Endogenous Controls Bias: Correction via Nonparametric Methods

Kaicheng Chen[1] Kyoo il Kim[2]

[1]Shanghai University of Finance and Economics

[2]Michigan State University

2026 AMES, Hong Kong
June 20, 2026

The message of the talk

- **Control variables** are routinely treated as exogenous, yet in many applications they are potentially **endogenous**.
- **Dilemma:** omitting controls from the model may leave the treatment endogenous, while including them may contaminate identification.
- **Endogenous control bias** in parametric models: e.g., linear (IV) and probit models.
- Correction via **nonparametric methods**.
  - **Intuition:** control endogeneity is generally innocuous in nonparametric models.
  - **We show:** (1) not exactly true; (2) when it is true; (3) what is recovered; (4) how to test.
- **Takeaway:** when controls are endogenous, nonparametric methods should be used, with care.

Motivation

Example 1: simple linear model

- To illustrate, consider a simple linear model:

$$Y = \tau W + \xi X + \varepsilon, \;\; E[\varepsilon | W, X] = E[\varepsilon | X].$$

- ε may not be mean-independent of X.
  - Without X, it may cause omitted variable bias;
  - With X, the dependence between X and ε contaminates the identification of τ.

- Let $D = (W, X)'$ and $\theta = (\tau, \xi)$, then the linear projection parameters are

$$\gamma_{\text{LP}} \equiv E[DD']^{-1} E[DY] = \theta + E[DD']^{-1} E\left[E[D\varepsilon | W, X]\right] = \theta + E[DD']^{-1} E\left[D E[\varepsilon | X]\right].$$

- However, we observe that

$$\tau = \partial_W E[Y | W, X].$$

Example 2: linear IV model

- Consider a linear model with an instrument Z:

$$Y = \tau W + \xi X + \varepsilon, \;\; E[\varepsilon | W, X] \neq E[\varepsilon | X]$$
$$W = \pi_Z Z + \pi_X X + \eta, \;\; E[\eta | Z, X] = 0$$

- Z is conditionally valid: $E[\varepsilon | Z, X] = E[\varepsilon | X]$.

- Let $Q = (Z, X)'$, $D = (W, X)'$, and $\theta = (\tau, \xi)$. The linear IV projection is given by

$$\gamma_{\mathrm{IV}} = E[QD']^{-1} E[QY] = \theta + E[QD']^{-1} E[E[Q\varepsilon | Z, X]] = \theta + E[QD']^{-1} E[QE[\varepsilon | X]].$$

- However, we observe that

$$\tau = \partial_Z E[Y | Z, X] / \partial_Z E[W | Z, X].$$

Example 3: probit model

- For a binary outcome Y, consider a threshold-crossing model

$$Y = 1\{\tau W + \xi X + \varepsilon > 0\}, \ \varepsilon | W, X \sim N(\mu(X), \sigma(X)^2)$$

  where $\mu(X)$ and $\sigma(X)$ are unknown.

- A parameter of interest is

$$\beta := E\left[\frac{\tau}{\sigma(X)}\phi\left(\frac{\tau W + \xi X - \mu(X)}{\sigma(X)}\right)\right],$$

  which is in general not equal to the the APE estimand of probit QMLE: $\tau^* E[\phi(\tau^* W + \xi^* X)]$ where (τ^*, ξ^*) are the estimands of the probit QMLE.

- However, we observe that

$$E[\partial_W E(Y|W, X)] = \beta.$$

The general idea

- The parametric estimators above are all biased for the corresponding parameters due to **lack of identification** under the endogenous control.

- But these parameters are **recovered through some nonparametric methods**.

- Is there a **general approach** to identify the parameter of interest with endogenous controls?

- What is the **targeted parameter** that can be properly defined when covariates are endogenous?

Literature

- There is **not** a big literature devoted to endogenous controls.
- Frölich (2008)–nonparametric methods for endogenous controls in **linear (IV) models**.
- **“Bad control”**–potential outcome counterpart of endogenous control:
  - Wooldridge (2005)–including these would cause the conditional unconfoundedness to **fail**.
  - Lechner (2008)–Same argument and **reformulate the sufficient conditions** for binary treatment.
  - Angrist and Pischke (2009)–**selection bias** and loss of causal interpretation.
- Kim (2013)–for **RDD**, endogenous control in **kernel/local linear 2SLS** yields asymptotic bias, yet is still consistent.
- Andrews, Barahona, Gentzkow, Rambachan, and Shapiro (2025): **endogenous controls due to misspecification** in structural models; their **“strong exclusion” IV** method amounts to including more flexible functional forms of X as the control function.

