Estimating Causal Effects of Discrete and Continuous Treatments with Binary Instruments

Victor Chernozhukov, Iván Fernández-Val,
Sukjin Han, Kaspar Wüthrich

MIT, Boston U, U of Bristol, UC San Diego

November 2024

Endogeneity and Heterogeneity

Endogeneity and heterogeneity are key challenges in causal inference

- accounting for them in estimating treatment effects is crucial to answer policy questions
- e.g. how to allocate social resources and combat inequalities

This paper proposes a flexible IV modeling framework for identifying heterogeneous treatment effects under endogeneity

- that yields straightforward semiparametric estimation and inference procedures

Example: Effects of Sleep on Well-Being

- Y: well-being index of workers in a developing country
- D: sleep hours per night
- Y_d: counterfactual well-being with sleep level d
  - causal object of interest (e.g., $\partial E[Y_d]/\partial d$)
- Z: randomly assigned sleep support from RCT
  - affects D but independent of Y_d
- D_z: counterfactual sleep level with assignment z
- X: observed characteristics of worker (e.g., gender, age)

Example: Effects of Sleep on Well-Being

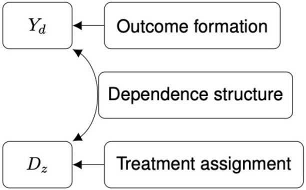
- Y: well-being index of workers in a developing country
- D: sleep hours per night
- Y_d: counterfactual well-being with sleep level d
  - causal object of interest (e.g., $\partial E[Y_d]/\partial d$)
- Z: randomly assigned sleep support from RCT
  - affects D but independent of Y_d
- D_z: counterfactual sleep level with assignment z
- X: observed characteristics of worker (e.g., gender, age)

$\Rightarrow$ D is endogenous (e.g., underlying health conditions)

- Y_d and D_z are dependent, even after controlling for X

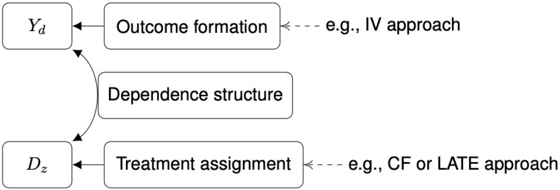
Previous Approaches

IV alone cannot point-ID meaningful treatment effects



Previous Approaches

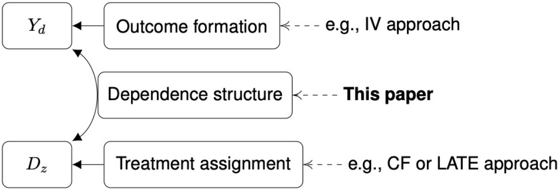
IV alone cannot point-ID meaningful treatment effects



Two approaches:

1. Restricting structure/heterogeneity of potential outcomes
   - IV approach: Ai & Chen 03; Newey & Powell 03; Chernozhukov & Hansen 05; Blundell, Chen & Kristensen 07; Vuong & Xu 17
2. Restricting structure/heterogeneity of treatment assignment
   - CF approach: Newey et al 99; Imbens & Newey 09
   - LATE/MTE approach: Imbens & Angrist 94; Heckman & Vytlacil 05

This Paper's Approach



We explore an intermediate route:

$\Rightarrow$ imposing structure on relationship between treatment assignment and potential outcomes

- achieve point ID of various heterogeneous treatment effects

This Paper: Local Copula Representation

Basis of our approach: Local Gaussian Representation (LGR)

- *copula* representing the joint distribution of the potential outcomes Y_d and treatment assignment unobservables D_z
- this representation is fully *nonparametric* (Chernozhukov, Fernández-Val & Luo 24)
  - by treating the correlation parameter as an implicit function
  - not require (Y_d, D_z) being jointly or marginally Gaussian

Use this representation to introduce an assumption that has not been previously considered for ID of treatment effects:

- *copula invariance*
- restricts the shape of local dependence

This Paper: Expands Modeling Trade-Offs

We show that, even with a binary IV, copula invariance identifies...

- quantile and average treatment effects (QTE and ATE) of binary and ordered treatments
- quantile and average structural functions (QSF and ASF) of continuous treatment

We expand the directions of modeling trade-offs:

- compared to IV, CF, LATE approaches...
- we impose more restrictions on the dependence structure (i.e., the form of endogeneity),
- while allowing for richer patterns of effect heterogeneity

Our identification strategy is constructive

- leads to simple semiparametric estimation procedures

Related Literature

Identification and estimation in nonparametric models with endogenous explanatory variables:

- nonparametric IV approach: Ai & Chen 03; Newey & Powell 03; Hall & Horowitz 05; Chernozhukov & Hansen 05; Blundell, Chen & Kristensen 07; Chen & Pouzo 09, 12, 15; Vuong & Xu 17; Chen & Christensen 18
- nonparametric CF approach: Newey, Powell & Vella 99; Das, Newey & Vella 03; Blundell & Powell 04; Imbens & Newey 09; D'Haultfoeuille, Hoderlein & Sasaki 21; Newey & Stouli 21
- related approaches:
  - Chesher 03
  - D'Haultfoeuille & Février 15; Torgovitsky 15
- monotonicity assumption with binary or discrete D: Imbens & Angrist 94; Abadie et al 02; Heckman & Vytlacil 05

Related Literature

Copula for identification and estimation in econometrics:

- Chen et al 06: semiparametric copula models for distributions
- Han & Vytlacil 17; Han & Lee 19: a class of single-parameter copulas to model endogeneity for binary outcome & treatment
- Han & Lee 24: dynamic treatment effect models using copula
- Chen & Fan 06a, b; Chen et al 22, Chen et al 24; Ghanem, Kédagni & Mourifié 24: use of copula in TS and DiD settings
- Arellano & Bonhomme 17: real analytical copula and continuous instrument in sample selection model
- Chernozhukov, Fernández-Val & Luo 24: use of LGR in sample selection model

Related Literature

$\Rightarrow$ **this paper**:

