$U_0 = Q_0(\eta; D, X, V) \equiv F^{-1}_{U_0|D,X,V}(\eta_0|D, X, V), \tag{3.2}$$

$$U_1 = Q_1(\eta; D, X, V) \equiv F^{-1}_{U_1|U_0,D,X,V}(\eta_1|U_0, D, X, V), \tag{3.3}$$

where $(\eta_0, \eta_1) \sim U[0,1]^2$. Then, by Assumption 1, we may write

$$U_0 = Q_0(\eta; X, V) \equiv F^{-1}_{U_0|X,V}(\eta_0|X, V),$$
$$U_1 = Q_1(\eta; X, V) \equiv F^{-1}_{U_1|U_0,X,V}(\eta_1|U_0, X, V).$$

[7] A singleton-valued control function in the literature is a special case of Assumption 2.

In general, under Assumption 1, we represent U as

$$U = Q(\eta; X, V), \tag{3.4}$$

for a map $Q : [0,1]^{d_U} \times \mathcal{X} \times \mathcal{V} \to \mathcal{U} \subseteq \mathbb{R}^{d_U}$ that depends on the conditional distribution F of $U|X, V$. We may view η as the remaining source of randomness in the potential outcome after controlling for (X, V). Note that η_0 and η_1 are mutually independent even though U_0 and U_1 are not.

Now consider the model's prediction given the structural parameter $\theta \equiv (\mu, F, \pi)$. For the moment, suppose we condition on (X, V). We represent U as $U = Q(\eta; X, V)$. Under Assumption 1, the treatment D is generated independently of U conditional on (X, V). The outcome is determined by

$$Y = \mu(D, X, U) = \mu(D, X, Q(\eta; X, V)). \tag{3.5}$$

One can view the right-hand side of (3.5) as an outcome equation augmented by an adjustment term $Q(\eta; X, V)$, which involves the control variable V and a "clean" error term η that is independent of D.[8] It is useful to note that Q is a function of F.

Using the fact that V is a measurable selection of $\boldsymbol{V}$, we define the following random closed set

$$\boldsymbol{Y}(\eta, D, X, \boldsymbol{V}; \mu, F) \equiv \mathrm{cl}\big\{y \in \mathcal{Y} : y = \mu(D, X, Q(\eta; X, V)), V \in \mathrm{Sel}(\boldsymbol{V})\big\}. \tag{3.6}$$

This set collects all outcome values (and their closure) compatible with the model structure. Lemma 1 in the appendix establishes $\boldsymbol{Y}$ is a well-defined random closed set. Representing the model's prediction in this way has several advantages. First, $\boldsymbol{Y}$ collects all outcome values given all *observable exogenous variables* $(D, X, \boldsymbol{V})$ and latent variables η. It represents the prediction of an *incomplete model* and fits into the framework of Jovanovic (1989) (see Remark 2 below). The model is silent about how Y gets selected from $\boldsymbol{Y}$ because we do not observe the true control variable V. Nonetheless, there are systematic ways to obtain sharp identifying restrictions in such a model and eventually sharp bounds on parameters of interest. Second, $\boldsymbol{Y}$ can often be simplified, which also helps derive closed-form bounds; e.g., see the discussion of the next paragraph. Finally, the framework can accommodate both continuous and discrete outcomes. We provide further details in Section 5.

Assumption 1 plays an important role in obtaining identifying restrictions for structural

[8] This is analogous to an additive model, in which the error term can be decomposed into a control function and an error term that is independent of the treatment.

parameters via $\boldsymbol{Y}$. Each measurable selection of $\boldsymbol{Y}$ is represented by the *augmented outcome equation* $Y = \mu(D, X, Q(\eta; X, V))$, which involves two functions with different features: the structural function μ takes D as its argument, while the adjustment term Q excludes D but accounts for (X, V). This separation is possible due to Assumption 1. Since Q does not depend on D, it facilitates recovering structural parameters from the equation $\mu(d, x, Q(\eta, x, v))$. For example, we may represent the potential outcome using the augmented outcome equation: $Y_d = \mu(d, X, Q(\eta; X, V))$. This allows us to express structural quantities such as the *average conditional response* $E[Y_d|X = x, V = v]$ by integrating out η

$$E[Y_d|X = x, V = v] = \int_{[0,1]^{d_U}} \mu(d, x, Q(\eta; x, v))d\eta. \tag{3.7}$$

After characterizing the sharp identification region for θ, we use this property to obtain bounds on various structural functions of interest (see Section 4.2).

Remark 2: The general formulation of Jovanovic (1989) is characterized by observed endogenous variables y, latent variables η, and a *structure* (ν, ϕ), where ν is the distribution of η, and ϕ is a relation such that $(y, \eta) \in \phi$. The observable exogenous variables are allowed to shift ϕ. In our setting, the graph of $\boldsymbol{Y}$ corresponds to ϕ. The representation of U in (3.4) allows us to incorporate the structural parameter (μ, F) into the model's incomplete prediction (ϕ in Jovanovic (1989)), whereas the remaining randomness is captured by $\eta \sim U[0, 1]^{d_U}$.

4 Identification

Let P_0 be the joint distribution of the observable variables (Y, D, X, Z). Recall $\theta = (\mu, F, \pi)$, and it belongs to a parameter space $\Theta \equiv \mathsf{M} \times \mathsf{F} \times \Pi$, which embodies a priori restrictions on the structural parameter. It is common to have additional restrictions on the parameter in the selection equation. We let $\Pi_r(P_0) \subset \Pi$ be the set of selection parameters that satisfy the additional restrictions. As we show below, π can be point identified in some examples.

We define the sharp identification region for θ as follows.

Definition 3 (Sharp Identification Region under Full Independence): *The* sharp identification region $\Theta_I(P_0) \subset \mathsf{M} \times \mathsf{F} \times \Pi_r(P_0)$ *is a set such that each of its elements* $\theta = (\mu, F, \pi)$ *satisfies the following statement: (i) For any* $Y|D, X, Z \sim P_0(Y|D, X, Z)$*, one can represent the outcome as* $Y = \mu(D, X, U)$*, where* U*'s conditional law* F *satisfies Assumptions 1* and *3 for some* $V : \Omega \to \mathcal{V}$*. (ii) The control variable* V *is a measurable selection of a set-valued control function* $\boldsymbol{V}$ *satisfying Assumption 2.*

The main result (Theorem 1) of this section characterizes $\Theta_I(P_0)$ through inequality restrictions on θ. For this, we introduce the *containment functional* $\mathbb{C}_\theta$ of a random set. Note that the (conditional) distribution of η satisfies $F_\eta(\eta|D = d, X = x, Z = z) = F_\eta(\eta|D = d, X = x, V = v) = \eta$ by construction. Therefore, for any closed set $A \subset \mathcal{Y}$ and $(d, x, z) \in \mathcal{D} \times \mathcal{X} \times \mathcal{Z}$, we let

$$\mathbb{C}_\theta(A|D = d, X = x, Z = z) \equiv \int_{[0,1]^{d_U}} 1\big\{\boldsymbol{Y}(\eta, D, X, \boldsymbol{V}; \mu, F) \subseteq A\big\} d\eta \tag{4.1}$$

be the containment functional associated with $\boldsymbol{Y}$. This functional uniquely determines the distribution of $\boldsymbol{Y}$ (Molchanov, 2017). We note that, conditional on (D, X, Z), the remaining source of randomness in $\boldsymbol{Y}$ is η, which is a d_U-dimensional random vector distributed uniformly over $[0, 1]^{d_U}$ independently of (D, X, Z). As such, it is straightforward to compute the right-hand side of (4.1) analytically or by simulation (see Section 5). The following theorem characterizes the sharp identification region.

THEOREM 1: *Suppose Assumptions 1-3 hold. Then, the sharp identification region for the structural parameter $\theta = (\mu, F, \pi)$ is*

$$\Theta_I(P_0) = \{\theta \in \Theta : P_0(A|D, X, Z) \geq \mathbb{C}_\theta(A|D, X, Z),\ a.s.\ \forall A \in \mathcal{F}(\mathcal{Y}),\ \pi \in \Pi_r(P_0)\}. \tag{4.2}$$

The restrictions (4.2) are known as *Artstein's inequalities* (Molchanov and Molinari, 2018, Theorem 2.13). We use them to convert the model's set-valued prediction into a system of inequality restrictions that do not involve the unobserved control variable V, making the resulting restrictions amenable to estimation. More specifically, the left-hand side $P_0(A|D, X, Z)$ of the inequality can be recovered from a large sample of (Y, D, X, Z). We demonstrate, through examples, that one may compute the right-hand side $\mathbb{C}_\theta(A|D, X, Z)$ from model primitives (see Sections 5.2 and 5.4).

REMARK 3: For a given (d, x, z), the number of the inequalities in (4.2) is finite when Y is a discrete variable with a finite support. Furthermore, it often suffices to impose a subset of inequalities to characterize $\Theta_I(P_0)$. Such a class $\mathcal{A} \subseteq \mathcal{F}(\mathcal{Y})$ is called the *core determining class*. The smallest core determining class only depends on the graph representation of $\boldsymbol{Y}(\cdot, D, X, Z; \mu, F)$ and does not depend on P_0 (Luo and Wang, 2017; Ponomarev, 2022).[9] If Y is a continuous variable, (4.2) involves infinitely many inequalities. However, if the outcome equation is separable between (D, X) and U, one can work with a finite number of inequalities under a weaker conditional mean independence assumption; see Section 4.1.

[9] Ponomarev (2022) provides an algorithm to determine the smallest core determining class.

Practitioners can use (4.2) to make inference for $\Theta_I(P_0)$ or its elements. For example, one may use inference methods for conditional moment inequalities (Andrews and Shi, 2013; Chernozhukov et al., 2013) or likelihood-based inference methods (Chen et al., 2018; Kaido and Molinari, 2022).

