Set-Valued Control Functions

Sukjin Han & Hiro Kaido

University of Bristol & Boston University

September 2024

Endogeneity and Control Function Approach

Control function approach has been a valuable tool...

- in addressing endogeneity and recovering various causal parameters,
- esp. for nonparametric models that allow for heterogeneity

Control Function Approach

Construct control variables V, which defines a latent type

- conditional on V, endogenous explanatory variables D is unconfounded

Often, V is constructed by inverting treatment selection processes

- so that V is written as a function of observables
- thus a control function (CF)

Many empirical studies build on this insight to construct CF

- Kline & Walters 16; Card et al 19; Abdulkadiroglu et al 20; Bishop et al 22...

Challenges of Control Function Approach

Challenge: CF approach relies on invertibility of selection models

- e.g., in nonparametric triangular model, it requires...
  - continuous D and
  - strict monotonicity w.r.t scalar unobservable

$\Rightarrow$ most important limitation of CF approach (Blundell & Powell 03)

Challenges of Control Function Approach

Challenge: CF approach relies on invertibility of selection models

- e.g., in nonparametric triangular model, it requires...
  - continuous D and
  - strict monotonicity w.r.t scalar unobservable

$\Rightarrow$ most important limitation of CF approach (Blundell & Powell 03)

Challenge: CF is required to be "point identified"

- sometimes does not hold
- e.g., interval data, controls involving strategic behaviors

This Paper: Set-Valued Control Functions

This paper: allows the CF to be set-valued

- when only coarse information of controls is available
- e.g., selection models without invertibility

$\Rightarrow$ expands the CF approach to broader applications

This Paper: General Selection Processes

We allow complex treatment selection processes, such as...

1. continuous or discrete decisions with rich heterogeneity
2. censored decisions
3. strategic interaction of multiple agents
4. dynamically optimizing behavior

$\Rightarrow$ these processes typically violate invertibility

- mapping from observables to V is only a correspondence

This Paper: Partially Observed Controls without Selection

We also allow control variables that are partially observed/identified without any selection process

5. controls being interval data (e.g., wealth, debt, biometric measures, psychological traits)
6. strategically reported preference in school matching (Bertanha et al 24)
7. link information as controls (Auerbach 22)

This Paper: Partially Observed Controls without Selection

We also allow control variables that are partially observed/identified without any selection process

5. controls being interval data (e.g., wealth, debt, biometric measures, psychological traits)
6. strategically reported preference in school matching (Bertanha et al 24)
7. link information as controls (Auerbach 22)

We show that the CF approach can still be used...

- to (partially) identify structural (i.e., causal) parameters
- e.g., average and quantile structural functions for outcomes

Outcome Equation

Consider an outcome equation:

$$Y = \mu(D, U)$$

- $Y \in \mathcal{Y} \subseteq \mathbb{R}^{d_Y}$ outcome of interest
- $D \in \mathcal{D} \subseteq \mathbb{R}^{d_D}$ vector of endogenous treatments
- $U \in \mathcal{U} \subseteq \mathbb{R}^{d_U}$ vector of latent variables
- $(X \in \mathcal{X} \subseteq \mathbb{R}^{d_X}$ vector of covariates, suppressed)
- μ structural function

Causal Parameters

Consider a potential outcome:

$$Y_d = \mu(d, U)$$

- many policy-relevant parameters are features of Y_d
  - hence functionals of μ
- e.g., the *average structural function* and the *distributional structural function*:

$$\begin{aligned} \text{ASF}(d) &\equiv E[\mu(d, U)] = E[Y_d] \\ \text{DSF}(d) &\equiv F_{\mu(d,U)} = F_{Y_d} \end{aligned}$$

Control Function Assumption

A vector of control variables $V \in \mathcal{V}$ (e.g. $\subseteq \mathbb{R}^{d_V}$) is such that

$$D \perp U | V$$

For CF approach to work, V needs to be identified or expressed as a function of observables

- when selection is involved, let $Z \in \mathcal{Z} \subseteq \mathbb{R}^{d_Z}$ be vector of IVs
- Newey et al 99: $D = \pi(Z) + V$, then $V = D - \pi(Z)$
- Imbens & Newey 09: $D = h(Z, V)$ with continuous *scalar* V and h strictly monotonic in V, then $V = h^{-1}(Z, D)$
  - when V is continuous, invertibility requires D to be continuous
  - scalar V limits heterogeneity in selection mechanism