- two-way sample selection in the binary treatment case
- general selection model without threshold-crossing
- completely new results with ordered and continuous treatments

I. Setup and Assumptions

Variables

$Y \in \mathcal{Y} \subseteq \mathbb{R}$ scalar outcome (continuous, discrete or mixed)

$D \in \mathcal{D} \subseteq \mathbb{R}$ scalar treatment

- $\mathcal{D} = \{0, 1\}$ for binary D
- $\mathcal{D} = \{1, \ldots, K\}$ for ordered D
- $\mathcal{D}$ uncountable for continuous D

$Z \in \{0, 1\}$ binary IV

- most challenging case; extends to discrete or continuous Z

Y_d potential outcome given $d \in \mathcal{D}$; and $Y = Y_D$

D_z potential treatment given $z \in \{0, 1\}$; and $D = D_Z$

$X \in \mathcal{X} \subseteq \mathbb{R}^{d_x}$ vector of covariates (explicit in estimation)

Generalized Treatment Equation

General treatment assignment equation:

$$D_z = h(z, V_z)$$

- $V_z \sim U[0, 1]$ as normalization (assuming h is weakly monotonic)
- permit D to be a function of *vector* (V_0, V_1)
- (even this is not necessary but simplifies the exposition)

Parameters of Interest

Interested in identifying F_{Y_d} for $d \in \mathcal{D}$ and functionals of F_{Y_d}

- quantile and average structural functions:

$$QSF_\tau(d) \equiv Q_{Y_d}(\tau) = \mathcal{Q}_\tau(F_{Y_d}),$$
$$ASF(d) \equiv E[Y_d] = \mathcal{E}(F_{Y_d}),$$

- $QSF_\tau(d) - QSF_\tau(d')$ and $ASF(d) - ASF(d')$ for binary or ordered treatment
- $\partial QSF_\tau(d)/\partial d$ and $\partial ASF_\tau(d)/\partial d$ for continuous treatment

Local Gaussian Representation

Let $C(u_1, u_2; \rho)$ be Gaussian copula

Lemma (LGR) (Chernozhukov et al 24)

For any r.v.'s Y, V and Z, the joint distribution admits the representation:

$$F_{Y,V|Z}(y, v \mid z) = C(F_{Y|Z}(y \mid z), F_{V|Z}(v \mid z); \rho_{Y,V;Z}(y, v; z))$$

for all (y, v, z), where $\rho_{Y,V;Z}(y, v; z)$ is the unique solution in ρ to

$$F_{Y,V|Z}(y, v \mid z) = C(F_{Y|Z}(y \mid z), F_{V|Z}(v \mid z); \rho).$$

- Gaussianity is not essential for the local representation, but convenient
- other (comprehensive) copulas can be used for representation
  - e.g., Clayton copula, Frank copula, t copula

Assumptions

Assumption EX

For $d \in \mathcal{D}$ and $z \in \{0, 1\}$, $Z \perp\!\!\!\perp Y_d$ and $Z \perp\!\!\!\perp V_z$.

Assumption REL

(i) $Z \in \{0, 1\}$; (ii) $0 < \Pr(Z = 1) < 1$; and (iii) Z is relevant.

Assumption CI

For $d \in \mathcal{D}$, $\rho_{Y_d, V_z; Z}(y, v; z)$ is a constant function of (v, z), that is

$$\rho_{Y_d, V_z; Z}(y, v; z) = \rho_{Y_d}(y),$$

and $\rho_{Y_d}(y) \in (-1, 1)$.

- Under joint independence of Z and rank invariance in selection, CI holds if $\tilde{C}(u_1 \mid u_2) = C(u_1 \mid u_2; \rho(u_1))$ (more later)

Examples of Distributions under Copula Invariance

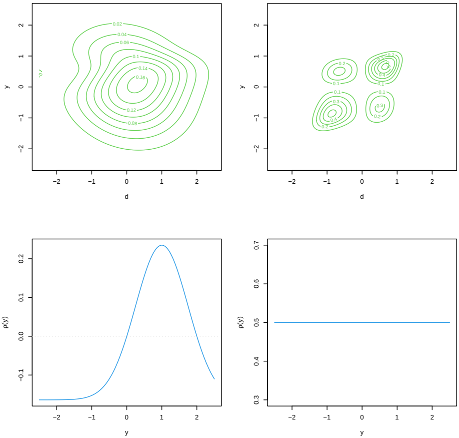


Figure: Joint Distributions under Copula Invariance

Notes: We depict joint distributions of (Y_d, V) under CI with Gaussian marginals (left) and nonparametric marginals (right).

II. Identification Analysis

Binary Treatment

Suppose $D \in \mathcal{D} = \{0, 1\}$ and consider

$$D_z = h(z, V_z) = 1\{V_z \leq \pi(z)\}$$

with propensity score (by EX)

$$\Pr[D = 1 \mid Z = z] = \Pr[D_z = 1 \mid Z = z] = \Pr[V_z \leq \pi(z)] = \pi(z)$$

- LATE monotonicity if $V_1 = V_0$

For ID analysis, consider

$$\begin{aligned}
\Pr[Y \leq y, D = 1 \mid Z = z] &= \Pr[Y_1 \leq y, D_z = 1 \mid Z = z] \\
&= C(F_{Y_1|Z}(y|z), \pi(z); \rho_{Y_1,V_z;Z}(y, \pi(z); z)) \\
&= C(F_{Y_1}(y), \pi(z); \rho_{Y_1}(y))
\end{aligned}$$

by LGR, EX and CI

Binary Treatment

By varying $Z \in \{0, 1\}$, a system nonlinear equations:

$$\Pr[Y \le y, D = 1 \mid Z = 0] = C(F_{Y_1}(y), \pi(0); \rho_{Y_1}(y))$$
$$\Pr[Y \le y, D = 1 \mid Z = 1] = C(F_{Y_1}(y), \pi(1); \rho_{Y_1}(y))$$

Then, the system has unique solution for $(F_{Y_1}(y), \rho_{Y_1}(y))$ by Gale & Nikaido 65's global univalence

▸ Gale & Nikaido 65

- because its Jacobian is P-matrix under REL

Theorem 1

Suppose $D_z = 1\{V_z \le \pi(z)\}$ for $z \in \{0, 1\}$. Under EX, REL and CI, the functions $y \mapsto F_{Y_d}(y)$ and $y \mapsto \rho_{Y_d}(y)$ are identified on $y \in \mathcal{Y}$ for $d \in \{0, 1\}$.