4.1 Conditional Mean Restrictions

In this section, we focus on a scalar outcome Y that is continuously distributed. Consider the following additive model:

$$Y = \mu(D, X) + U. \tag{4.3}$$

It nests the linear model $\mu(d, x) = \alpha d + x'\beta$ as a special case. The model allows us to write $U = Y - \mu(D, X)$.[10] One can generate identifying restrictions on μ under conditional mean independence assumption on U (Newey et al., 1999; Pinkse, 2000).

Below, we work with a more general model by relaxing the rank similarity assumption maintained in (4.3), but we keep the separability assumption. Let $U \equiv (U_d, d \in \mathcal{D})$, and suppose

$$Y = \mu(D, X) + U_D, \tag{4.4}$$

where $U_D = \sum_{d \in \mathcal{D}} U_d 1\{D = d\}$. This way, Y is a function of vector U, making this model a special case of (2.1). Suppose the following assumption holds.

Assumption 4: *For each $d \in \mathcal{D}$, $E[|U_d|] < \infty$, and $E[U_d|D, X, V] = E[U_d|X, V]$, a.s.*

For each $d \in \mathcal{D}$, let $\lambda_d(X, V) \equiv E[U_d|X, V]$ and $\eta_d \equiv U_d - E[U_d|X, V]$. Under Assumption 4, we may write

$$E[Y|D = d, X = x, V = v] = \mu(d, x) + \lambda_d(x, v). \tag{4.5}$$

We note that λ_d is a function of F. We define the sharp identification region as follows.

Definition 4 (Sharp Identification Region under Mean Independence): *The* sharp identification region *$\Theta_I(P_0) \subset \mathsf{M} \times \mathsf{F} \times \Pi_r(P_0)$ is a set such that each of its elements $\theta = (\mu, F, \pi) \in \Theta_I(P_0)$ satisfies the following statement: (i) For any Y whose conditional mean is $E_{P_0}[Y|D, X, Z]$, one can represent the outcome as in (4.4), where U's conditional law F sat-*

[10] A similar argument can be applied to a nonadditive model $Y = \mu(D, X, U)$ for which U is a scalar and μ is invertible with respect to U. We focus on the additive model only for notational simplicity later.

isfies Assumptions 3 and 4 for some $V : \Omega \to \mathcal{V}$. (ii) The control variable V is a measurable selection of a set-valued control function $\boldsymbol{V}$ satisfying Assumption 2.

Let $\eta \equiv (\eta_d, d \in \mathcal{D})$. Define

$$\boldsymbol{Y}(\eta, D, X, Z; \mu, F) \equiv \mathrm{cl}\big\{y \in \mathcal{Y} : y = \mu(D, X) + \lambda_D(X, V) + \eta_D, V \in \mathrm{Sel}(\boldsymbol{V})\big\}. \tag{4.6}$$

Let $E_{P_0}[Y|D, X, Z]$ be the conditional mean of the outcome recovered from data. Then, it should be the conditional mean of a measurable selection of $\boldsymbol{Y}(\eta, D, X, Z; \mu, F)$ for some $\theta \in \Theta_I(P_0)$. The following theorem characterizes the sharp identified set.

Theorem 2: *Suppose Assumptions 2-4 hold. Suppose $E_{P_0}[|Y|] < \infty$. Then, the sharp identification region for the structural parameter is*

$$\begin{aligned}\Theta_I(P_0) = \{\theta \in \Theta : \mu(d, x) + \lambda_L(d, x, z) \le E_{P_0}[Y|D = d, X = x, Z = z] \\ \le \mu(d, x) + \lambda_U(d, x, z),\ \pi \in \Pi_r(P_0)\},\end{aligned} \tag{4.7}$$

where

$$\lambda_L(d, x, z) = \inf_{v \in \boldsymbol{V}(d,x,z;\pi)} \lambda_d(x, v), \quad \lambda_U(d, x, z) = \sup_{v \in \boldsymbol{V}(d,x,z;\pi)} \lambda_d(x, v). \tag{4.8}$$

4.2 Functionals of θ

Based on the identification region of θ obtained in Theorems 1 and 2, one can also construct bounds on functionals of θ. Let $W \equiv (X, V)$ and let F_W be its distribution. Structural estimands can be obtained from θ. For example, the average structural function $\mathrm{ASF}(d) \equiv E[\mu(d, X, U)] = E[Y_d]$ considered by Blundell and Powell (2003) can be expressed as a function of (μ, F, F_W):

$$\mathrm{ASF}(d) = \int\int \mu(d, x, u) dF(u|w) dF_W(w) = \int\int \mu(d, x, Q(\eta; w)) d\eta dF_W(w). \tag{4.9}$$

The average treatment effect (ATE) is then $\mathrm{ATE}(d, d') = \mathrm{ASF}(d) - \mathrm{ASF}(d')$. Another example is the distributional structural function (Chernozhukov et al., 2020), which is the CDF of the potential outcome Y_d is also a function of (μ, F, F_W):

$$\begin{aligned}\mathrm{DSF}(y, d) \equiv \int\int 1\{\mu(d, x, u) \le y\} dF_U(u|w) dF_W(w) \\ = \int\int 1\{\mu(d, x, Q(\eta; w)) \le y\} d\eta dF_W(w).\end{aligned} \tag{4.10}$$

The *quantile structural function* (QSF), the τ-th quantile of Y_d, can be obtained using $\text{QSF}(d) = \text{DSF}^{-1}(\tau, d)$ (Imbens and Newey, 2002).

Given a function $\varphi : \mathbb{R} \to \mathbb{R}$, let

$$\kappa(d) \equiv E[\varphi(Y_d)] = \int \varphi(\mu(d, x, Q(\eta; w))d\eta dF_W(w). \tag{4.11}$$

The average and distributional structural functions are special cases of κ with $\varphi(Y_d) = Y_d$ and $\varphi(Y_d) = 1\{Y_d \leq y\}$ respectively. In general, F_W is only partially identified. The following proposition characterizes the identification region for κ.

THEOREM 3: *Suppose the conditions of Theorem 1 or 2 hold. Suppose φ is bounded, and the underlying probability space is non-atomic. Then, the sharp identification region for κ is*

$$\mathfrak{K}_I(d) = \bigcup_{\theta \in \Theta_I(P_0)} [\underline{\kappa}(d; \theta), \overline{\kappa}(d; \theta)], \tag{4.12}$$

where

$$\overline{\kappa}(d; \theta) = E[\sup_{v \in \boldsymbol{V}(D,X,Z;\pi)} \int \varphi(\mu(d, X, Q(\eta; X, v))d\eta], \tag{4.13}$$

$$\underline{\kappa}(d; \theta) = E[\inf_{v \in \boldsymbol{V}(D,X,Z;\pi)} \int \varphi(\mu(d, X, Q(\eta; X, v))d\eta], \tag{4.14}$$

and the expectation above is with respect to the distribution of (η, D, X, Z).

The identification region for κ is expressed as a union of intervals. Practically, one may be interested in the upper and lower bounds of κ. They are given by

$$\overline{\kappa}(d) = \sup_{\theta \in \Theta_I(P)} E[\sup_{v \in \boldsymbol{V}(D,X,Z;\pi)} \int \varphi(\mu(d, X, Q(\eta; X, v))d\eta],$$
$$\underline{\kappa}(d) = \inf_{\theta \in \Theta_I(P)} E[\inf_{v \in \boldsymbol{V}(D,X,Z;\pi)} \int \varphi(\mu(d, X, Q(\eta; X, v))d\eta].$$

In some examples, F_W is point identified even if V itself is unobserved and is not uniquely recovered.[11] If so, for each $\theta \in \Theta_I(P_0)$, $\overline{\kappa}(d; \theta) = \underline{\kappa}(d; \theta)$. Then, we can simplify the bounds

[11]In Example 1, the distribution of V is normalized to $U[0, 1]$.

on $\kappa(d)$ as follows:

$$\overline{\kappa}(d) = \sup_{\theta \in \Theta_I(P)} E_{\eta,W}[\varphi(\mu(d, X, Q(\eta; W))] \tag{4.15}$$

$$\underline{\kappa}(d) = \inf_{\theta \in \Theta_I(P)} E_{\eta,W}[\varphi(\mu(d, X, Q(\eta; W))]. \tag{4.16}$$

In addition to the structural parameter (4.11), one can consider a policy that only changes the selection behavior. Suppose a policy sets Z (e.g., a tuition subsidy) to z, and the treatment selection under this policy is $D_z = \pi(z, X, V)$. The *policy-relevant structural function (PRSF)* would be

$$\kappa(z) \equiv E[\varphi(Y_{D_z})] = \int \varphi(\mu(\pi(z, w), x, Q(\eta; w))d\eta dF_W(w). \tag{4.17}$$

The PRSF is related to the policy-relevant treatment effect (PRTE) and marginal PRTE introduced in Heckman and Vytlacil (2005) and Carneiro et al. (2010).

One can consider another related structural function. Suppose $D = (D_1, D_2)$, and let Y_{d_1,d_2} denote the counterfactual outcome given (d_1, d_2) and D_{2,d_1} denote the counterfactual treatment of D_1 given d_1. Then the *mediated structural function (MSF)* would be

$$\kappa(d_1, d_1') \equiv E[\varphi(Y_{d_1,D_{2,d_1'}})] = \int \varphi(\mu(d_1, \pi_2(d_1', z, w), x, Q(\eta; w))d\eta dF_{Z,W}(z, w), \tag{4.18}$$

where we allow $d_2 \neq d_2'$. The MSF can be used to define the direct causal effect of one treatment and the indirect causal effect mediated by another treatment. This scenario is relevant in Example 3 on strategic interaction (e.g., a player's decision being mediated by the opponent's decision) and Example 4 on dynamic treatment effects (e.g., a previous treatment being mediated by the previous outcome; Han and Lee, 2023).

Again, one can derive bounds on these objects in a similar manner as before.

5 Applications of the Identification Results

We illustrate the use of Theorems 1 and 2 through examples.