Main result

A general framework: what is the parameter of interest

- Consider a generic parametric model with endogenous control:

$$Y = h(W, X, \varepsilon; \theta), E[\varepsilon|W, X] = E[\varepsilon|X]$$

- In all examples above, the parameters of interest can be defined as the averages of **local average responses (LAR)**.
- For a continuous Y, LAR is defined as

$$\beta(w, x) = \int \partial_w h(w, x, \epsilon; \theta) f_{\varepsilon|W=x,X=x}(\epsilon) d\epsilon$$

- For a binary Y, write $Y = 1\{h^*(W, X, \varepsilon; \theta) > 0\}$ with implicitly defined h^*. let (u, v) by some partition of ε such that u^* is uniquely defined by $h^*(W, X, u, v; \theta) = 0$. Then, define

$$\beta(w, x) = \int -\partial_w u^*(w, x, v) d_{u,v|W=x,X=x}(u^*, v) dv$$

- The average LAR is defined as $\beta = E[\beta(W, X)]$.

Benchmark result: identification of β via APE

- Let $\widetilde{X}$ denote a collection of control variables such that $X \subseteq \widetilde{X}$.
- Assume (i) $\varepsilon \perp\!\!\!\perp W \mid \widetilde{X}$ (**conditional independence**), and (ii) enough support to vary W holding $\widetilde{X}$ fixed (**measurable separability**).
- We show that under (i), (ii), and regularity conditions (smoothness and boundedness of m/m^*), for both continuous and binary Y, (**Theorem 1**)

$$\beta = \mathbb{E}\left[\partial_W \mathbb{E}(Y \mid W, \widetilde{X})\right].$$

Why is the rank constraint important

- **Measurable separability (MS)**: W and X are measurably separated if any function of W, almost surely equal to a function of X, must be almost surely equal to a constant.

- In Example 1, Theorem 1 implies $\tau = \int \partial_w E[Y|W = w, X = x] dF(w, x)$.

- If violated, $\exists l, g$ s.t. $l(W) = g(X)$ a.s., and it is not a constant. Then, on the event $\{W = w, X = x\}$, it is necessary that $l(w) = g(x)$ a.e.

- **To find** $\partial_w E[Y|W = w, X = x]$, we need to fix $X = x$ while letting $W = w + \delta$ for some $\delta \to 0$, which necessitates $l(w + \delta) = g(x) = l(w)$ for all δ. But this requires $l(w)$ to be constant locally.

- In other words, $\partial_w E[Y|W = w, X = x]$ is not identified. What can be identified is

$$\frac{d}{dw} E[Y|W = w, X = x(w)] = \tau + \frac{d}{dw}(\xi x(w) + E[\varepsilon|X = x(w)]) \neq \tau.$$

  where $x(w)$ is such that $l(w) = g(x(w))$.

Why is the rank constraint important

- **Measurable separability (MS)**: W and X are measurably separated if any function of W, almost surely equal to a function of X, must be almost surely equal to a constant.
- In Example 1, Theorem 1 implies $\tau = \int \partial_w E[Y|W = w, X = x] dF(w, x)$.
- If violated, $\exists l, g$ s.t. $l(W) = g(X)$ a.s., and it is not a constant. Then, on the event $\{W = w, X = x\}$, it is necessary that $l(w) = g(x)$ a.e.
- **To find** $\partial_w E[Y|W = w, X = x]$, we need to fix $X = x$ while letting $W = w + \delta$ for some $\delta \to 0$, which necessitates $l(w + \delta) = g(x) = l(w)$ for all δ. But this requires $l(w)$ to be constant locally.
- In other words, $\partial_w E[Y|W = w, X = x]$ is not identified. What can be identified is

$$\frac{d}{dw} E[Y|W = w, X = x(w)] = \tau + \frac{d}{dw}(\xi x(w) + E[\varepsilon|X = x(w)]) \neq \tau.$$

where $x(w)$ is such that $l(w) = g(x(w))$.