Ordered Treatment

Suppose $D \in \mathcal{D} = \{1, ..., K\}$ and consider

$$D_z = h(z, V_z) = \begin{cases} 1, & \pi_0(z) < V_z \leq \pi_1(z) \\ 2, & \pi_1(z) < V_z \leq \pi_2(z) \\ \vdots & \vdots \\ K, & \pi_{K-1}(z) < V_z \leq \pi_K(z) \end{cases}$$

where $\pi_0(z) = 0$ and $\pi_K(z) = 1$

- this model generalizes Heckman & Vytlacil 07 who consider

$$D_z = \begin{cases} 1, & \pi_0 < \mu(z) + V \leq \pi_1 \\ 2, & \pi_1 < \mu(z) + V \leq \pi_2 \\ \vdots & \vdots \\ K, & \pi_{K-1} < \mu(z) + V \leq \pi_K \end{cases}$$

Ordered Treatment

For ID analysis, consider

$$\begin{aligned}&\Pr[Y \leq y, D = d \mid Z = z] \\ &= \Pr[Y_d \leq y, \pi_{d-1}(z) < V_z \leq \pi_d(z) \mid Z = z] \\ &= C(F_{Y_d}(y), \pi_d(z); \rho_{Y_d}(y)) - C(F_{Y_d}(y), \pi_{d-1}(z); \rho_{Y_d}(y))\end{aligned}$$

by LGR, EX and CI

- for $d \in \{1, K\}$, REL identifies $F_{Y_d}(y)$ and ρ_{Y_d} (as before)
- but, for $d \in \mathcal{D} \backslash \{1, K\}$, Gale & Nikaido 65 doesn't apply

Ordered Treatment

To apply different global univalence, we assume:

Assumption U_{OC}

Either $F_{D|Z}(d \mid 0) > F_{D|Z}(d \mid 1)$ for all $d \in \mathcal{D} \backslash \{K\}$ or $F_{D|Z}(d \mid 0) < F_{D|Z}(d \mid 1)$ for all $d \in \mathcal{D} \backslash \{K\}$.

- U_{OC} is directly testable from data
- Heckman & Vytlacil 07's model satisfies U_{OC}
- when V_0 and V_1 are exchangeable, U_{OC} (with $>$) implies

$$\Pr[\text{ all complier groups }] > \Pr[\text{ all defier groups }]$$

  - cf. de Chaisemartin 17 with binary D
- when $V_0 = V_1$, U_{OC} (with $>$) implies

$$\Pr[\text{ all defier groups }] = 0$$

Ordered Treatment

Then, we apply the inverse theorem in Ambrosetti & Prodi 95 by showing...

1. the system has a unique solution when $\rho_{Y_d}(y) = 0$ (locally no endogeneity)
2. the function that defines the system is proper
3. the Jacobian has full-rank (by U_{OC})

▸ Ambrosetti & Prodi 95

Theorem 2

Suppose D_z, $z \in \{0, 1\}$, satisfies the ordered selection model. Under EX, REL, CI and U_{OC}, the functions $y \mapsto F_{Y_d}(y)$ and $y \mapsto \rho_{Y_d}(y)$ are identified on $y \in \mathcal{Y}$ for $d \in \mathcal{D}$.

Continuous Treatment

Suppose $D \in \mathcal{D} \subseteq \mathbb{R}$ and $F_{D|Z}(\cdot \mid z)$ is strictly increasing on $\mathcal{D}$

Consider

$$D_z = h(z, V_z) = F_{D|Z}^{-1}(V_z \mid z)$$

For ID analysis, consider

$$F_{Y|D,Z}(y \mid d, z) = F_{Y_d|D_z,Z}(y \mid d, z) = F_{Y_d|V_z,Z}(y \mid F_{D|Z}(d \mid z), z)$$

By LGR, EX and properties of cond'l CDF and Gaussian copula,

$$F_{Y_d|V_z,Z}(y \mid v, z) = \frac{(\partial/\partial v) F_{Y_d,V_z|Z}(y, v \mid z)}{(\partial/\partial v) F_{V_z|Z}(v \mid z)} = \Phi\left(\frac{\mu_{d,y} - \rho_{Y_d,V_z;Z}(y, v; z)\eta_v}{\sqrt{1 - \rho_{Y_d,V_z;Z}(y, v; z)^2}}\right)$$
$$+ \phi_2(\mu_{d,y}, \eta_v; \rho_{Y_d,V_z;Z}(y, v; z))(\partial/\partial v)\rho_{Y_d,V_z;Z}(y, v; z)$$

where $\mu_{d,y} \equiv \Phi^{-1}(F_{Y_d}(y))$ and $\eta_v \equiv \Phi^{-1}(v)$

Continuous Treatment

$$F_{Y|D,Z}(y \mid d,z) = \Phi\left(\frac{\mu_{d,y} - \rho_{Y_d,V_z;Z}(y, F_{D|Z}(d \mid z); z)\eta_v}{\sqrt{1 - \rho_{Y_d,V_z;Z}(y, F_{D|Z}(d \mid z); z)^2}}\right)$$
$$+ \phi_2(\mu_{d,y}, \eta_v; \rho_{Y_d,V_z;Z}(y, F_{D|Z}(d \mid z); z))(\partial/\partial v)\rho_{Y_d,V_z;Z}(y, F_{D|Z}(d \mid z); z)$$