5.1 Generalized Roy Model with a Continuous Outcome

We revisit Example 1. Let $U \equiv (U_1, U_0)$, and recall that

$$D = 1\{\pi(Z, X) \geq V\};$$

hence U's conditional mean independence from D holds as long as U is mean independent of the instrument Z. Let $\lambda_d(X,V) \equiv E[U_d|X,V]$ for $d \in \mathcal{D}$. The model's prediction is

$$\boldsymbol{Y}(\eta, D, X, \boldsymbol{V}; \mu, F) = \{y \in \mathcal{Y} : y = \mu(D,X) + \lambda_D(X,V) + \eta_D, V \in \mathrm{Sel}(\boldsymbol{V})\}. \tag{5.1}$$

Theorem 2 implies

$$E_{P_0}[Y|D=d, X=x, Z=z] \leq \mu(d,x) + \lambda_U(d,x,z) \tag{5.2}$$

$$E_{P_0}[Y|D=d, X=x, Z=z] \geq \mu(d,x) + \lambda_L(d,x,z). \tag{5.3}$$

We summarize the argument as a corollary.

Corollary 1: *Suppose $E_{P_0}[|Y|] < \infty$. Suppose $U_0, U_1|X,Z$ have a density with respect to Lebesgue measure, and $E[U_d|Z,X,V] = E[U_d|X,V], d = 0,1$. Then, $\Theta_I(P_0)$ is the set of parameter values $\theta = (\mu, F, \pi)$ such that, for almost all (d,x,z),*

$$\begin{aligned}\sup_{z\in\mathcal{Z}} \Big\{ E_{P_0}[Y|D=d, X=x, Z=z] - \lambda_U(d,x,z) \Big\} \\ \leq \mu(d,x) \leq \\ \inf_{z\in\mathcal{Z}} \Big\{ E_{P_0}[Y|D=d, X=x, Z=z] - \lambda_L(d,x,z) \Big\},\end{aligned} \tag{5.4}$$

where

$$\lambda_L(d,x,z) = \inf_{v \in \boldsymbol{V}(d,x,z;\pi)} \lambda_d(x,v), \quad \lambda_U(d,x,z) = \sup_{v \in \boldsymbol{V}(d,x,z;\pi)} \lambda_d(x,v). \tag{5.5}$$

The identifying restrictions (5.4) take the form of intersection bounds on μ. For each z, $E_{P_0}[Y|D=d, X=x, Z=z] - \lambda_U(d,x,z)$ defines a lower bound on $\mu(d,x)$. Since z is excluded from μ, we can intersect the lower bounds across all values of z. The upper bound is formed similarly. It is worth noting that (5.4) restricts the parameter vector $\theta = (\mu, F, \pi)$ jointly because λ_L, λ_U are functions of (F, π). Therefore, they are also useful for bounding (F, π). Furthermore, if $Z \perp V|X$, π is point identified as the propensity score $\pi(z,x) = P_0(D=1|Z=z, X=x)$ only using the model of selection. Hence, in this case, (5.4) gives joint restrictions on (μ, F).

The terms λ_U, λ_L can be seen as the *correction terms* to account for the effects of V. To see this, suppose $\boldsymbol{V}$ is a singleton $\{V(D,X,Z;\pi)\}$ (e.g., because $D = \pi(Z,X) + V$). Then,

$$\lambda_L(d,x,z) = \lambda_U(d,x,z) = \lambda_d(x,z) = E[U_d|X=x, V=v], \tag{5.6}$$

In this case, (5.2)-(5.3) reduce to

$$E[Y|D=d,X=x,Z=z]=\mu(d,x)+E[U_d|X=x,V=v]. \tag{5.7}$$

Hence, it justifies regressing Y on (D,X) with an additive correction term (Newey et al., 1999). This argument works only when $\boldsymbol{V}$ is singleton-valued. In the general setting with a set-valued control function, one can work with the intersection bounds (5.4).

5.2 Multinomial Choice with a Generalized Selection Model

Suppose an individual chooses an option Y out of mutually exclusive alternatives $\mathcal{Y}=\{1,\dots,J\}$ by maximizing her utility:[12]

$$Y\in\arg\max_{j\in\mathcal{Y}}\mu_j(D,X)+U_j. \tag{5.8}$$

The individual's utility from alternative j depends on whether she is enrolled in a certain program $(D=1)$ or not $(D=0)$.[13] For example, Sosa-Rubí et al. (2009) analyze the choice of pregnant women in Mexico who choose sites for their obstetric care. The treatment of interest is enrollment in a public health insurance program that provides access to health services for vulnerable populations.

Suppose there is a binary instrument (e.g., eligibility), and $D=D_1Z+D_0(1-Z)$ with

$$D_z=1\{\pi(z,X)\geq V_z\}. \tag{5.9}$$

This is the flexible selection model in Example 2. Allowing for the flexibility is relevant in this context, as the insurance program may not be mandatory for the eligible or exclusive against the non-eligible. The set-valued control function is as in (2.12).

Let $V=(V_0,V_1)$ and let $U_k=Q_k(\eta;X,V),k=1,\dots,J$. This model's prediction is

$$\begin{aligned}&\boldsymbol{Y}(\eta,D,X,\boldsymbol{V};\mu,F)\\&=\Big\{j\in\mathcal{Y}:\mu_j(D,X)\geq\inf_{V\in\mathrm{Sel}(\boldsymbol{V})}\Big(\max_{k\neq j}[\mu_k(D,X)+Q_k(\eta;X,V)]-Q_j(\eta;X,V)\Big)\Big\}.\end{aligned} \tag{5.10}$$

Each element of $\boldsymbol{Y}$ is the maximizer of the utility index $\mu_k(D,X)+Q_k(\eta;X,V),k=1,\dots,J$

[12] Similar to Section 5.1, one can allow further heterogeneity by replacing U_j with $U_{j,D}$ in this model.

[13] It is also possible to let μ be a function of individual-specific unobservables (e.g., random coefficients) and treat them as part of U. For simplicity, we do not pursue this extension here.

for some $V \in \mathrm{Sel}(\boldsymbol{V})$. When $\boldsymbol{V}$ is singleton-valued,

$$\boldsymbol{Y}(\eta, D, X, \boldsymbol{V}; \mu, F) = \arg\max_{j \in \mathcal{Y}} \mu_j(D, X) + Q_j(\eta; X, V). \tag{5.11}$$

The model prediction in (5.11) nests Petrin and Train's (2010) specification, which assumes the additive separability of Q_j between functions of V and η:[14]

$$Q_j(\eta; X, V) = g(V; \lambda) + Q_j(\eta_j). \tag{5.12}$$

Let $\theta \equiv (\mu_1, \ldots, \mu_J, \pi, F)$. Also, let $A \subseteq \mathcal{Y}$. One can show the containment functional is

$$\begin{aligned} &\mathbb{C}_\theta(A|D = d, X = x, Z = z) \\ &= \sum_{\{j_1,\ldots,j_m\} \subset A} F_\eta\Big(\mu_{j_\ell}(d, x) \geq \inf_{v \in \boldsymbol{V}(d,x,z;\pi)} \Big(\max_{k \neq j_\ell}[\mu_k(d, x) + Q_k(\eta; x, v)] - Q_{j_\ell}(\eta; x, v)\Big), \ell = 1, \ldots, m\Big). \end{aligned} \tag{5.13}$$

The containment functional can be computed by simulating $\eta \sim U[0, 1]^J$. The following corollary characterizes $\Theta_I(P_0)$, applying Theorem 1.

Corollary 2: *Suppose $U = (U_1, \ldots, U_J)$ has a strictly positive conditional density given (X, V), and $U \perp Z|X, V$. Then, $\Theta_I(P_0)$ is the set of parameter values $\theta = (\mu_1, \ldots, \mu_J, \pi, F)$ such that, for almost all (d, x, z),*

$$\begin{aligned} &P_0(Y \in A|D = d, Z = z, X = x) \geq \\ &\sum_{B \subseteq A} F_\eta\Big(\Big\{\mu_{j_\ell}(d, x) \geq \inf_{v \in \boldsymbol{V}(d,x,z;\pi)} \Big(\max_{k \neq j_\ell}[\mu_k(d, x) + Q_k(\eta; x, v)] - Q_{j_\ell}(\eta; x, v)\Big)\Big\} \\ &\cap \Big\{\mu_{j_m}(d, x) < \inf_{v \in \boldsymbol{V}(d,x,z;\pi)} \Big(\max_{k \neq j_m}[\mu_k(d, x) + Q_k(\eta; x, v)] - Q_{j_m}(\eta; x, v)\Big)\Big\}, j_\ell \in B, j_m \in A \setminus B\Big), \\ &A \subseteq \{1, \ldots, J\}. \end{aligned} \tag{5.14}$$

As in the previous example, (5.14) jointly restricts μ, F (through Q), and π (via $\boldsymbol{V}$). Suppose further that $V \perp Z|X$. Then, π is point identified as $\pi(z, x) = P_0(D = 1|Z = z, X = x)$, which also ensures $\tilde{\pi}$ in (2.12) is point identified.

[14] In their notation, $g(V; \lambda)$ is $CF(\mu_n; \lambda)$, and $Q_j(\eta_j)$ is $\tilde{\varepsilon}_{nj}$. Their specification only allows V to shift the location of the conditional distribution of U. They show that this specification holds for several parametric models of $U|V$.

5.3 Treatment Responses with Social Interactions

We revisit Example 3. Recall

$$Y_{j,d_1,d_2} = \mu(d_1, d_2, X) + U_{d_1,d_2}, \tag{5.15}$$

We work with the conditional mean-independence assumption. Recall $D = (D_1, D_2)$. Define the model prediction

$$\boldsymbol{Y}(\eta, D, X, \boldsymbol{V}; \mu, F) = \{y \in \mathcal{Y} : y = \mu(D, X) + \lambda_D(X, V) + \eta_D, V \in \mathrm{Sel}(\boldsymbol{V})\}, \tag{5.16}$$

where $\lambda_d(x, v)$ is the conditional mean function of $U_d|X, V$.