Why is the rank constraint important

- **Measurable separability (MS)**: W and X are measurably separated if any function of W, almost surely equal to a function of X, must be almost surely equal to a constant.
- In Example 1, Theorem 1 implies $\tau = \int \partial_w E[Y|W = w, X = x] dF(w, x)$.
- If violated, $\exists l, g$ s.t. $l(W) = g(X)$ a.s., and it is not a constant. Then, on the event $\{W = w, X = x\}$, it is necessary that $l(w) = g(x)$ a.e.
- **To find** $\partial_w E[Y|W = w, X = x]$, we need to fix $X = x$ while letting $W = w + \delta$ for some $\delta \to 0$, which necessitates $l(w + \delta) = g(x) = l(w)$ for all δ. But this requires $l(w)$ to be constant locally.
- In other words, $\partial_w E[Y|W = w, X = x]$ is not identified. What can be identified is

$$\frac{d}{dw} E[Y|W = w, X = x(w)] = \tau + \frac{d}{dw}(\xi x(w) + E[\varepsilon|X = x(w)]) \neq \tau.$$

where $x(w)$ is such that $l(w) = g(x(w))$.

Why is the rank constraint important

- **Measurable separability (MS)**: W and X are measurably separated if any function of W, almost surely equal to a function of X, must be almost surely equal to a constant.
- In Example 1, Theorem 1 implies $\tau = \int \partial_w E[Y|W = w, X = x] dF(w, x)$.
- If violated, $\exists l, g$ s.t. $l(W) = g(X)$ a.s., and it is not a constant. Then, on the event $\{W = w, X = x\}$, it is necessary that $l(w) = g(x)$ a.e.
- **To find** $\partial_w E[Y|W = w, X = x]$, we need to fix $X = x$ while letting $W = w + \delta$ for some $\delta \to 0$, which necessitates $l(w + \delta) = g(x) = l(w)$ for all δ. But this requires $l(w)$ to be constant locally.
- In other words, $\partial_w E[Y|W = w, X = x]$ is not identified. What can be identified is

$$\frac{d}{dw} E[Y|W = w, X = x(w)] = \tau + \frac{d}{dw}(\xi x(w) + E[\varepsilon|X = x(w)]) \neq \tau.$$

where $x(w)$ is such that $l(w) = g(x(w))$.

Why is the rank constraint important

- **Measurable separability (MS)**: W and X are measurably separated if any function of W, almost surely equal to a function of X, must be almost surely equal to a constant.
- In Example 1, Theorem 1 implies $\tau = \int \partial_w E[Y|W = w, X = x] dF(w, x)$.
- If violated, $\exists l, g$ s.t. $l(W) = g(X)$ a.s., and it is not a constant. Then, on the event $\{W = w, X = x\}$, it is necessary that $l(w) = g(x)$ a.e.
- **To find** $\partial_w E[Y|W = w, X = x]$, we need to fix $X = x$ while letting $W = w + \delta$ for some $\delta \to 0$, which necessitates $l(w + \delta) = g(x) = l(w)$ for all δ. But this requires $l(w)$ to be constant locally.
- In other words, $\partial_w E[Y|W = w, X = x]$ is not identified. What can be identified is

$$\frac{d}{dw} E[Y|W = w, X = x(w)] = \tau + \frac{d}{dw}(\xi x(w) + E[\varepsilon|X = x(w)]) \neq \tau.$$

  where $x(w)$ is such that $l(w) = g(x(w))$.

Why is the rank constrain innocuous in most applications

- Florens, Heckman, Meghir, and Vytlacil (2008, ECTA) give primitive conditions, which essentially requires that holding X fixed, there is enough support to vary W.

- As long as there is sufficient variation in W sourced outside X, this condition holds.

Intuition: why nonparametric approach works

- Including X parametrically blocks out some channels that cause endogeneity of W.
- It amounts to holding a particular form of X constant, while allowing other forms of X to vary when we move W around.
- However, these extra variations in X are not orthogonal to the error term.
- Nonparametric methods hold all forms of X constant while moving W.

Main result with IV

- Consider

$$Y = h(W, X, \varepsilon; \theta), \qquad W = q(Z, X, \eta).$$

- An IV condition that allows for endogenous control:

$$Z \perp\!\!\!\perp (\varepsilon, \eta) \mid X.$$

- By extending Imbens and Newey (2009), we show, for a continuous W,

$$V = F_{W|Z,X}(W)$$

can restore conditional independence (**Theorem 2**):

$$\varepsilon \perp\!\!\!\perp W \mid X, V.$$

- If, additionally, W and (X, V) are MS, then Theorem 1 implies:

$$\beta = \mathbb{E}\left[\partial_W \mathbb{E}(Y \mid W, \widetilde{X})\right], \ \widetilde{X} = (X, V).$$

Test for endogenous control

Three parameters to distinguish

- The **parameter of interest** is $\beta_0 := E[\partial_W h(W, X, \varepsilon; \theta)]$.
- The **estimand of the proposed nonparametric method** (average derivative estimator, ADE) is $\beta_{np} = E[\partial_W E[Y|W, X]]$.
- The **estimand of the parametric method** is denoted by β_p.
  - In Examples 1&2, β_p is the linear projection parameter associated with W.
  - In Example 3, β_p is $\tau^* E[\phi(\tau^* W + \xi^* X)]$.
- If X is exogenous, β_p and β_{np} coincide; otherwise, they differ–A **Hauseman-type** (1978) test.