CI implies

$$\rho_{Y_d,V_z;Z}(y, F_{D|Z}(d \mid z); z) = \rho_{Y_d}(y)$$
$$(\partial/\partial v)\rho_{Y_d,V_z;Z}(y, F_{D|Z}(d \mid z); z) = 0$$

Therefore, for $z \in \{0,1\}$,

$$\Phi^{-1}\left(F_{Y|D,Z}(y \mid d,z)\right) = a_{d,y} + b_{d,y}\Phi^{-1}(F_{D|Z}(d \mid z))$$

with $a_{d,y} \equiv \mu_{d,y}/\sqrt{1-\rho_{Y_d}(y)^2}$, $b_{d,y} \equiv -\rho_{Y_d}(y)/\sqrt{1-\rho_{Y_d}(y)^2}$

Continuous Treatment

$$\Phi^{-1}\left(F_{Y|D,Z}(y \mid d,z)\right) = a_{d,y} + b_{d,y}\Phi^{-1}(F_{D|Z}(d \mid z)) \text{ for } z \in \{0,1\}$$

This is a linear system of two equations on two unknowns, which has solution

$$a_{d,y} = \frac{\Phi^{-1}(F_{Y|D,Z}(y \mid d,0))\Phi^{-1}(F_{D|Z}(d \mid 1)) - \Phi^{-1}(F_{Y|D,Z}(y \mid d,1))\Phi^{-1}(F_{D|Z}(d \mid 0))}{\Phi^{-1}(F_{D|Z}(d \mid 1)) - \Phi^{-1}(F_{D|Z}(d \mid 0))}$$

$$b_{d,y} = \frac{\Phi^{-1}(F_{Y|D,Z}(y \mid d,1)) - \Phi^{-1}(F_{Y|D,Z}(y \mid d,0))}{\Phi^{-1}(F_{D|Z}(d \mid 1)) - \Phi^{-1}(F_{D|Z}(d \mid 0))}$$

Then, we can ID $\mu_{d,y} \equiv \Phi^{-1}(F_{Y_d}(y))$ and $\rho_{Y_d}(y)$ from

$$a_{d,y} \equiv \mu_{d,y}/\sqrt{1-\rho_{Y_d}(y)^2}, \quad b_{d,y} \equiv -\rho_{Y_d}(y)/\sqrt{1-\rho_{Y_d}(y)^2}$$

Continuous Treatment

Theorem 3

Suppose D_z, $z \in \{0,1\}$, satisfies $D_z = F^{-1}_{D|Z}(V_z \mid z)$. Under EX, REL and CI, the functions $y \mapsto F_{Y_d}(y)$ and $y \mapsto \rho_{Y_d}(y)$ are identified on $y \in \mathcal{Y}$ for $d \in \mathcal{D}$ by

$$F_{Y_d}(y) = \Phi\left(\frac{a_{d,y}}{\sqrt{1+b^2_{d,y}}}\right), \quad \rho_{Y_d}(y) = \frac{-b_{d,y}}{\sqrt{1+b^2_{d,y}}}.$$

- unlike Imbens & Newey 09, this approach does not require large support IV nor rank invariance in selection ($V_1 = V_0$)
  - instead, it imposes CI
- unlike D'Haultfoeuille & Février 15; Torgovitsky 15, CI does not impose any structural models for Y and D nor restrictions on the dimension of unobservables

III. Discussions on Copula Invariance

Sufficient Conditions for CI

Recall

- EX: $Z \perp\!\!\!\perp Y_d$ and $Z \perp\!\!\!\perp V_z$
- CI: $\rho_{Y_d, V_z; Z}(y, v; z) = \rho_{Y_d}(y)$

Assumption EX2

For $d \in \mathcal{D}$ and $z \in \{0, 1\}$, $Z \perp\!\!\!\perp (Y_d, V_z)$.

Assumption RI_S

$V_1 = V_0 = V$ a.s.

Assumption CI2

$\rho_{Y_d, V}(y, v) = \rho_{Y_d}(y)$.

- CI2 is CI in treatment propensity

Proposition 1

Under EX2 and RI_S, CI2 implies CI.

Equivalent Condition for CI

Recall CI2: $\rho_{Y_d,V}(y,v) = \rho_{Y_d}(y)$

Assumption SI

For $d \in \mathcal{D}$,

$$F_{Y_d|V}(y \mid v) = \Phi\left(a_{d,y} + b_{d,y}\Phi^{-1}(v)\right), \quad (y,v) \in \mathcal{Y} \times \mathcal{V},$$

where $a_{d,y} = \Phi^{-1}(F_{Y_d}(y))/\sqrt{1-\rho_{Y_d}(y)^2}$ and $b_{d,y} = -\rho_{Y_d}(y)/\sqrt{1-\rho_{Y_d}(y)^2}$.

- SI is single index restriction on local relationship btw (Y_d, V)
- SI *does not* require Gaussianity
- still, e.g., sign of $(\partial/\partial v)F_{Y_d|V,Z}(y \mid v)$ should not depend on v, but can change with y

Proposition 1

CI2 is equivalent to SI.

Local Dependence as Implicit Function

LGR can be expressed for arbitrary copula $\tilde{C}$:

$$\tilde{C}(u_1, u_2 \mid z) = C(u_1, u_2; \rho(u_1, u_2; z))$$

where C is Gaussian copula

For simplicity, maintain EX2 so that

$$\tilde{C}(u_1, u_2) = C(u_1, u_2; \rho(u_1, u_2))$$

By implicit function theorem, ρ is differentiable and

$$\tilde{C}(u_1 \mid u_2) = C(u_1 \mid u_2; \rho(u_1, u_2)) + C_\rho(u_1, u_2; \rho(u_1, u_2))\frac{\partial \rho(u_1, u_2)}{\partial u_2}$$

Proposition 2

Under EX2, CI2 is equivalent to $\tilde{C}(u_1 \mid u_2) = C(u_1 \mid u_2; \rho(u_1))$.