Let us rewrite the set-valued control function in (2.17) as a union of two random sets.

$$\boldsymbol{V}(D, X, Z; \pi) = [\tilde{\boldsymbol{V}}_0(D, X, Z; \pi) \times \{0\}] \cup [\tilde{\boldsymbol{V}}_1(D, X, Z; \pi) \times \{1\}], \tag{5.17}$$

where

$$\tilde{\boldsymbol{V}}_0(D, X, Z; \pi) \equiv \begin{cases} S_{\pi,(0,1)}(Z) \cup S_{\pi,\{(1,0),(0,1)\}}(Z) & \text{if } D = (0, 1) \\ S_{\pi,(d_1,d_2)}(Z) & \text{if } D = (d_1, d_2),\ (d_1, d_2) \neq (0, 1), \end{cases} \tag{5.18}$$

and

$$\tilde{\boldsymbol{V}}_1(D, X, Z; \pi) \equiv \begin{cases} S_{\pi,(1,0)}(Z) \cup S_{\pi,\{(1,0),(0,1)\}}(Z) & \text{if } D = (1, 0) \\ S_{\pi,(d_1,d_2)}(Z) & \text{if } D = (d_1, d_2),\ (d_1, d_2) \neq (1, 0). \end{cases} \tag{5.19}$$

As in the previous examples, the sharp identification region for $\theta = (\mu, \pi, F)$ involves the supremum and infimum of a function f over $v \in \boldsymbol{V}(d, x, z; \pi)$. Eq. (5.17) suggests that the supremum, for example, can be written as

$$\sup_{v \in \boldsymbol{V}(d,x,z;\pi)} f(v_1, v_2, v_s) = \max\{ \sup_{(v_1,v_2) \in \tilde{\boldsymbol{V}}_0(d,x,z)} f(v_1, v_2, 0), \sup_{(v_1,v_2) \in \tilde{\boldsymbol{V}}_1(d,x,z)} f(v_1, v_2, 1)\}. \tag{5.20}$$

We use (4.7) from Theorem 2 and argue as in Example 1 to characterize the sharp identification region.

Corollary 3: *Suppose $E_{P_0}[|Y|] < \infty$. Suppose $U = (U_{00}, U_{10}, U_{01}, U_{11})$ has a strictly positive conditional density given (X, V). Suppose, for each $(d_1, d_2) \in \mathcal{D}$, $E[U_{d_1,d_2}|Z, X, V] = E[U_{d_1,d_2}|X, V]$, a.s. Then, $\Theta_I(P_0)$ is the set of parameter values $\theta = (\mu, \pi, F)$ such that, for* *almost all* (d,x),

$$\sup_{z\in\mathcal{Z}}\Big\{E_{P_0}[Y|D=d,X=x,Z=z]-\lambda_U(d,x,z)\Big\}$$
$$\le \mu(d,x)\le$$
$$\inf_{z\in\mathcal{Z}}\Big\{E_{P_0}[Y|D=d,X=x,Z=z]-\lambda_L(d,x,z)\Big\}, \quad (5.21)$$

where

$$\lambda_U(d,x,z)=\max\big\{\sup_{(v_1,v_0)\in\tilde{\boldsymbol{V}}_0(d,x,z;\pi)}\lambda_d(x,v_1,v_2,0),\sup_{(v_1,v_0)\in\tilde{\boldsymbol{V}}_1(d,x,z;\pi)}\lambda_d(x,v_1,v_2,1)\big\}, \quad (5.22)$$
$$\lambda_L(d,x,z)=\min\big\{\inf_{(v_1,v_0)\in\tilde{\boldsymbol{V}}_0(d,x,z;\pi)}\lambda_d(x,v_1,v_2,0),\inf_{(v_1,v_0)\in\tilde{\boldsymbol{V}}_1(d,x,z;\pi)}\lambda_d(x,v_1,v_2,1)\big\}. \quad (5.23)$$

Remark 4: One may impose further restrictions on the relationship between U and V_s via a priori restrictions on F. A leading example is to assume U is independent (or mean independent) of the selection mechanism conditional on other control variables, i.e., $U \perp V_s|X,V_0,V_1$. This assumption may be plausible if V_s is viewed as a signal that is only relevant for the treatment decision e.g., firms' profitability but irrelevant for the outcome e.g., pollution level. One can impose this restriction by restricting F to the space of conditional distributions such that $\lambda_d(x,v_1,v_2,1)=\lambda_d(x,v_1,v_2,0)=\lambda_d(x,v_1,v_2)$. This type of restriction helps simplify λ_U and λ_L.

Remark 5: With additional assumptions, it is possible to point identify π using Tamer's (2003) result. Specifically, suppose $(V_1,V_2)\perp(Z_1,Z_2)|X$, and for each j, $Z_j=(Z_{j,k},Z_{j,-k})$ contains a continuous component Z_{jk} supported on $\mathbb{R}$. Furthermore, $supp(Z_j,X|Z_{-j})=supp(Z_j,X)$, *a.s.* Tamer (2003) shows, if one can vary the continuous component to push the choice probabilities toward extreme values, i.e., π_j is such that $\lim_{z_{j,k}\to-\infty}\pi_j(0,z_{j,k},z_{j,-k},x)=0$, and $\lim_{z_{j,k}\to\infty}\pi_j(1,z_{j,k},z_{j,-k},x)=1$, then, π is point identified. This argument, however, requires a variable with a large support (for each player), which may be hard to find in practice.

5.4 Dynamic Treatment Effects

Depending on parameters of interest, each layer of the triangular dynamic system (2.19)–(2.21) can be written as

$$Y=1\{\mu(D,X)\ge U\}, \quad (5.24)$$

by properly labeling the variables. For example, we may write (2.21) as above with $\mu(D,X) = \mu_2(Y_1,D_1,D_2,X)$, $D=(Y_1,D_1,D_2)$, and $U=U_2$. In this case, we take $V=(U_1,V_1,V_2)$ as a vector of control variables and $\pi=(\mu_1,\pi_1,\pi_2)$.

One can derive $\boldsymbol{Y}$ as follows. First, let $U=Q(\eta|X,V)=F^{-1}(\eta|X,V)$. Then,

$$\begin{aligned}
Y &= 1\{\mu(D,X) \geq Q(\eta|X,V)\} \\
&= 1\{F(\mu(D,X)|X,V) \geq \eta\} \\
&= 1\{H(D,X,V) \geq \eta\},
\end{aligned}$$

where $H(d,x,v) \equiv F(\mu(d,x)|x,v)$. The model prediction $\boldsymbol{Y}$ has the following *incomplete threshold-crossing structure* (Kaido, 2022)[15]:

$$\boldsymbol{Y}(\eta,D,X,\boldsymbol{V};\mu,F) = \begin{cases} \{0\} & \eta > \sup_{V\in \mathrm{Sel}(\boldsymbol{V})} H(D,X,V) \\ \{0,1\} & \inf_{V\in \mathrm{Sel}(\boldsymbol{V})} H(d,x,v) < \eta \leq \sup_{V\in \mathrm{Sel}(\boldsymbol{V})} H(D,X,V) \\ \{1\} & \eta \leq \inf_{V\in \mathrm{Sel}(\boldsymbol{V})} H(D,X,V). \end{cases} \tag{5.25}$$

For a given (d,x,z), the model prediction depends on the realization of η relative to two thresholds $\sup_{v\in \boldsymbol{V}(d,x,z;\pi)} H(d,v)$ and $\inf_{v\in \boldsymbol{V}(d,x,z;\pi)} H(d,v)$. If η is below the lower threshold, the model predicts $\boldsymbol{Y}=\{1\}$, whereas $\boldsymbol{Y}=\{0\}$ if η is above the upper threshold. The model predicts $\boldsymbol{Y}=\{0,1\}$ if η is between the two thresholds (see Figure 2).

The containment functional of $\boldsymbol{Y}$ in (5.25) for $A=\{1\}$ is

$$\begin{aligned}
\mathbb{C}_\theta(\{1\}|D=d,X=x,Z=z) &= F_\eta(\boldsymbol{Y}(\eta,D,X,\boldsymbol{V};\mu,F) \subseteq \{1\}|D=d,X=x,Z=z) \\
&= \inf_{v\in \boldsymbol{V}(d,x,z;\pi)} H(d,x,v).
\end{aligned}$$

Similarly, the containment functional for $A=\{0\}$ is

$$\mathbb{C}_\theta(\{0\}|D=d,X=x,Z=z) = 1 - \sup_{v\in \boldsymbol{V}(d,x,z;\pi)} H(d,x,v).$$

We apply this argument sequentially to (2.19)–(2.21) to characterize the sharp identification

[15] The incomplete threshold-crossing structure also appears in semiparametric binary choice models with interval-valued covariates (Manski and Tamer, 2002). Manski and Tamer's (2002) model is $Y=1\{W'\theta+\delta X+\epsilon>0\}$, where W is exogenous, X is interval-valued (i.e. $X\in[X_L,X_U]$), $\delta>0$ and ϵ satisfies a quantile independence condition. See also Molinari (2020) (Section 3.1.1) for an extensive discussion of their model.

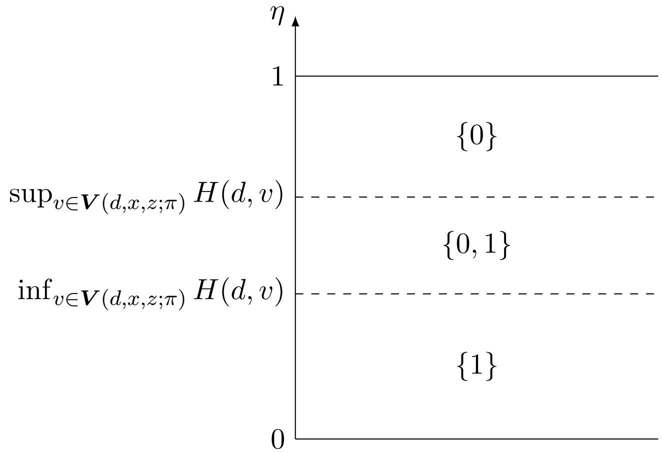


Figure 2: An incomplete threshold-crossing structure.