Simulation

Simulation design

- What to demonstrate:
  - Omitting endogenous controls creates bias;
  - Including endogenous controls in a restricted way is still biased.
  - ADE recovers the targeted parameter under CIA and MS.
  - ADE is biased when CIA holds but MS fails.
  - The empirical coverage and power of the test.
- Implementation: ADE through polynomial series approximation.

DGPs and results

Application

Application: China Syndrome revisit

- We revisit the import-competition design of Autor, Dorn, and Hanson (2013).

$$\Delta L_{it} = \gamma_t + \tau \Delta I_{US,it} + X_i' \xi + e_{it}.$$

- L_{it} is the regional share of manufacturing employment.
- $I_{US,it}$ is a *share-shift* measure: manufacturing industry *share* at the start of the observation period $\times$ growth in US imports from China (*shift*).
- Δ denotes a decade-long first-difference.

Endogeneity in treatment and controls

- Two sources of endogeneity in treatment $I_{US,it}$.
  - For *shift*, use import growth in other high-income countries as exogenous shift. The resulting IV is $I_{O,it}$
  - For *share*: use first-differenced treatment and outcome; include other start-of-period measures of regional market conditions and demographics.
- Controls are relevant: the results in Autor et al. (2013) show that adding more controls reduces the size of the estimated (negative) impact.

Application results: pattern

Controls	IV(se)	ADE(se)	Difference(se)
(1) Time dummies only	−0.746(0.070)	−1.036(0.144)	0.290(0.111)
(2) Start-of period employment + (1)	−0.610(0.102)	−0.498(0.082)	−0.113(0.124)
(3) Census division + (2)	−0.538(0.094)	−0.420(0.069)	−0.118(0.115)
(4) Start-of period demographics + (3)	−0.508(0.082)	−0.182(0.083)	−0.326(0.111)
(5) Industry controls + (3)	−0.562(0.085)	−0.126(0.075)	−0.436(0.130)
(6) All controls	−0.596(0.095)	−0.037(0.068)	−0.559(0.127)

Main takeaways

1. When controls are endogenous, including the control in a restricted way may fail to identify treatment parameters.
   - This is the case even if the model is correctly specified.
2. We consider a general class of models and show identification of average LAR through $E[\partial_W E[Y|W, \tilde{X}]]$ under CIA and a rank condition.
3. The implementation is straightforward through with existing nonparametric methods.
4. The gap between our method and a parametric method leads to a test:
   - Given correct specification, it tests for endogenous controls.
   - If the parametric model is misspecified, then the test can be rejected for both reasons.
5. In a word: if controls are potentially endogenous, nonparametric methods should be used, but only when we know **what is of interest**, **what is recovered**, and **under what conditions**.

Thank you

Backup slides

Simulation design: DGP 1

$$\text{DGP(1)}: \; Y = \tau W + X_1 + X_2 + U, \quad W = \frac{1}{2}\exp(a) + N(0,1)$$
$$X_1 = a + \frac{1}{2}\exp(p), \quad X_2 = p, \quad U = \frac{1}{2}\exp(b) + \frac{1}{2}\exp(q) + N(0,1),$$

- (a, b) and (p, q) are independent of each other, and each are jointly normal with mean zero, variance one, and covariance ρ;
- $N(0, 1)$ denotes a random draw from a standard normal distribution, independent of (a, b) and (p, q).
- We observe that (1) $X = (X_1, X_2)'$ are relevant for both Y and W; (2) W and X are dependent on U; (3) Conditional on X, W is independent of U; and (4) W and X are MS.