Examples of Distributions under Copula Invariance

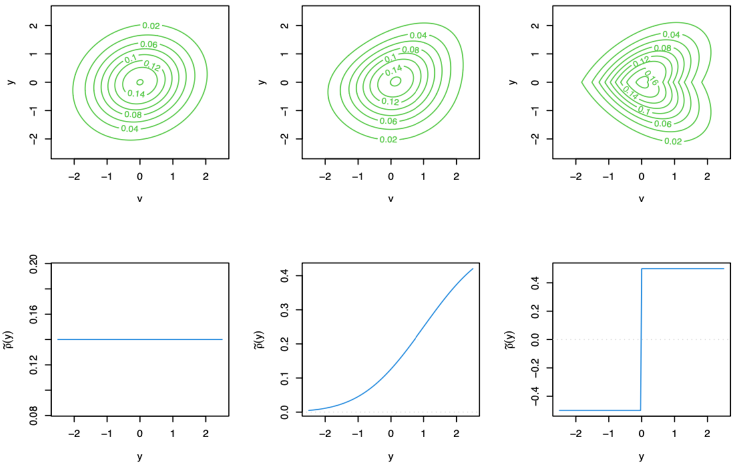


Figure: Joint Distributions under CI2

Notes: We depict joint distributions of (Y_d, V) under CI with Gaussian marginals.

Examples of Distributions under Copula Invariance

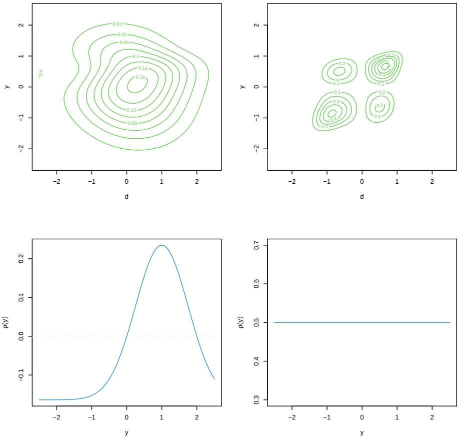


Figure: Joint Distributions under CI2

Notes: We depict joint distributions of (Y_d, V) under CI with Gaussian marginals (left) and nonparametric marginals (right).

Examples of Selection Patterns under Copula Invariance

Suppose $Y = \mu + \varepsilon$ and $D = 1\{V \leq \pi(Z)\}$

- which yields $E[Y|D=1,Z] = \mu + E[\varepsilon|V \leq \pi(Z)]$

We depict $E[\varepsilon|V \leq \pi]$ as a function of π...

- under Gaussian joint distribution (left) and CI (right)

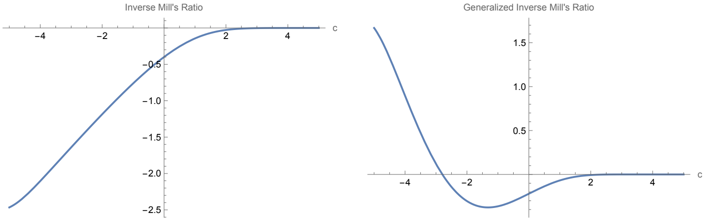


Figure: Control Functions under CI2

Partial Identification and Copula Invariance

When the copula of (Y_d, V_z) is completely unrestricted (true under LGR), it lies between Fréchet-Hoeffding (FH) bounds

- under LGR and EX2, FH lower and upper bounds correspond to $\rho_{Y_d,V_z}(y,v) = -1$ and 1 resp. for all (y,v)
- then, worst-case bounds on F_{Y_d} and its functionals can be obtained
    - e.g., Manski 90's bounds

CI2 achieves intersection of two manifolds: for given y,

$$\{(F_{Y_1}(y), \rho(y, \pi_1(1)))\}$$
$$\{(F_{Y_1}(y), \rho(y, \pi_1(0)))\}$$

generated by the system of equations

- leading to point-identification of $F_{Y_1}(y)$

Comparison to Previous Approaches

Chernozhukov & Hansen 05's IVQR model assumes:

- $Y_d = Q_{Y_d}(U_d)$ for $U_d \sim U[0,1]$
- rank similarity: $U_1 \overset{d}{=} U_0 \mid Z, V$

Then, IVQR yields a conditional moment restriction:

$$\tau = \Pr[Y_1 \leq Q_{Y_1}(\tau), D_z = 1|z] + \Pr[Y_0 \leq Q_{Y_0}(\tau), D_z = 0|z]$$

which can be rewritten as

$$\tau = \Pr[Y_1 \leq Q_{Y_1}(\tau), V_z \leq \pi(z)|z] + \tau - \Pr[Y_0 \leq Q_{Y_0}(\tau), V_z \leq \pi(z)|z]$$

or equivalently

$$\Pr[Y_1 \leq Q_{Y_1}(\tau), V_z \leq \pi(z)|z] = \Pr[Y_0 \leq Q_{Y_0}(\tau), V_z \leq \pi(z)|z]$$

Comparison to Previous Approaches

$$\Pr[Y_1 \leq Q_{Y_1}(\tau), V_z \leq \pi(z)|z] = \Pr[Y_0 \leq Q_{Y_0}(\tau), V_z \leq \pi(z)|z]$$

Using LGR, we can further rewrite above as

$$C(\tau, \pi(z); \rho_{Y_1, V_z; Z}(Q_{Y_1}(\tau), \pi(z); z)) = C(\tau, \pi(z); \rho_{Y_0, V_z; Z}(Q_{Y_0}(\tau), \pi(z); z))$$

This shows that the IVQR also relies on copula invariance:

$$\rho_{Y_1, V_z; Z}(Q_{Y_1}(\tau), \pi(z); z) = \rho_{Y_0, V_z; Z}(Q_{Y_0}(\tau), \pi(z); z), \quad z \in \{0, 1\}$$

In the paper, we also make comparison to other approaches, such as D'Haultfoeuille & Février 15; Torgovitsky 15

More

IV. Estimation and Inference

Estimation Algorithms

Assume a random sample $\{(Y_i, D_i, Z_i, X_i)\}_{i=1}^n$

Notation:

- $B(X_i)$, $B(X_i, Z_i)$, and $B(D_i, X_i, Z_i)$: vectors of transformations
- $I_i(y) \equiv 1\{Y_i \leq y\}$ and $J_i(d) \equiv 1\{D_i \leq d\}$
- $\bar{\mathcal{D}}$ and $\bar{\mathcal{Y}}$: finite grids covering $\mathcal{D}$ and $\mathcal{Y}$
- Φ_2 and Φ are bivariate and univariate Gaussian CDFs

We provide an algorithm for each case

- two-step ML estimation based on distribution regression

Estimation Algorithm: Binary D

Algorithm 1 (Binary D)

1. (Treatment eq.) Estimate π using a Probit regression

$$\widehat{\pi} = \arg\max_{c} \sum_{i=1}^{n} \left[D_i \log \Phi(B(X_i, Z_i)'c) + (1 - D_i) \log(1 - \Phi(B(X_i, Z_i)'c)) \right].$$

2. (Outcome eq.) For $y \in \bar{\mathcal{Y}}$ and $d \in \{0, 1\}$,

$$\widehat{F}_{Y_d|X}(y|x) = \Phi(B(x)'\widehat{\beta}_d(y)) \text{ and } \widehat{\rho}_{Y_d;X}(y;x) = \rho(B(x)'\widehat{\gamma}_d(y)),$$

where $\rho(u) = \tanh(u) \in (-1, 1)$ and

$$\begin{aligned}(\widehat{\beta}_1(y), \widehat{\gamma}_1(y)) = \arg\max_{b,g} \sum_{i=1}^{n} D_i[&I_i(y) \log \Phi_2(B(X_i)'b, B(X_i, Z_i)'\widehat{\pi}, \rho(B(X_i)'g)) \\ &+ (1 - I_i(y)) \log \Phi_2(-B(X_i)'b, B(X_i, Z_i)'\widehat{\pi}, \rho(B(X_i)'g))], \\ (\widehat{\beta}_0(y), \widehat{\gamma}_0(y)) = \arg\max_{b,g} \sum_{i=1}^{n} (1 - D_i)[&I_i(y) \log \Phi_2(B(X_i)'b, -B(X_i, Z_i)'\widehat{\pi}, -\rho(B(X_i)'g)) \\ &+ (1 - I_i(y)) \log \Phi_2(-B(X_i)'b, -B(X_i, Z_i)'\widehat{\pi}, -\rho(B(X_i)'g))].\end{aligned}$$

Estimation Algorithm: Ordered D

Algorithm 2 (Ordered D)

1. (Treatment eq.) Set $\widehat{\pi}_0(z,x) = 0$ and $\widehat{\pi}_K(z,x) = 1$ for all (z,x). For $d \in \{1,\dots,K-1\}$, $\widehat{\pi}_d(z,x) = \Phi(B(z,x)'\widehat{\pi}(d))$, where

$$\widehat{\pi}(d) \in \arg\max_p \sum_{i=1}^n \left[J_i(d) \log \Phi(B(Z_i,X_i)'p) + (1 - J_i(d)) \log \Phi(-B(Z_i,X_i)'p) \right].$$

2. (Outcome eq.) for $y \in \bar{\mathcal{Y}}$ and $d \in \bar{\mathcal{D}}$,

$$\widehat{F}_{Y_d|X}(y|x) = \Phi(B(x)'\widehat{\beta}_d(y)) \text{ and } \widehat{\rho}_{Y_d;X}(y;x) = \rho(B(x)'\widehat{\gamma}_d(y)),$$

where

$$(\widehat{\beta}_d(y), \widehat{\gamma}_d(y)) \in \arg\max_{b,g} \sum_{i=1}^n 1\{D_i = d\} \left[I_i(y) \log g_{d,i}(b,g) + (1 - I_i(y)) \log \bar{g}_{d,i}(b,g) \right],$$

$$\begin{aligned} g_{d,i}(b,g) &\equiv \Phi_2(B(X_i)'b, \Phi^{-1}(\widehat{\pi}_d(Z_i,X_i)), \rho(B(X_i)'g)) \\ &\quad - \Phi_2(B(X_i)'b, \Phi^{-1}(\widehat{\pi}_{d-1}(Z_i,X_i)), \rho(B(X_i)'g)), \\ \bar{g}_{d,i}(b,g) &\equiv \widehat{\pi}_d(Z_i,X_i) - \widehat{\pi}_{d-1}(Z_i,X_i) - g_{d,i}(b,g). \end{aligned}$$

Estimation Algorithm: Continuous D

Algorithm 3 (Continuous D)

1. (Observable conditional dist.) For $y \in \bar{\mathcal{Y}}$ and $d \in \bar{\mathcal{D}}$,
$\widehat{F}_{Y|D,Z,X}(y|d,z,x) = \Phi(B(d,z,x)'\widehat{\beta}(y))$ and
$\widehat{F}_{D|Z,X}(d|z,x) = \Phi(B(z,x)'\widehat{\pi}(d))$, where

$$\widehat{\beta}(y) = \arg\max_b \sum_{i=1}^n \left[I_i(y)\log\Phi(B(D_i,Z_i,X_i)'b) + (1-I_i(y))\log(1-\Phi(B(D_i,Z_i,X_i)'b))\right]$$

$$\widehat{\pi}(d) = \arg\max_p \sum_{i=1}^n \left[J_i(d)\log\Phi(B(Z_i,X_i)'p) + (1-J_i(d))\log(1-\Phi(B(Z_i,X_i)'p))\right]$$