Note: The figure shows the value of $\boldsymbol{Y}(\eta|d, \boldsymbol{V}; \nu, F)$ as a function of η.

region. For this, define the following objects

$$\begin{aligned} H_{Y_2}(y_1, d_1, d_2, x, v; \theta) &\equiv F_{U_2|X,U_1,V_1,V_2}(\mu_2(y_1, d_1, d_2, x)|x, u_1, v_1, v_2) \\ H_{D_2}(y_1, d_1, z_2, x, \tilde{v}; \theta) &\equiv F_{V_2|X,U_1,V_1}(\pi_2(y_1, d_1, z_2, x)|x, u_1, v_1) \\ H_{Y_1}(d_1, x, v_1; \theta) &\equiv F_{U_1|X,V_1}(\mu_1(d_1, x)|x, v_1), \end{aligned}$$

where $\tilde{v} = (u_1, v_1)$. Recall $\boldsymbol{V}$ was defined as in (2.22). Also define

$$\begin{aligned} \boldsymbol{V}_{D_2}(Y_1, D_1, Z, X; \theta) &= \boldsymbol{V}_{U_1}(D, X; \mu_1) \times \boldsymbol{V}_1(D, Z_1, X; \pi_1) \\ \boldsymbol{V}_{Y_1}(D_1, Z_1, X; \theta) &= \boldsymbol{V}_1(D, Z_1, X; \pi_1). \end{aligned}$$

Then, we can use (4.2) from Theorem 1 and characterize the sharp identification region.

Corollary 4: *Suppose $U_2, U_1, V_1, V_2|X$ has a positive density with respect to Lebesgue measure. Suppose (i) $U_1 \perp Z_1|X, V_1$ (ii) $V_2 \perp Z_1|X, U_1, V_1$ and (iii) $U_2 \perp (Z_1, Z_2)|X, V_2, U_1, V_1$. Then, $\Theta_I(P_0)$ is the set of parameter values $\theta = (\mu_1, \mu_2, \pi_1, \pi_2, F)$ such that, for almost all*

(d, x, z),

$$\inf_{v \in \boldsymbol{V}(d,x,z;\theta)} H_{Y_2}(y_1, d_1, d_2, x, v; \theta)$$
$$\leq P_0(Y_2 = 1 | Y_1 = y_1, D_1 = d_1, D_2 = d_2, X = x, Z = z)$$
$$\leq \sup_{v \in \boldsymbol{V}(d,x,z;\theta)} H_{Y_2}(y_1, d_1, d_2, x, v; \theta), \quad (5.26)$$

$$\inf_{\tilde{v} \in \boldsymbol{V}_{D_2}(y_1,d_1,x,z;\theta)} H_{D_2}(y_1, d_1, x, z_2, \tilde{v}; \theta)$$
$$\leq P_0(D_2 = 1 | Y_1 = y_1, D_1 = d_1, X = x, Z = z)$$
$$\leq \sup_{\tilde{v} \in \boldsymbol{V}_{D_2}(y_1,d_1,x,z;\theta)} H_{D_2}(y_1, d_1, x, z_2, \tilde{v}; \theta), \quad (5.27)$$

and

$$\inf_{v_1 \in \boldsymbol{V}_{Y_1}(d_1,x,z_1;\theta)} H_{Y_1}(d_1, x, v_1; \theta)$$
$$\leq P_0(Y_1 = 1 | D_1 = d_1, X = x, Z_1 = z_1)$$
$$\sup_{v_1 \in \boldsymbol{V}_{Y_1}(d_1,x,z_1;\theta)} H_{Y_1}(d_1, x, v_1; \theta). \quad (5.28)$$

In the corollary, the conditional independence assumptions (i), (ii), (iii) for the IVs are useful in constructing informative bounds on μ_1, π_2, and μ_2, respectively. As in Example 1, π_1 can be point identified from the selection equation in period 1 if $V_1 \perp Z_1 | X$.

Remark 6: In the above illustrations, we paired continuous outcome variables with the generalized Roy model and strategic treatment decisions. We paired discrete (multinomial and binary) outcomes with other examples of selection processes. These choices were arbitrary. Theorems 1 and 2 are flexible. They allow the researcher to combine various outcome variable types and selection models.

6 Concluding remarks

Observational data are often generated through complex decision processes. Allowing control functions to be set-valued, this paper expands the scope of the control function approach. The proposed framework accommodates, for example, selection processes that involve rich heterogeneity, dynamic optimizing behavior, or social interaction. Our identifying restrictions are inequalities on the conditional choice probabilities. One can conduct inference using moment-based methods or likelihood-based inference methods. Practitioners can use the results of this paper for various purposes. First, they can evaluate social programs nonparametrically, while taking into account potentially complex treatment selection processes. Second, the bounds in our main identification results can easily be combined with a range of shape restrictions and parametric assumptions, allowing practitioners to conduct a sensitivity analysis to assess the additional identifying power of specific assumptions. The tools from random set theory enable us to guarantee sharpness of bounds one obtains in such a sensitivity analysis, without needing to prove sharpness case after case.

References

Abadie, A., J. Angrist, and G. Imbens (2002): "Instrumental variables estimates of the effect of subsidized training on the quantiles of trainee earnings," *Econometrica*, 70, 91–117.

Abdulkadiroğlu, A., P. A. Pathak, J. Schellenberg, and C. R. Walters (2020): "Do Parents Value School Effectiveness?" *American Economic Review*, 110, 1502–39.

Andrews, D. W. K. and X. Shi (2013): "Inference based on conditional moment inequalities." *Econometrica*, 81, 609–666.

Aronow, P. M. and C. Samii (2017): "Estimating average causal effects under general interference, with application to a social network experiment," *The Annals of Applied Statistics*, 11, 1912–1947.

Auerbach, E. (2022): "Identification and estimation of a partially linear regression model using network data," *Econometrica*, 90, 347–365.

Balat, J. F. and S. Han (2022): "Multiple treatments with strategic substitutes," *Journal of Econometrics*.

Bishop, K. C., J. D. Ketcham, and N. V. Kuminoff (2022): "Hazed and Confused: The Effect of Air Pollution on Dementia," *The Review of Economic Studies*, 90, 2188–2214.

Blundell, R. and J. L. Powell (2003): "Endogeneity in nonparametric and semiparametric regression models," *Econometric society monographs*, 36, 312–357.

Blundell, R. W. and J. L. Powell (2004): "Endogeneity in Semiparametric Binary Response Models," *The Review of Economic Studies*, 71, 655–679.

Blundell, R. W. and R. J. Smith (1989): "Estimation in a class of simultaneous equation limited dependent variable models," *The Review of Economic Studies*, 56, 37–57.

Card, D., S. DellaVigna, P. Funk, and N. Iriberri (2019): "Are Referees and Editors in Economics Gender Neutral?*," *The Quarterly Journal of Economics*, 135, 269–327.

Carlier, G., A. Galichon, and F. Santambrogio (2010): "From Knothe's Transport to Brenier's Map and a Continuation Method for Optimal Transport," *SIAM Journal on Mathematical Analysis*, 41, 2554–2576.

Carneiro, P., J. J. Heckman, and E. Vytlacil (2010): "Evaluating marginal policy changes and the average effect of treatment for individuals at the margin," *Econometrica*, 78, 377–394.

Chen, X. and T. M. Christensen (2018): "Optimal sup-norm rates and uniform inference on nonlinear functionals of nonparametric IV regression," *Quantitative Economics*, 9, 39–84.

Chen, X., T. M. Christensen, and E. Tamer (2018): "Monte Carlo Confidence Sets for Identified Sets," *Econometrica*, 86, 1965–2018.

Chernozhukov, V., I. Fernández-Val, W. Newey, S. Stouli, and F. Vella (2020): "Semiparametric estimation of structural functions in nonseparable triangular models," *Quantitative Economics*, 11, 503–533.

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

Chernozhukov, V., S. Lee, and A. M. Rosen (2013): "Intersection bounds: estimation and inference." *Econometrica*, 81, 667–737.

Chesher, A. (2003): "Identification in nonseparable models," *Econometrica*, 71, 1405–1441.

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

Chesher, A. and A. M. Rosen (2013): "What Do Instrumental Variable Models Deliver with Discrete Dependent Variables?" *American Economic Review*, 103, 557–62.

——— (2017): "Generalized instrumental variable models," *Econometrica*, 85, 959–989.

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

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

Das, M., W. K. Newey, and F. Vella (2003): "Nonparametric estimation of sample selection models," *The Review of Economic Studies*, 70, 33–58.

D'Haultfoeuille, X. and P. Février (2015): "Identification of Nonseparable Triangular Models With Discrete Instruments," *Econometrica*, 83, 1199–1210.

D'Haultfœuille, X., S. Hoderlein, and Y. Sasaki (2021): "Testing and relaxing the exclusion restriction in the control function approach," *Journal of Econometrics*.

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

Gautier, E. and S. Hoderlein (2011): "A triangular treatment effect model with random coefficients in the selection equation," .

Graham, B. S. (2011): "Econometric Methods for the Analysis of Assignment Problems in the Presence of Complementarity and Social Spillovers," .

Hall, P. and J. L. Horowitz (2005): "Nonparametric methods for inference in the presence of instrumental variables," *The Annals of Statistics*, 33, 2904–2929.

Han, S. (2021): "Identification in nonparametric models for dynamic treatment effects," *Journal of Econometrics*, 225, 132–147.

Han, S. and S. Lee (2023): "Semiparametric Models for Dynamic Treatment Effects and Mediation Analyses with Observational Data," .

Han, S. and S. Yang (2024): "A computational approach to identification of treatment effects for policy evaluation," *Journal of Econometrics*, 240, 105680.