$\Rightarrow$ We aim to remove these restrictions

Generalized Selection Equation

Consider generalized selection equation:

$$D = \pi(Z, V)$$

- (X suppressed)
- in general, $(D, Z) \mapsto V$ is only a correspondence
- e.g., $D = 1\{\pi(Z) \geq V\}$

The selection process restricts V to the following set a.s.:

$$\{v : D = \pi(Z, v)\} \subseteq \mathbb{R}^{d_V}$$

This Paper: Control Function as Random Set

A set-valued CF $\boldsymbol{V}$ is a *random closed set*, constructed from observable variables

$$\boldsymbol{V}(D, Z; \pi) = \text{cl}\{v : D = \pi(Z, v)\} \subseteq \mathbb{R}^{d_V}$$

- it contains the true control variable V a.s.
- it can be used to construct a *set-valued predictions* of outcome that are compatible with the model
- then, using the *containment functional* or *Aumann expectation* (Molchanov 17) associated with the set, we generate sharp identifying restrictions,
- which then yield the (sharp) identified set for structural parameters

Related Literature

Identification and estimation in nonparametric models with endogenous explanatory variables:

- nonparametric IV approach: Newey & Powell 03; Hall & Horowitz 05; Chernozhukov & Hansen 05; Darolles et al 11; D'Haultfoeuille & Fevrier 15; Torgovitsky 15; Vuong & Xu 17; Chen & Christensen 18
- nonparametric CF approach: Newey et al 99; Chesher 03; Das et al 03; Blundell & Powell 04; Imbens & Newey 09; D'Haultfoeuille et al 21; Newey & Stouli 21
- monotonicity assumption with binary or discrete D: Imbens & Angrist 94; Abadie et al 02; Heckman & Vytlacil 05

Related Literature

Partial identification:

- generalization of IV approach: Chesher & Rosen 17; Chesher & Smolinski 12; Chesher & Rosen 13; Chesher et al 23
- related approaches: Beresteanu et al 11; Galichon & Henry 11
- partial identification without invertibility in selection: Chesher 05; Shaikh & Vytlacil 11; Jun et al 11; Mourifie 15; Mogstad et al 18; Machado et al 19; Han & Yang 24
- interval data: Manski & Tamer 02; Molinari 20

This paper:

- different way of applying random set theory
- generalization of CF approach
  - CF and IV assumptions are non-nested
- a wide range of models where controls are partially identified

I. Motivating Examples

Example 1: Generalized Roy Model

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

can be motivated by

$$\begin{aligned} D &= 1\{Y_1 - Y_0 - C \geq 0\} \\ Y_d &= \mu(d) + U_d \quad \text{for } d = 0, 1 \\ C &= \mu_c(Z) + U_c \end{aligned}$$

- Z a vector of cost-shifters
- $\pi(Z) \equiv \mu(1) - \mu(0) - \mu_c(Z)$ and $V \equiv U_c - U_1 + U_0$

Note $U \equiv (U_1, U_0)$ is a vector in $Y = \mu(D, U)$

Example 1: Generalized Roy Model

Suppose we are interested in the causal effect of D on Y

Suppose $Z \perp U|V$

Then, V is valid control variable, because $D \perp U|V$

Example 1: Generalized Roy Model

Suppose we are interested in the causal effect of D on Y

Suppose $Z \perp U | V$

Then, V is valid control variable, because $D \perp U | V$

We cannot recover V by inverting the selection equation

Nonetheless, the model restricts V to the following set a.s.:

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

which is a set-valued CF

Example 2: Non-Monotonic Treatment Decisions

Example 1 satisfies LATE monotonicity, eliminating either defiers or compliers (Imbens & Angrist 94; Vytlacil 02)

Consider instead

$$D_z = 1\{\pi(z) \geq V_z\} \quad \text{for } z \in \mathcal{Z}$$

- suppose Z is binary
- both compliers and defiers can have nonzero shares:

$$\{D_0 = 0, D_1 = 1\} = \{V_0 > \pi(0), V_1 \leq \pi(1)\}$$
$$\{D_0 = 1, D_1 = 0\} = \{V_0 \leq \pi(0), V_1 > \pi(1)\}$$