2. (Potential outcome dist.) For $y \in \bar{\mathcal{Y}}$ and $d \in \bar{\mathcal{D}}$,
$\widehat{F}_{Y_d|X}(y|x) = \Phi(\widehat{\mu}_{d,y;x})$ and $\widehat{\rho}_{Y_d;X}(y;x) = -\widehat{b}_{d,y;x}/\sqrt{1+\widehat{b}^2_{d,y;x}}$,
where $\widehat{\mu}_{d,y;x} = \widehat{a}_{d,y;x}/\sqrt{1+\widehat{b}^2_{d,y;x}}$ and

$$\widehat{a}_{d,y;x} = \frac{(B(d,0,x)'\widehat{\beta}(y))(B(1,x)'\widehat{\pi}(d)) - (B(d,1,x)'\widehat{\beta}(y))(B(0,x)'\widehat{\pi}(d))}{B(1,x)'\widehat{\pi}(d) - B(0,x)'\widehat{\pi}(d)},$$

$$\widehat{b}_{d,y;x} = \frac{B(d,1,x)'\widehat{\beta}(y) - B(d,0,x)'\widehat{\beta}(y)}{B(1,x)'\widehat{\pi}(d) - B(0,x)'\widehat{\pi}(d)}.$$

Estimation Algorithm: F_{Y_d}, QSF and ASF

Algorithm 4 (F_{Y_d}, QSF and ASF)

1. Unconditional distribution: for $y \in \bar{\mathcal{Y}}$ and $d \in \bar{\mathcal{D}}$,

$$\widehat{F}_{Y_d}(y) = \frac{1}{n}\sum_{i=1}^{n} \widehat{F}_{Y_d|X}(y \mid X_i).$$

For $y \in \mathcal{Y} \setminus \bar{\mathcal{Y}}$ and $d \in \bar{\mathcal{D}}$,

$$\widehat{F}_{Y_d}(y) = \max\{\widehat{F}_{Y_d}(\bar{y}) : \bar{y} < y, \bar{y} \in \bar{\mathcal{Y}}\}.$$

2. Quantile and average structural functions:

$$\widehat{QSF}_{\tau}(d) = \widehat{Q}_{Y_d}(\tau) = \mathcal{Q}_{\tau}(\widehat{F}_{Y_d}),$$
$$\widehat{ASF}(d) = \widehat{E}[Y_d] = \mathcal{E}(\widehat{F}_{Y_d}).$$

Inference

Denote the functional parameters by

$$u \mapsto \delta_u, \quad u \in \mathcal{U}$$

- e.g., if we are interested in $\tau \mapsto QSF_\tau(d)$ on $[.05, .95]$, then $u = \tau$, $\delta_u = QSF_u(d)$ and $\mathcal{U} = [.05, .95]$
- in practice, we approximate $\mathcal{U}$ using a fine grid $\bar{\mathcal{U}}$

Let $\widehat{\delta}_u$ be the estimator of δ_u obtained from Algorithms 1–4

Then, we establish FCLT that

$$\sqrt{n}(\widehat{\delta}_u - \delta_u) \rightsquigarrow Z_\delta \text{ in } \ell^\infty(\mathcal{U})$$

where Z_δ is a mean-zero Gaussian process and that the bootstrap is consistent for estimating Z_δ

Inference

Algorithm 5 (Bootstrap for Uniform Confidence Band)

1. For $u \in \bar{\mathcal{U}}$, obtain B bootstrap draws $\{\widehat{\delta}_u^{(b)} : 1 \leq b \leq B\}$ of the estimator $\widehat{\delta}_u$.
2. For $u \in \bar{\mathcal{U}}$, compute the robust standard error,

$$SE(\widehat{\delta}_u) = (\widehat{Q}_\delta(0.75, u) - (\widehat{Q}_\delta(0.25, u))/(\Phi^{-1}(0.75) - (\Phi^{-1}(0.25)),$$

where $\widehat{Q}_\delta(\tau, u)$ is the τ-quantile of $\{\widehat{\delta}_u^{(b)} : 1 \leq b \leq B\}$.
3. Compute the critical value as

$$cv(1-\alpha) = (1-\alpha)\text{-quantile of } \left\{ \max_{u \in \bar{\mathcal{U}}} \frac{|\widehat{\delta}_u^{(b)} - \widehat{\delta}_u|}{SE(\widehat{\delta}_u)} : 1 \leq b \leq B \right\}.$$

4. Compute the $(1-\alpha)$ uniform confidence band as

$$CB_{(1-\alpha)}(\delta_u) = [\widehat{\delta}_u \pm cv(1-\alpha) SE(\widehat{\delta}_u)], \quad u \in \bar{\mathcal{U}}.$$

Inference

The uniform confidence bands $CB_{(1-\alpha)}(\delta_u)$ satisfies

$$\lim_{n\to\infty} \Pr[\delta_u \in CB_{(1-\alpha)}(\delta_u) \text{ for all } u \in \mathcal{U}] = 1 - \alpha$$

For bootstrap in Step 1, we recommend...

- binary and ordered D: multiplier bootstrap (based on influence function)
  - as nonlinear optimization is involved
- continuous D: standard empirical bootstrap

V. Empirical Application with Continuous Treatment

Distributional Effects of Sleep on Well-Being

Bessone et al 2021 analyzed the effects of randomized interventions to increase sleep of low-income adults in India

Bessone et al 2021; Dong & Lee 2023 used TSLS

- we estimate the distributional effects of sleep on well-being

Y: overall index of individual well-being

D: sleep per night, in hours (continuous)

Z: randomly assigned experimental treatments (binary)

- Z_1: devices + encouragement
- Z_2: devices + incentives
- $Z = Z_1 + Z_2$ ($= 1$: any treatment; $= 0$ none)

X: gender, three age indicators, baseline well-being index

Distributional Effects of Sleep on Well-Being

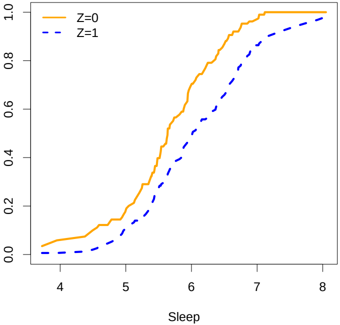


Figure: Distributional First Stage

Notes: Control for gender, three age indicators, and baseline well-being index. $n = 226$.