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

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

Heckman, J. J. and E. J. Vytlacil (2007): "Chapter 71 Econometric Evaluation of Social Programs, Part II: Using the Marginal Treatment Effect to Organize Alternative Econometric Estimators to Evaluate Social Programs, and to Forecast their Effects in New Environments," Elsevier, vol. 6 of *Handbook of Econometrics*, 4875–5143.

Imbens, G. W. and J. D. Angrist (1994): "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 62, 467–475.

Imbens, G. W. and W. K. Newey (2002): "Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity," Working Paper 285, National Bureau of Economic Research.

——— (2009): “Identification and estimation of triangular simultaneous equations models without additivity,” *Econometrica*, 77, 1481–1512.

Joe, H. (2014): *Dependence modeling with copulas*, CRC press.

Jovanovic, B. (1989): “Observable implications of models with multiple equilibria,” *Econometrica: Journal of the Econometric Society*, 1431–1437.

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

Kaido, H. (2022): “A Control Function Approach to Binary Response Models with Discrete Endogenous Variables,” Working Paper, Boston University.

Kaido, H. and F. Molinari (2022): “Information Based Inference In Models with Set-Valued Predictions and Misspecification,” Working Paper, Boston University.

Kline, P. and C. R. Walters (2016): “Evaluating Public Programs with Close Substitutes: The Case of Head Start*,” *The Quarterly Journal of Economics*, 131, 1795–1848.

——— (2019): “On Heckits, LATE, and Numerical Equivalence,” *Econometrica*, 87, 677–696.

Luo, Y. and H. Wang (2017): “Core Determining Class: Construction, Approximation and Inference,” Working Paper, University of Florida.

Machado, C., A. M. Shaikh, and E. J. Vytlacil (2019): “Instrumental variables and the sign of the average treatment effect,” *Journal of Econometrics*, 212, 522–555.

Manski, C. F. (2013): “Identification of treatment response with social interactions,” *The Econometrics Journal*, 16, S1–S23.

Manski, C. F. and E. Tamer (2002): “Inference on Regressions with Interval Data on a Regressor or Outcome,” *Econometrica*, 70, 519–546.

Mogstad, M., A. Santos, and A. Torgovitsky (2018): “Using Instrumental Variables for Inference About Policy Relevant Treatment Parameters,” *Econometrica*, 86, 1589–1619.

Molchanov, I. (2017): *Theory of random sets*, vol. 19, Springer London, 2 ed.

Molchanov, I. and F. Molinari (2018): *Random sets in econometrics*, vol. 60, Cambridge University Press.

Molinari, F. (2020): "Chapter 5 - Microeconometrics with partial identification," in *Handbook of Econometrics, Volume 7A*, ed. by S. N. Durlauf, L. P. Hansen, J. J. Heckman, and R. L. Matzkin, Elsevier, vol. 7 of *Handbook of Econometrics*, 355–486.

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

Newey, W. K. (1987): "Efficient estimation of limited dependent variable models with endogenous explanatory variables," *Journal of Econometrics*, 36, 231–250.

Newey, W. K. and J. L. Powell (2003): "Instrumental variable estimation of nonparametric models," *Econometrica*, 71, 1565–1578.

Newey, W. K., J. L. Powell, and F. Vella (1999): "Nonparametric Estimation of Triangular Simultaneous Equations Models," *Econometrica*, 67, 565–603.

Newey, W. K. and S. Stouli (2021): "Control variables, discrete instruments, and identification of structural functions," *Journal of Econometrics*, 222, 73–88.

Petrin, A. and K. Train (2010): "A Control Function Approach to Endogeneity in Consumer Choice Models," *Journal of Marketing Research*, 47, 3–13.

Pinkse, J. (2000): "Nonparametric two-step regression estimation when regressors and error are dependent," *Canadian Journal of Statistics*, 28, 289–300.

Ponomarev, K. (2022): "Essays in Econometrics," Ph.D. thesis, UCLA.

Ponomareva, M. and E. Tamer (2011): "Misspecification in moment inequality models: back to moment equalities?" *The Econometrics Journal*, 14, 186–203.

Rivers, D. and Q. H. Vuong (1988): "Limited information estimators and exogeneity tests for simultaneous probit models," *Journal of Econometrics*, 39, 347–366.

Robins, J. M. (1997): "Latent Variable Modeling and Applications to Causality," *Lecture Notes in Statistics*, 69–117.

Rubin, D. B. (1978): "Bayesian Inference for Causal Effects: The Role of Randomization," *The Annals of Statistics*, 6, 34–58.

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

Smith, R. J. and R. W. Blundell (1986): "An exogeneity test for a simultaneous equation Tobit model with an application to labor supply," *Econometrica: journal of the Econometric Society*, 679–685.

Sosa-Rubí, S. G., O. Galárraga, and J. E. Harris (2009): "Heterogeneous impact of the "Seguro Popular" program on the utilization of obstetrical services in Mexico, 2001–2006: A multinomial probit model with a discrete endogenous variable," *Journal of Health Economics*, 28, 20–34.

Tamer, E. (2003): "Incomplete Simultaneous Discrete Response Model with Multiple Equilibria," *The Review of Economic Studies*, 70, 147–165.

——— (2010): "Partial Identification in Econometrics," *Annual Review of Economics*, 2, 167–195.

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

Villani, C. (2008): *Optimal Transport: Old and New*, Grundlehren der mathematischen Wissenschaften, Springer Berlin Heidelberg.

Vuong, Q. and H. Xu (2017): "Counterfactual mapping and individual treatment effects in nonseparable models with binary endogeneity," *Quantitative Economics*, 8, 589–610.

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

A Comparison with Chesher and Rosen (2017)

In characterizing identified sets for structural parameters, Chesher and Rosen (2017) utilize random sets of unobservables. In this section, we compare our approach with theirs. The main propose of the comparison is to illustrate that the two methods are non-nested and complementary.

We summarize the characterization of the identified set in Chesher and Rosen (2017) with notation close to ours. Let Y be a vector of endogenous variables, Z be a vector of exogenous variables (e.g., IVs), and U be a vector of structural unobservables. Then define a random closed set of U as

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

where h is a structural function, the features of which are of interest. Assume $(h, F_{U|Z}) \in \mathcal{M}$ where $\mathcal{M}$ incorporates identifying assumptions. Chesher and Rosen (2017) use the following Artstein's inequality: For $F_{U|Z}(\cdot|z) \in \mathsf{F}_{U|Z}$ being the distribution of one of the measurable selections of $\boldsymbol{U}(Y,Z;h)$,

$$F_{U|Z}(B|z) \geq \mathbb{C}_h(B|z)$$

holds for all closed sets $B \in \mathcal{F}(\mathcal{U})$ where $\mathbb{C}_h(B|z) \equiv P[\boldsymbol{U}(Y,Z;\mu) \subseteq B|Z=z]$. Then, the identified set can be characterized as

$$\{(h, F_{U|Z}) \in \mathcal{M} : F_{U|Z}(B|Z) \geq \mathbb{C}_h(B|Z),\ a.s.\ \forall B \in \mathcal{F}(\mathcal{U})\}.$$

They also provide characterization using the Aumman expectation of $\boldsymbol{U}(Y,Z;h)$.

On the other hand, we characterize the identified set as

$$\{(\mu, F_{U|X,V}) : P_0(A|D,X,Z) \geq \mathbb{C}_{\mu,F}(A|D,X,Z),\ a.s.\ \forall A \in \mathcal{F}(\mathcal{Y}),\ \pi \in \Pi_r(P_0)\},$$

where $\mathbb{C}_{\mu,F}(A|D,X,Z) \equiv F_\eta(\boldsymbol{Y}(\eta,D,X,\boldsymbol{V};\mu,F_{U|X,V}) \subseteq A)$. We also provide characterization using the Aumman expectation of $\boldsymbol{Y}$.

The two approaches share similar features in that only *sets* of unobservables can be recovered from observed data, while stochastic restrictions are imposed on true unobservables. However, the two differ in several ways. First, the CF assumption is imposed on $F_{U|D,X,V}$ whereas the IV assumption is imposed on $F_{U|Z}$. Note that the two stochastic assumptions are not nested. Even if we were to use the IV approach by having our (U,V) as their U and μ as part of h, their framework is not suitable to impose the CF assumption. Second, the containment functional is compared to the observed distribution to construct the identified set in our setting, while it is compared to the unobserved distribution in theirs. This difference may have implications on implementation in practice.

In sum, the two approaches offer complementary tools to practitioners for robust causal analyses. Practitioners can select the most appropriate approach based on the specific model at hand and their belief on the stochastic nature of their problem.

B Proofs

B.1 Proofs of Theorems 1, 2, and 3

Proof of Theorem 1. By Assumptions 1 and 3, one may represent the outcome as $Y = \mu(D, X, U) = \mu(D, X, Q(\eta; X, V))$. By Assumption 2, V is a measurable selecition of $\boldsymbol{V}$, and therefore Y is a measurable selection of $\boldsymbol{Y}(\eta, D, X, \boldsymbol{V}; \mu, F)$. Therefore, under Assumptions 1-3, the model's prediction is summarized by

$$Y \in \boldsymbol{Y}(\eta, D, X, \boldsymbol{V}; \mu, F), \text{ a.s.} \tag{B.1}$$

By Assumption 2 (ii), $\boldsymbol{V}$ is a function of (D, X, Z). Hence, one may condition on $(D, X, \boldsymbol{V})$ by conditioning on (D, X, Z). By Artstein's inequality (see Molinari, 2020, Theorem A.1.), the distribution $P_0(A|D, X, Z)$ is the conditional law of a measurable selection of $\boldsymbol{Y}$ if and only if

$$P_0(A|D, X, Z) \geq \mathbb{C}_\theta(A|D, X, Z), \ \forall A \in \mathcal{F}(\mathbb{R}^{d_Y}). \tag{B.2}$$

This ensures the representation of the sharp identification region by the inequalities above. □

A random closed set $\boldsymbol{X}$ is said to be *integrable* if $\boldsymbol{X}$ has at least one integrable selection. We define the Aumann (or selection) expectation of an integrable random closed set as follows (Molinari, 2020). For this, we let $\mathrm{Sel}^1(\boldsymbol{X})$ denote the set of integrable selections of $\boldsymbol{X}$.