Example 2: Non-Monotonic Treatment Decisions

Observed $D = D_0 + (D_1 - D_0)Z$ satisfies

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

where $\tilde{\pi}(Z) \equiv \pi(0) + Z(\pi(1) - \pi(0))$

- a *random-coefficient* model for selection (Gautier & Hoderlein 11; Kline & Walters 19)

Suppose $Z \perp U | (V_0, V_1)$, then $V \equiv (V_0, V_1)$ are valid control variables

V belongs to the following set-valued CF a.s.:

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

Example 2: Non-Monotonic Treatment Decisions

For continuous D, consider a random coefficient model:

$$D = V_0 + V_1 Z$$

Then $V \equiv (V_0, V_1)$ belongs to the following set-valued CF a.s.:

$$\boldsymbol{V}(D, Z) = \mathrm{cl}\left\{(v_0, v_1) : v_0 = D - Z v_1\right\}$$

Example 3: Decisions as Corner Solutions

Latent treatment: $D^* = \pi^*(Z) + V$

Observed treatment: $D = \max\{D^*, 0\}$

- e.g., hours of training, amount of subsidy
- then $D = \pi(Z, V) \equiv \max\{\pi^*(Z) + V, 0\}$

Fix z, then

- if $d = 0$, it must be that $\pi^*(z) + V \leq 0$
- if $d > 0$, it must be that $\pi^*(z) + V > 0$

Then, we have

$$\boldsymbol{V}(D, Z; \pi^*) = \begin{cases} [-\pi^*(Z), \infty) & \text{if } D > 0 \\ (-\infty, -\pi^*(Z)] & \text{if } D = 0 \end{cases}$$

Example 4: Strategic or Dynamic Treatment Decisions

Let D be *vector* of decisions (across individuals or periods)

We are interested in the effect of the entire profile D on Y

- treatment effects with strategic interaction (Balat & Han 23)
- dynamic treatment effects (Han 21; Han 24; Han & Lee 24)

Suppose Z is vector of IVs (individual- or time- specific) and $\pi(\cdot)$ is the generalized selection function

We can construct $\boldsymbol{V}(D, Z; \pi)$ in multi-dimensional space (more later)

Example 5: Set-Valued Controls Without Selection

Bertanha, Luflade & Mourifié 24 estimate the causal effects of school assignment

- students' local preferences as control variables, then RD comparison
- i.e., $V \in \mathcal{V}$ where $\mathcal{V}$ is the set of preference relations
- but under capacity constraints, students have incentives to misreport their preferences
- based on reported partial order of preferences, they recover local preference sets ($\boldsymbol{V}$) that contain V a.s.

Example 5: Set-Valued Controls Without Selection

In social network setting, Auerbach 22 considers a partial linear model for an outcome

- with nonparametric $\lambda(V)$ where V is an unknown control variable (e.g., social characteristics)
  - V is seldom identified
- instead, use the link function $f(\cdot)$ in a link formation model
  - f is identified from the distribution of social links
- Assumption 3: individuals with similar f have similar $\lambda(V)$
- then, the linear parameters are identified
- want to relax Assumption 3: individuals with similar f have values of $\lambda(V)$ with discrepancy bounded by M
- then, we can recover a set of controls $(\lambda(\boldsymbol{V}_M))$

II. Model Predictions and Identification Analysis

Main Assumptions

Assumption 1 (CF)

$U|D,V \sim U|V$.

Assumption 2 (Set-Valued CF)

(i) There is a random closed set $\boldsymbol{V} : \Omega \to \mathcal{F}(\mathcal{V})$ such that $V \in \boldsymbol{V}$ with prob 1; (ii) $\boldsymbol{V}$ is a measurable function of observable variables and a parameter π.

Assumption 3 (Continuous U)

$U|D,V$ has strictly positive density w.r.t. Lebesgue measure on $\mathbb{R}^{d_U}$ a.s.