Distributional Effects of Sleep on Well-Being

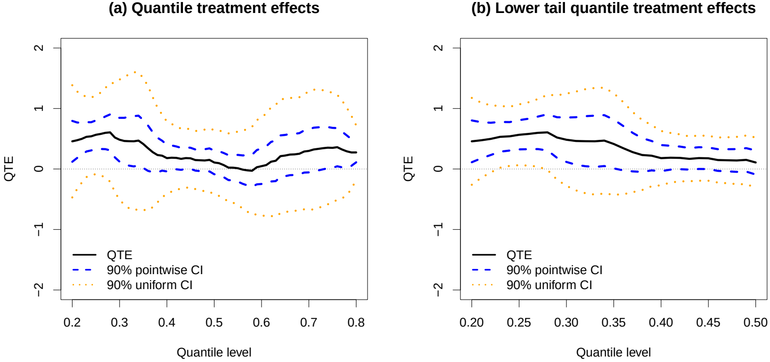


Figure: Quantile Treatment Effects

Notes: We report the normalized QTE, $(\mathcal{Q}_\tau(\widehat{F}_{Y_{d''}}) - \mathcal{Q}_\tau(\widehat{F}_{Y_{d'}}))/(d'' - d')$ with d'' and d' being 75% and 25% quantiles of sleep. Uniform and pointwise CIs are computed using empirical bootstrap with 5000 repetitions. We control for gender, three age indicators, and baseline well-being index. $n = 226$.

Distributional Effects of Sleep on Well-Being

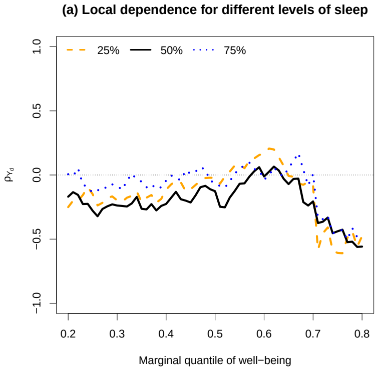


Figure: Local Dependence Functions

Notes: We report the average of $\widehat{\rho}_{Y_d}(y; X_i)$ with d being 25%, 50%, 75% quantiles of sleep. We control for gender, three age indicators, and baseline well-being index. $n = 226$.

Distributional Effects of Sleep on Well-Being

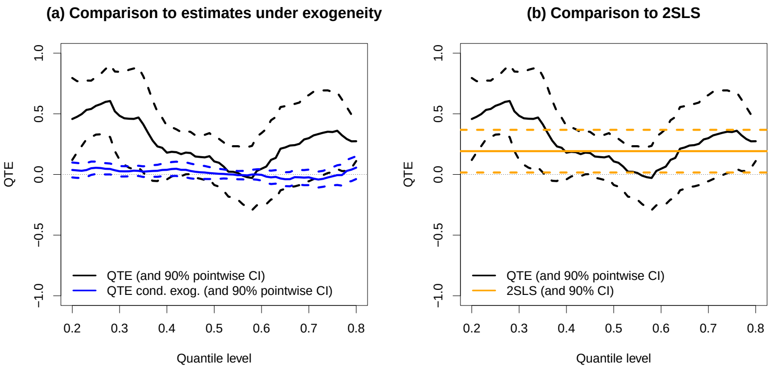


Figure: Comparison to Estimators under Conditional Exogeneity and 2SLS

Notes: We report the normalized QTE, $(\mathcal{Q}_\tau(\widehat{F}_{Y_{d''}}) - \mathcal{Q}_\tau(\widehat{F}_{Y_{d'}}))/(d'' - d')$ with d'' and d' being 75% and 25% quantiles of sleep. Conditional exogeneity assumes $Y_d \perp\!\!\!\perp D|X$. Pointwise CIs with empirical bootstrap with 5000 repetitions. We control for gender, three age indicators, and baseline well-being index. $n = 226$.

VI. Conclusions

Conclusions

In identifying treatment effects under endogeneity, researchers face modeling trade-offs

This paper proposes a new direction to explore modeling trade-offs

- based on LGR
- impose assumption on local dependence parameter
- allow rich heterogeneity in outcome and treatment processes
- lead to simple estimation and inference procedures, appealing to practitioners
- can also estimate the dependence function (which reveals patterns of endogeneity)

Thank You! ☺

Global Univalence by Gale & Nikaido 65

Definition (P-matrix)

A square matrix J is called a P-matrix if all its principal minors are positive.

- a *principal minor* is the determinant of a submatrix obtained from J when the same set of rows and columns are deleted

Theorem (Global Univalence by Gale & Nikaido 65)

If $F : \Omega \to \mathbb{R}^n$, where Ω is a closed rectangular region of $\mathbb{R}^n$, is a differentiable mapping such that the Jacobian matrix $J(x)$ is a P-matrix for all x in Ω, then F is univalent in Ω.

- Jacobian of our mapping $\Pi : \theta \to p$ is P-matrix by the properties of Gaussian copula

Global Identification Using Ambrosetti & Prodi 95

Theorem (Ambrosetti & Prodi 95)

Suppose $F : X \to Y$ is continuous, proper and locally invertible in X and let Y be connected. Then, the cardinality of $F^{-1}(\{y\})$ is constant for all $y \in Y$.

- our mapping $\Pi : \theta \to p$ is proper by the properties of copula
- local invertibility is guaranteed by full rank Jacobian of Π
- take the value of θ such that $\rho = 0$; then $\left|\Pi^{-1}(\{\Pi(\theta)\})\right| = 1$ for such θ

Return

Comparison to Torgovitsky 10

Both CCI and CI restrict the dependence of (Y_d, D) on Z...

- by requiring $\rho(\cdot)$ not to depend on $Z = z$

But Torgovitsky 10 maintains RI...

- so restricting the copula of (U, D) is sufficient

Our strategy does not depend on RI...

- such that we need to impose CI for both Y_1 and Y_0
- as trade-off of not assuming RI, we require CI that $\rho(\cdot)$ is not a function of $F_{D|Z}$

Return