Definition 5: *The* Aumann expectation *of an integrable random closed set* $\boldsymbol{X}$ *is given by*

$$\mathbb{E}[\boldsymbol{X}] = \mathrm{cl}\left\{E[X] :, X \in \mathrm{Sel}^1(\boldsymbol{X})\right\}. \tag{B.3}$$

For each sub σ-algebra $\mathfrak{B} \subset \mathfrak{F}$, the conditional Aumann expectation of X given $\mathfrak{B}$ is the $\mathfrak{B}$-measurable random closed set $\boldsymbol{Y} = \mathbb{E}(\boldsymbol{X}|\mathfrak{B})$ such that the family of $\mathfrak{B}$-measurable integrable *selections of* $\boldsymbol{Y}$*, denoted* $\mathrm{Sel}^1_{\mathfrak{B}}(\boldsymbol{Y})$*, satisfies*

$$\mathrm{Sel}^1_{\mathfrak{B}}(\boldsymbol{Y}) = \mathrm{cl}\left\{E[X|\mathfrak{B}] :, X \in \mathrm{Sel}^1(\boldsymbol{X})\right\}. \tag{B.4}$$

where the closure in the right-hand side is taken in L^1*.*

Proof of Theorem 2. Let $\mathfrak{B} \equiv \sigma(D, X, Z)$ be the σ-algebra generated by (D, X, Z). By Assumptions 2 and 4, we may represent the model's set-valued prediction by $\boldsymbol{Y}$ in (4.6), the random set of outcomes $Y = \mu(D, X) + \lambda_D(X, V) + \eta_D$, where $\eta = (\eta_d, d \in \mathcal{D})$ is conditionally mean independent of D. $\boldsymbol{Y}$ is integrable because its measurable selection Y is assumed to be integrable. Because of $Y \in \mathrm{Sel}^1(\boldsymbol{Y})$, the model's prediction on the conditional mean is summarized by

$$E_{P_0}[Y|\mathfrak{B}] \in \mathbb{E}[\boldsymbol{Y}(\eta, D, X, Z; \mu, F)|\mathfrak{B}],\ a.s., \tag{B.5}$$

where the right-hand side is the conditional Aumann expectation of $\boldsymbol{Y}$. Let $b \in \{-1, 1\}$. Then, (B.5) is equivalent to

$$bE_{P_0}[Y|\mathfrak{B}] \leq s(b, \mathbb{E}[\boldsymbol{Y}(\eta, D, X, Z; \mu, F)|\mathfrak{B}]), \tag{B.6}$$

where $s(b, K) \equiv \sup_{x \in K} bx$ is the support function of K.

Now, we use the convexification property of the Aumann (selection) expectation of random closed sets. Technically, Aumann expectation depends on the probability space used to define $\boldsymbol{Y}$ (Molchanov, 2017, Sec. 2.1.2). Hence, we proceed as follows. Let $\Omega = \mathbb{R}^{d_U} \times \mathbb{R}^{d_D} \times \mathbb{R}^{d_X} \times \mathbb{R}^{d_Z}$ be the sample space, and let $\mathfrak{F} = \mathfrak{F}_{\mathbb{R}^{d_U}} \otimes \mathfrak{F}_{\mathbb{R}^{d_D}} \otimes \mathfrak{F}_{\mathbb{R}^{d_X}} \otimes \mathfrak{F}_{\mathbb{R}^{d_Z}}$ be the product σ-algebra, where $\mathfrak{F}_E$ is the Borel σ-algebra over E. Let $\mathbb{F}$ be a probability measure on $(\Omega, \mathfrak{F})$. Measurable maps (η, D, X, Z) are defined on this space. Consider a measurable rectangle $A = A_\eta \times A_{D,X,Z}$, where $A_\eta \subset \mathbb{R}^{d_U}$ and $A_{D,X,Z} \subset \mathbb{R}^{d_D} \times \mathbb{R}^{d_X} \times \mathbb{R}^{d_Z}$. Then, $\mathbb{F}(A|\mathfrak{B}) = F_\eta(A_\eta)$. By Assumption 3 and the construction of η, F_η is atomless. Since any $A \in \mathfrak{F}$ can be approximated by a countable union of measurable rectangles, conclude that $\mathbb{F}$ is atomless over $\mathfrak{B}$.[16]

By the convexification theorem (Molinari, 2020, Theorem A.2.), $\mathbb{E}[\boldsymbol{Y}(\eta, D, X, Z; \mu, F)|\mathfrak{B}]$ is convex and

$$s(b, \mathbb{E}[\boldsymbol{Y}(\eta, D, X, Z; \mu, F)|\mathfrak{B}]) = E[s(b, \boldsymbol{Y}(\eta, D, X, Z; \mu, F))|\mathfrak{B}],\ b \in \{1, -1\}. \tag{B.7}$$

[16] An event $A' \in \mathfrak{B}$ is called a $\mathfrak{B}$-atom if $\mathbb{F}(0 < \mathbb{F}(A|\mathfrak{B}) < \mathbb{F}(A'|\mathfrak{B})) = 0$ for all $A \subset A'$ such that $A \in \mathfrak{F}$

For $b = 1$,

$$
\begin{aligned}
E[s(b, \boldsymbol{Y}(\eta, D, X, Z; \mu, F))|\mathfrak{B}] &= E[\sup_{y \in \boldsymbol{Y}(\eta, D, X, Z; \mu, F)} y|\mathfrak{B}] \\
&= \mu(d, x) + E[\sup_{v \in \boldsymbol{V}(D, X, Z; \pi)} \lambda_d(X, V)|\mathfrak{B}] \\
&= \mu(d, x) + \sup_{v \in \boldsymbol{V}(d, x, z; \pi)} \lambda_d(x, v). \qquad \text{(B.8)}
\end{aligned}
$$

By (B.6)-(B.8), $E_{P_0}[Y|D = d, X = x, Z = z] \leq \mu(d, x) + \sup_{v \in \boldsymbol{V}(d,x,z;\pi)} \lambda_d(x, v)$. For $b = -1$, the argument is similar. $\square$

Proof of Theorem 3. Let $\theta \in \Theta_I(P_0)$. Let $\boldsymbol{W} \equiv \{X\} \times \boldsymbol{V}$. Define

$$
\mathfrak{K}_I(d; \theta) \equiv \{\kappa(d) \in \mathbb{R} : \kappa(d) = \int \varphi(\mu(d, x, Q(\eta; w))d\eta dF_W(w),\ W \in \mathrm{Sel}(\boldsymbol{W})\}. \qquad \text{(B.9)}
$$

This set collects the values of $\kappa(d)$ compatible with θ for some measurable selection W of $\boldsymbol{W}$. The sharp identification region for $\kappa(d)$ is

$$
\mathfrak{K}_I(d) = \bigcup_{\theta \in \Theta_I(P_0)} \mathfrak{K}_I(d; \theta). \qquad \text{(B.10)}
$$

Hence, for the conclusion of the theorem, it suffices to show $\mathfrak{K}_I(d; \theta) = [\underline{\kappa}(d; \theta), \overline{\kappa}(d; \theta)]$.

For this, we represent $\mathfrak{K}_I(d; \theta)$ as the Aumann expectation of a random set and apply the convexification theorem. Define

$$
\boldsymbol{K}(d; \theta) \equiv \Big\{r \in \mathbb{R} : r = \int \varphi(\mu(d, x, Q(\eta; W))d\eta,\ W \in \mathrm{Sel}(\boldsymbol{W})\Big\}. \qquad \text{(B.11)}
$$

Then, by construction, $\mathfrak{K}_I(d; \theta)$ is the Aumann expectation of $\boldsymbol{K}(d; \theta)$. Under the assumption that the underlying probability space is non-atomic, we may apply the convexification theorem (Molinari, 2020, Theorem A.2.). It ensures $\mathfrak{K}_I(d; \theta) = \mathbb{E}[\boldsymbol{K}(d; \theta)]$ is a convex closed set. Since φ is bounded, $\mathfrak{K}_I(d; \theta)$ is a bounded closed interval. Again, by Theorem A.2. of Molinari (2020), its upper bound is

$$
\begin{aligned}
&s(1, \boldsymbol{K}(d; \theta)) = s(1, \mathbb{E}[\boldsymbol{K}(d; \theta)]) = E[s(1, \boldsymbol{K}(d; \theta))] \\
&\quad = E[\sup_{w \in \boldsymbol{W}} \int \varphi(\mu(d, x, Q(\eta; W))d\eta] = E[\sup_{v \in \boldsymbol{V}} \int \varphi(\mu(d, x, Q(\eta; X, v))] = \overline{\kappa}(d; \theta), \qquad \text{(B.12)}
\end{aligned}
$$

where we used $\boldsymbol{W} = \{X\} \times \boldsymbol{V}$. The argument for the lower bound is similar and is omitted. $\square$

B.2 Lemmas

LEMMA 1: *Suppose* μ *is a measurable function. Then,* $\boldsymbol{Y}(\eta, D, X, \boldsymbol{V}; \mu, F)$ *is a random closed set.*

Proof. $\boldsymbol{Y}(\eta, D, X, \boldsymbol{V}; \mu, F)$ being closed is immediate from the definition. We show its measurability below. Write $\boldsymbol{Y}(\eta(\omega), D(\omega), X(\omega), \boldsymbol{V}(\omega); \mu, F)$ as $\boldsymbol{Y}(\omega)$ for short. Since $\boldsymbol{V}$ is a random closed set, there is a sequence $\{V_n\}$ such that $\boldsymbol{V} = \mathrm{cl}(\{V_n, n \geq 1\})$ by Theorem 1.3.3 in Molchanov (2017). Let $\upsilon_n(\omega) = \mu(D(\omega), X(\omega), Q(\eta(\omega), V_n(\omega)))$ and note that $\boldsymbol{Y} = \mathrm{cl}(\{\upsilon_n, n \geq 1\})$ by Lemma 2. Then, for any $x \in \mathcal{Y}$, the distance function

$$\rho(x, \boldsymbol{Y}(\omega)) = \inf\{\|x - y\|, y \in \boldsymbol{Y}(\omega)\} = \inf\{\|x - \upsilon_n(\omega)\|, n \geq 1\} \tag{B.13}$$

is a random variable in $[0, \infty]$. Again, by Theorem 1.3.3 in Molchanov (2017), the conclusion follows. $\square$

Consider a random closed set $\boldsymbol{X}$ that is nonempty almost surely. A countable family of selections $\xi_n \in \mathrm{Sel}(\boldsymbol{X}), n \geq 1$ is called the *Castaing representation* of $\boldsymbol{X}$ if $\boldsymbol{X} = \mathrm{cl}(\{\xi_n, n \geq 1\})$. Such representation exists for any random closed set (Molchanov, 2017).