Back

Simulation design: DGP 2

$$\text{DGP(2)}: \; Y = \tau W + X_1 + X_2 + U, \quad W = X_1 + X_2 + Z + \eta$$
$$X_1 = a + \frac{1}{2}\exp(p), \quad X_2 = p, \quad Z = \frac{1}{2}\exp(a) + N(0,1)$$
$$\eta = \frac{1}{2}\exp(\xi), \quad U = \frac{1}{2}\exp(b) + \frac{1}{2}\exp(q) + \frac{1}{2}\exp(\zeta) + N(0,1),$$

- (ξ, ζ) are also jointly normal with mean zero, variance one, and covariance ρ.
- We observe that (i) W is not conditionally independent given X; (ii) Z is a valid IV only when conditional on X, but X is endogenous; (iii) Z and X are MS. (iv) The MS between Z and X here also implies the MS between W and (X, η).

Back

Simulation results

Table: Comparing parametric and nonparametric estimators under endogenous controls

ρ	Monte Carlo Statistics	DGP(1) OLS w.o. X	DGP(1) OLS w. X	DGP(1) ADE w. X	DGP(2) IV w.o. X	DGP(2) IV w. X	DGP(2) ADE w. $(X, \widehat{V})$
0.75	Bias	0.732	0.156	0.013	0.528	0.156	0.020
	SD	0.091	0.080	0.048	0.057	0.083	0.063
	MSE	0.545	0.031	0.002	0.282	0.031	0.004
0.5	Bias	0.588	0.088	0.004	0.424	0.089	0.001
	SD	0.083	0.063	0.054	0.052	0.068	0.069
	MSE	0.352	0.012	0.003	0.182	0.012	0.005
0	Bias	0.383	0.000	0.000	0.277	0.009	0.004
	SD	0.065	0.045	0.058	0.044	0.052	0.073
	MSE	0.151	0.002	0.003	0.079	0.003	0.005

Back

Simulation results: empirical coverage

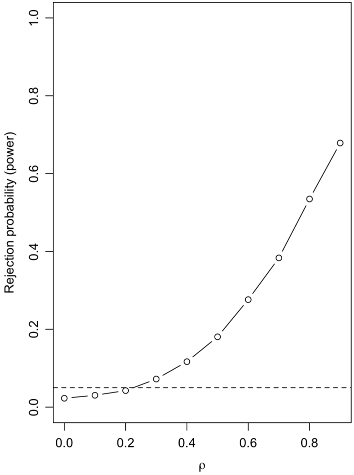
Table: Empirical Coverage with a nominal rate 0.95

Sample size	1000		500		100	
Bootstrap reps	1000	500	1000	500	1000	500
Coverage (%)	96.6	95.9	97.3	97.7	99.9	99.9

Note: Simulation results are based on 10,000 Monte Carlo replications.

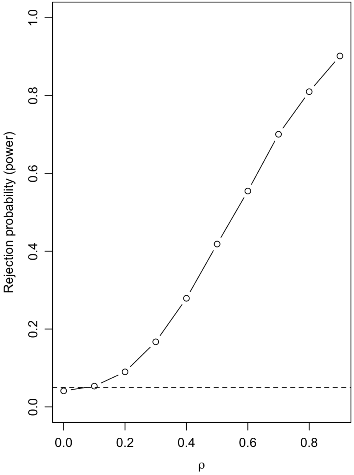
Back

Simulation results: Empirical power with $n = 500$



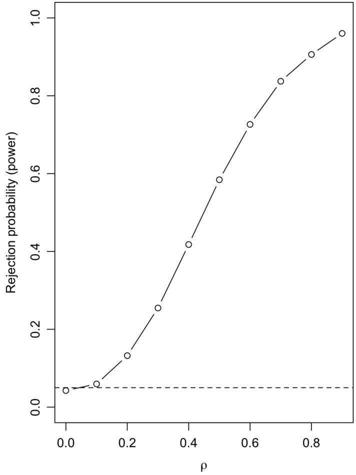
Back

Simulation results: Empirical power with $n = 1000$



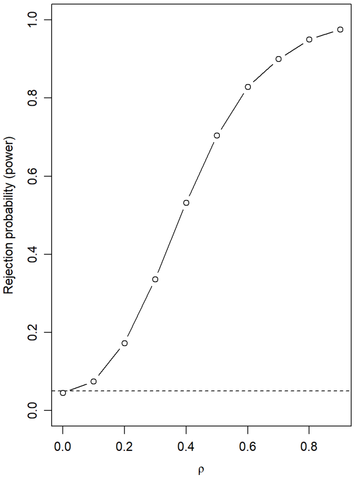
Back

Simulation results: Empirical power with $n = 1500$



Back

Simulation results: Empirical power with $n = 2000$



Back