Main Assumptions

By Assumption 3, one may represent

$$U = Q(\eta; D, V) \in \mathbb{R}^{d_U}$$

- random vector $\eta \in \mathbb{R}^{d_U}$ where $\eta \perp (D, V)$ and $\eta \sim U[0,1]^{d_U}$
- Knothe-Rosenblatt transform (Villani 08; Carlier et al 10; Joe 14)
- e.g., we can represent $U = (U_0, U_1) \sim F_{U|D,V}$ sequentially as

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

where $(\eta_0, \eta_1) \sim U[0,1]^2$

Main Assumptions

$$U = Q(\eta; D, V) \in \mathbb{R}^{d_U}$$

Then, by Assumption 1,

$$U = Q(\eta; V)$$

- Q is a *known* function of distribution F on $U|V$
  - if U is scalar, Q is conditional quantile of $U|V$

Model's Prediction

Now we can write

$$Y = \mu(D, U) = \mu(D, Q(\eta; V))$$

- $Q(\eta; V)$ is adjustment term
  - involves V and "clean" error term η (independent of D)

Consider model's prediction given parameter $\theta \equiv (\mu, F, \pi)$:

- define a random closed set

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

- this set collects all Y values compatible with the model
- a function of *observable exogenous* $(D, \boldsymbol{V})$ and latent η

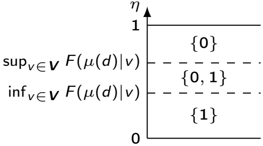
Illustrative Example

Consider binary Y and

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

Model's prediction given (μ, F):

$$\begin{aligned} &\boldsymbol{Y}(\eta, D, \boldsymbol{V}; \mu, F) \\ &= \begin{cases} \{0\} & \eta > \sup_{v \in \boldsymbol{V}} F(\mu(D)|v) \\ \{0,1\} & \inf_{v \in \boldsymbol{V}} F(\mu(D)|v) < \eta \leq \sup_{v \in \boldsymbol{V}} F(\mu(D)|v) \\ \{1\} & \eta \leq \inf_{v \in \boldsymbol{V}} F(\mu(D)|v) \end{cases} \end{aligned}$$



Identification Analysis

We aim to characterizes $\Theta_I(P_0)$ through inequality restrictions on θ

We introduce the *containment functional* $\mathbb{C}_\theta$ of random set $\boldsymbol{Y}$:

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

for any closed set $A \subset \mathcal{Y}$ and (d, z)

- $\mathbb{C}_\theta$ uniquely determines the distribution of $\boldsymbol{Y}$ (Molchanov 17)
- conditional on (D, Z), the remaining randomness in $\boldsymbol{Y}$ is η
- it is straightforward to compute the right-hand side, as η is uniform over $[0, 1]^{d_U}$ independent of (D, Z)

Illustrative Example (continued)

$$\boldsymbol{Y}(\eta, D, \boldsymbol{V}; \mu, F) = \begin{cases} \{0\} & \eta > \sup_{v \in \boldsymbol{V}} F(\mu(D)|v) \\ \{0,1\} & \inf_{v \in \boldsymbol{V}} F(\mu(D)|v) < \eta \leq \sup_{v \in \boldsymbol{V}} F(\mu(D)|v) \\ \{1\} & \eta \leq \inf_{v \in \boldsymbol{V}} F(\mu(D)|v) \end{cases}$$

For example, for $A = \{1\}$,

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

and, for $A = \{0\}$,

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

Identification Analysis

Theorem 1 (Identified Set)

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

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

- above restrictions are known as *Artstein's inequalities* (Molchanov & Molinari 18)
- model's set-valued prediction $\Rightarrow$ a system of inequalities that do not involve unobservable V
- thus amenable to estimation:
  - $P_0(A|D, Z)$ can be recovered from sample of (Y, D, Z)
  - $\mathbb{C}_\theta(A|D, Z)$ can be computed from model primitives

Identification Analysis with Mean Restrictions

Suppose Y is continuous with

$$Y_d = \mu(d) + U_d$$

Then, instead of Assumption 1 (CF), we can assume the following:

Assumption 1$'$ (Mean CF)

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

Let $\lambda_d(V) \equiv E[U_d|V]$ and $\eta_d \equiv U_d - E[U_d|V]$

Under Assumption $1^{'}$, we may write

$$E[Y|D = d, V = v] = \mu(d) + \lambda_d(v)$$

and $Y = \mu(d) + \lambda_d(v) + \eta_d$

- λ_d is a known function of F (the distribution of $U|V$)

Identification Analysis with Mean Restrictions

With $\eta \equiv (\eta_d, d \in \mathcal{D})$, define

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

Then use *Aumann expectations* and *support functions* to derive:

Theorem 2 (Identified Set)

Suppose Assumptions $1'$, 2, 3 hold. Suppose $E_{P_0}[|Y|] < \infty$. Then, the sharp identification region for the structural parameter is