LEMMA 2: *Let* $\boldsymbol{X}$ *be a random closed set, and let* $\{\xi_n, n \geq 1\}$ *be its Castaing representation. For each* $\omega \in \Omega$*, let* $\boldsymbol{Y}(\omega) \equiv cl\{y \in \mathcal{Y} : y = f(\omega, \xi(\omega)), \xi \in \mathrm{Sel}(\boldsymbol{X})\}$ *for a measurable map* $f : \Omega \times \mathcal{X} \to \mathcal{Y}$*. Then* $\boldsymbol{Y}$ *is a random closed set with a Castaing representation* $\{\upsilon_n\}$ *with* $\upsilon_n(\omega) = f(\omega, \xi_n(\omega))$ *for* $n \geq 1$*.*

Proof. Le $\{y_n, n \geq 1\}$ be an enumeration of a countable dense set in $\mathcal{Y}$. For each $n, k \geq 1$ and $\omega \in \Omega$, let $C_{k,n}(\omega) = \{x \in \mathcal{X} : f(\omega, x) \cap B_{2^{-k}}(y_n) \neq \emptyset\}$. Let $\boldsymbol{X}_{k,n} \equiv \boldsymbol{X} \cap C_{k,n}$ if the intersection is nonempty and let $\boldsymbol{X}_{k,n} = \boldsymbol{X}$ otherwise. Note that $\boldsymbol{X}_{k,n}$ itself is a random closed set. For each k, n, there is $m \in \mathbb{N}$ such that ξ_m is a measurable selection of $\boldsymbol{X}_{k,n}$. For each ω with $y \in \boldsymbol{Y}(\omega)$, we have $y \in B_{2^{-k}}(y_n)$ for some k, n, and

$$\|y - \upsilon_{k,n}\| \leq \|y - y_n\| + \|y_n - \upsilon_{k,n}\| \leq 2^{-k+1}, \tag{B.14}$$

where $\upsilon_{k,n} = f(\omega, \xi_m)$. Therefore, the conclusion follows. $\square$

B.3 Proofs of Corollaries

Proof of Corollary 1. We show Assumptions 2-4 and invoke Theorem 2. First, define

$$\boldsymbol{V}(D,Z,X;\pi)=\begin{cases}[0,\pi(Z,X)] & \text{if } D=1\\ [\pi(Z,X),1] & \text{if } D=0.\end{cases} \tag{B.15}$$

Then Assumption 2 (i) holds by the selection equation (2.7)–(2.8). Assumption 2 (ii) holds because $\boldsymbol{V}$ is a function of (D,Z,X) and π. Assumptions 3-4 hold by hypothesis. By Theorem 2, each $\theta=(\mu,F,\pi)$ in the sharp identified set satisfies

$$\mu(d,x)+\lambda_L(d,x,z)\le E_{P_0}[Y|D=d,X=x,Z=z]\le\mu(d,x)+\lambda_U(d,x,z).$$

Rearranging them yields

$$E_{P_0}[Y|D=d,X=x,Z=z]-\lambda_U(d,x,z)\le\mu(d,x)\le E_{P_0}[Y|D=d,X=x,Z=z]-\lambda_L(d,x,z).$$

Note that μ does not depend on z. Taking the supremum of the lower bounds and taking the infimum of the upper bounds over $z\in\mathcal{Z}$ yields the desired result. □

Proof of Corollary 2. We show Assumptions 1-3 and invoke Theorem 1. Assumption 1 holds because D is a function of (Z,X,V) in (2.11), and $U\perp Z|X,V$. Assumption 3 holds by hypothesis. Define

$$\boldsymbol{V}(D,Z,X;\pi)=\begin{cases}\left\{v\in\mathbb{R}^2:\tilde{\pi}(Z,X)+(1-Z)v_0+Zv_1\ge 0\right\} & \text{if } D=1\\ \left\{v\in\mathbb{R}^2:\tilde{\pi}(Z,X)+(1-Z)v_0+Zv_1\le 0\right\} & \text{if } D=0,\end{cases} \tag{B.16}$$

where $\tilde{\pi}(z,x)=\pi(0,x)+z(\pi(1,x)-\pi(0,x))$. Assumption 2 (i) holds by (2.10) and (2.11). Also, Assumption 2 (ii) holds because $\boldsymbol{V}$ is a function of (D,Z,X) and π. By Theorem 1, θ is in the sharp identified set iff $P_0(A|D,X,Z)\ge\mathbb{C}_\theta(A|D,X,Z)$ holds.

For the main result, it remains to show (5.13). Let $A\subset\{1,\dots J\}$ and write the model's prediction as $\boldsymbol{Y}$ for short. Then,

$$\begin{aligned}
&\{\boldsymbol{Y} \subseteq A\} \\
&= \bigcup_{B \subseteq A} \{\boldsymbol{Y} = B\} \\
&= \bigcup_{B \subseteq A} \Big(\bigcap_{j_\ell \in B} \Big\{ \mu_{j_\ell}(D, X) \geq \inf_{V \in \mathrm{Sel}(\boldsymbol{V})} \Big(\max_{k \neq j_\ell} [\mu_k(D, X) + Q_k(\eta; X, V)] - Q_{j_\ell}(\eta; X, V) \Big) \Big\} \Big) \\
&\qquad \cap \Big(\bigcap_{j_m \in A \setminus B} \Big\{ \mu_{j_n}(D, X) < \inf_{V \in \mathrm{Sel}(\boldsymbol{V})} \Big(\max_{k \neq j_m} [\mu_k(D, X) + Q_k(\eta; X, V)] - Q_{j_m}(\eta; X, V) \Big) \Big\} \Big).
\end{aligned}$$

Conditioning on (D, X, Z) and evaluating the probability on the right-hand side by F_η yields (5.13).

$\square$

Proof of Corollary 3. The main argument is essentially the same as the proof of Corollary 1. Hence, we omit it. Here, we derive (5.22). By Theorem 2,

$$\lambda_U(d, x, z) = \sup_{v \in \boldsymbol{V}(d,x,z;\pi)} \lambda_d(x, v), \tag{B.17}$$

where $v = (v_1, v_2, v_s)$. By (5.20), the identifying restrictions in (5.22) follow. One can show (5.23) by a similar argument.

$\square$

Proof of Corollary 4. We first note that, conditional on (X, U_1, V_1, V_2), the endogenous variables (Y_1, D_1, D_2) are a function of the instruments determined by the following triangular system:

$$\begin{aligned}
D_2 &= 1\{\pi_2(Y_1, D_1, Z_2, X) \geq V_2\} \\
Y_1 &= 1\{\mu_1(D_1, X) \geq U_1\} \\
D_1 &= 1\{\pi_1(Z_1, X) \geq V_1\}.
\end{aligned}$$

Therefore, Assumption 1 follows from $U_2 \perp (Z_1, Z_2)|X, U_1, V_1, V_2$, which in turn is implied by $(U_1, U_2, V_1, V_2) \perp (Z_1, Z_2)|X$. The sets $\boldsymbol{V}_{U_1}, \boldsymbol{V}_1, \boldsymbol{V}_2$ are defined by inverting the triangular system above with respect to (U_1, V_1, V_2), which ensures Assumption 2 (i). Assumption 2 (ii) also holds because these sets are functions of (D, Z, X) and π. Assumption 3 holds by hypothesis and noting that $U = U_2$ and $V = (U_1, V_1, V_2)$ in this example.

By Theorem 1, $\theta \in \Theta_I(P_0)$ iff

$$P_0(Y = 1|D = d, X = x, Z = z) \geq \mathbb{C}_\theta(\{1\}|D = d, X = x, Z = z) \tag{B.18}$$

$$P_0(Y = 0|D = d, X = x, Z = z) \geq \mathbb{C}_\theta(\{0\}|D = d, X = x, Z = z). \tag{B.19}$$

By (5.25) (and as argued in the text),

$$\mathbb{C}_\theta(\{1\}|D = d, X = x, Z = z) = \inf_{v \in \boldsymbol{V}(d,x,z;\pi)} H(d, x, v) \tag{B.20}$$

$$\mathbb{C}_\theta(\{0\}|D = d, X = x, Z = z) = 1 - \sup_{v \in \boldsymbol{V}(d,x,z;\pi)} H(d, x, v) \tag{B.21}$$

The identifying restriction (5.26) follows from (B.18)-(B.21) and noting that $P_0(Y = 0|D = d, X = x, Z = z) = 1 - P_0(Y = 1|D = d, X = x, Z = z)$.

The identifying restrictions (5.27)-(5.28) follow from applying the same argument sequentially. For example, letting $Y = D_2$, $D = (Y_1, D_1)$, $U = V_2$, and $V = (U_1, V_1)$ and applying the argument above yields (5.27). $\square$