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

where

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

Identification Analysis: Functionals of θ

Based on $\Theta_I(P_0)$ of θ obtained in Theorems 1 and 2, we can construct bounds on functionals of θ

Identification Analysis: Functionals of θ

Based on $\Theta_I(P_0)$ of θ obtained in Theorems 1 and 2, we can construct bounds on functionals of θ

Structural estimands can be obtained as functionals of (μ, F, F_V):

- e.g., $\text{ASF}(d) \equiv E[\mu(d, U)] = E[Y_d]$ (Blundell & Powell 03):

$$\text{ASF}(d) = \int \int \mu(d, Q(\eta; v)) d\eta dF_V$$

  - $\text{ATE}(d, d') = \text{ASF}(d) - \text{ASF}(d')$

- e.g., $DSF(y, d) \equiv F_{\mu(d,U)} = F_{Y_d}$ (Chernozhukov et al 20):

$$\text{DSF}(y, d) = \int \int 1\{\mu(d, Q(\eta; v)) \leq y\} d\eta dF_V$$

  - $\text{QSF}(d) = \text{DSF}^{-1}(\tau, d)$ (Imbens & Newey 02)

Identification Analysis: Functionals of θ

- e.g., *policy-relevant structural function*:

$$\kappa(z) \equiv E[Y_{D_z}] = \int\int \mu(\pi(z,v), Q(\eta;v))d\eta dF_V$$

- e.g., *mediated structural function*:

$$\kappa(d_1, d_1') \equiv E[Y_{d_1, D_{2,d_1'}}] = \int\int \mu(d_1, \pi_2(d_1', z, v), Q(\eta;v))d\eta dF_{Z,V}$$

where we allow $d_1 \neq d_1'$

Identification Analysis: Functionals of θ

In general, given a function $\varphi : \mathbb{R} \to \mathbb{R}$, let

$$\kappa(d) \equiv E[\varphi(Y_d)] = \int\int \varphi(\mu(d, Q(\eta; v))d\eta dF_V$$

- ASF and DSF are special cases of κ

Theorem 3 (Identified Set)

Suppose the conditions of Theorem 1 or 2 hold. Then, the sharp identification region for κ is $\mathfrak{K}_I(d) \equiv \bigcup_{\theta\in\Theta_I(P_0)}[\underline{\kappa}(d;\theta), \overline{\kappa}(d;\theta)]$, where

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

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

Discussion: Inference

Practitioners can use the restrictions in Theorem 1 or 2 to make inference for $\Theta_I(P_0)$, its elements, or $\mathfrak{K}_I(d)$

- inference methods based on conditional moment inequalities (Andrews & Shi 13; Chernozhukov et al 13)
- likelihood-based inference methods (Chen et al 18; Kaido & Molinari 22; Kaido & Zhang 24)

III. Applications of the Identification Results

Example 4.1: Treatment Responses with Social Interactions

Consider individuals $j = 1, \ldots, J$

Let $D \equiv (D_1, \ldots, D_J)$ be vector of decisions across individuals

We are interested in the effect of the entire profile D on Y

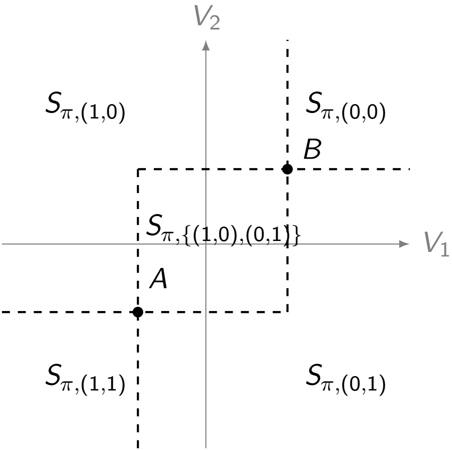
Suppose that observed D satisfies

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

where D_{-j} is vector D without D_j

- Balat & Han 23; Ciliberto, Murry & Tamer 21
- can be motivated by relaxing SUTVA (Rubin 78) and introducing Roy-type decisions
- multiple solutions to (1) may exist
  - the selection process is *incomplete* (Tamer 03)

Example 4.1: Treatment Responses with Social Interactions



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

Example 4.1: Treatment Responses with Social Interactions

Let $V_s : \Omega \to \{0,1\}$ represent an unknown *selection mechanism*

- if $(V_1, V_2) \in S_{\pi,\{(1,0),(0,1)\}}(Z)$,...
  - $D = (1,0)$ is selected when $V_s = 1$, and
  - $D = (0,1)$ is selected when $V_s = 0$
- V_s is another source of possible endogeneity

Suppose $E[U|Z,V] = E[U|V]$ for $V \equiv (V_1, V_2, V_s)$, then V is valid control variable

Example 4.1: Treatment Responses with Social Interactions

We define the following set-valued CF as a union of two random sets:

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

where

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

and

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

Example 4.1: Treatment Responses with Social Interactions

For fixed j,

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

Recall $D \equiv (D_1, D_2)$ and define the model prediction

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

where $\lambda_d(v) \equiv E[U_d|V]$

Example 4.1: Treatment Responses with Social Interactions

Corollary 2 (Identified Set)

Suppose $E_{P_0}[|Y|] < \infty$. Suppose, for each $(d_1, d_2) \in \mathcal{D}$, $E[U_{d_1,d_2}|Z, V] = E[U_{d_1,d_2}|V]$, *a.s.* Then, $\Theta_I(P_0)$ is the set of values $\theta = (\mu, \pi, F)$ such that

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

where

$$\lambda_U(d, z) \equiv \max\{ \sup_{(v_1,v_2)\in\tilde{\boldsymbol{V}}_0(d,z;\pi)} \lambda_d(v_1, v_2, 0), \sup_{(v_1,v_2)\in\tilde{\boldsymbol{V}}_1(d,z;\pi)} \lambda_d(v_1, v_2, 1)\},$$

$$\lambda_L(d, z) \equiv \min\{ \inf_{(v_1,v_2)\in\tilde{\boldsymbol{V}}_0(d,z;\pi)} \lambda_d(v_1, v_2, 0), \inf_{(v_1,v_2)\in\tilde{\boldsymbol{V}}_1(d,z;\pi)} \lambda_d(v_1, v_2, 1)\}.$$

Example 4.2: Dynamic Treatment Effects

Consider

$$
\begin{aligned}
D_1 &= 1\{\pi_1(Z_1) \geq V_1\} \\
Y_1 &= 1\{\mu_1(D_1) \geq U_1\} \\
D_2 &= 1\{\pi_2(Y_1, D_1, Z_2) \geq V_2\} \\
Y_2 &= 1\{\mu_2(Y_1, D_1, D_2) \geq U_2\}
\end{aligned}
$$

- Han 21; Han 24; Han & Lee 24

Focus on the effect of $D \equiv (Y_1, D_1, D_2)$ on Y_2

- recovering the effect is not straightforward
- U_2 may depend on (U_1, V_1, V_2)
- e.g., U_1 and U_2 may share a time invariant component
- e.g., U_2 may be related to (V_1, V_2) through the agent's dynamic treatment decisions

Example 4.2: Dynamic Treatment Effects

Recall $D \equiv (Y_1, D_1, D_2)$ and let $Z \equiv (Z_1, Z_2)$

Let $U \equiv U_2$ and $V \equiv (U_1, V_1, V_2)$

Can show $D \perp U | V$ if $Z \perp U | V$

Let $\pi \equiv (\mu_1(\cdot), \pi_1(\cdot), \pi_2(\cdot))$

Can construct the following set-valued control function:

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

where

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

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

Then, construct $\boldsymbol{Y}$ and apply Theorem 1 (see the paper)

IV. Impact of Learning HIV Status

Impact of Learning HIV Status

Revisit Thornton 08 who studies the impacts of learning HIV status in Malawi by using RCT data and estimating LATE

- binary D: learning HIV status and receiving counseling
- ordered Y: HIV preventive behavior (condom purchases)
- IVs Z: voucher amount and distance to test center
- discrete/continuous X: HIV diagnosis, gender, age, district, simulated distance to center

Impact of Learning HIV Status

Consider an ordered choice model of condom purchases:

$$Y = \begin{cases} 0 & \text{if } \mu(D, X) + U \leq c_L \\ 3 & \text{if } c_L < \mu(D, X) + U \leq c_U \\ 6 & \text{if } \mu(D, X) + U > c_U \end{cases}$$

and a model for selection: $D = 1\{\pi(Z, X) \geq V\}$

Proposition 1 (Identified Set)

Suppose Assumptions 1–3 hold. Then, $\theta = (\mu, c_L, c_U, F, \pi)$ is in the sharp identification region $\Theta_I(P_0)$ if and only if $\pi \in \Pi_r(P_0)$ and

$$\mathbb{C}_\theta(\{0\}|D, X, Z) \leq P_0(Y = 0|D, X, Z) \leq \mathbb{C}^*_\theta(\{0\}|D, X, Z)$$
$$\mathbb{C}_\theta(\{6\}|D, X, Z) \leq P_0(Y = 6|D, X, Z) \leq \mathbb{C}^*_\theta(\{6\}|D, X, Z), \ a.s.$$

Impact of Learning HIV Status

Our objects of interest:

- (conditional) average structural function:

$$ASF(d, x_{HIV}) = E[Y_d | x_{HIV}]$$

- (conditional) switching probability:

$$P(Y_0 = 0, Y_1 > 0 | x_{HIV})$$

  i.e., the share of "switchers" induced by intervention

- (in progress) (conditional) policy-relevant structural function:

$$PRSF(z^*_{amt}, x_{HIV}) = E[Y_{D_{z^*_{amt}}} | x_{HIV}]$$

  by giving \$3 ($z^*_{amt} = 3$) to encourage them to learn HIV status

Impact of Learning HIV Status

Additional identifying assumptions:

- MTS (Manski & Pepper 00): For each $d = 0, 1$,

$$E[Y(d)|D = 1] \geq E[Y(d)|D = 0]$$

  - those who choose to learn their HIV status are more likely to buy condoms (e.g., health-conscious individuals)

- MTR (Manski 90, 97) for the HIV- group:

$$Y(1)|X_{HIV} = 0 \geq Y(0)|X_{HIV} = 0$$

  - HIV- group may have stronger incentive for preventive behavior (Thornton 08)

Impact of Learning HIV Status

		HIV+	HIV-
Baseline			
	ASF(1)	[0.030, 5.578]	[0.090, 2.864]
	ASF(0)	[0.030, 5.216]	[0.151, 3.256]
	ATE	[-4.794, 4.492]	[-2.563, 2.141]
MTS			
	ASF(1)	[0.030, 5.307]	[0.090, 1.779]
	ASF(0)	[0.030, 5.276]	[0.181, 3.286]
	ATE	[-4.854, 3.889]	[-2.563, 1.176]

95% CIs are calculated based on Kaido & Zhang 24

Impact of Learning HIV Status

		HIV+	HIV-
MTR for HIV-			
	ASF(1)	[0.060, 5.397]	[0.271, 2.683]
	ASF(0)	[0.030, 4.070]	[0.151, 1.568]
	ATE	[-2.985, 4.372]	[0.030, 2.020]
MTS & MTR for HIV-			
	ASF(1)	[0.030, 5.276]	[0.241, 1.719]
	ASF(0)	[0.030, 4.281]	[0.181, 1.719]
	ATE	[-3.045, 3.950]	[0.030, 1.176]

95% CIs are calculated based on Kaido & Zhang 24

Impact of Learning HIV Status

		HIV+	HIV-
Baseline			
	Switch Pr.	[0, 0.829]	[0, 0.508]
MTS			
	Switch Pr.	[0, 0.724]	[0, 0.291]
MTR for HIV-			
	Switch Pr.	[0, 0.824]	[0, 0.492]
MTS & MTR for HIV-			
	Switch Pr.	[0, 0.729]	[0, 0.286]

95% CIs are calculated based on Kaido & Zhang 24

V. Conclusions

Concluding Remarks

Allowing control functions to be set-valued, this paper expands the scope of the control function approach

We accommodate selection processes that involve...

- rich heterogeneity,
- dynamic optimizing behavior,
- social interaction, or
- censoring,...
- and cases without selection process

We derive sharp identifying restrictions...

- that are inequalities on the conditional choice probabilities

Concluding Remarks

Practitioners can use the results of this paper for various purposes:

1. evaluate social programs nonparametrically, when only coarse information of controls is available
2. conduct a sensitivity analysis to assess identifying power of specific assumptions (e.g., shape restrictions)
   - random set theory guarantees sharpness of bounds, without needing to prove sharpness case after case

More empirical results to come...

- random coefficient model for selection

Thank You! ☺