Improving IV Estimates When the Direction of Endogeneity Bias is Known

Kirill Ponomarev

University of Chicago

September 22, 2024

Outline

Motivation and Literature

Estimation in the Normal Experiment

Inference in the Normal Experiment

Asymptotic Results

Empirical Applications

Motivation

- In many settings, researchers may have strong prior beliefs about endogeneity bias.
- Examples:
  - Omitted variables: regression of income on education with unobserved ability $\implies$ OLS biased upwards;

Motivation

- In many settings, researchers may have strong prior beliefs about endogeneity bias.
- Examples:
  - Omitted variables: regression of income on education with unobserved ability $\Longrightarrow$ OLS biased upwards;
  - Simultaneity: preexisting skills facilitate new technology and new technology attracts skills $\Longrightarrow$ OLS biased upwards;

Motivation

- In many settings, researchers may have strong prior beliefs about endogeneity bias.
- Examples:
  - Omitted variables: regression of income on education with unobserved ability $\implies$ OLS biased upwards;
  - Simultaneity: preexisting skills facilitate new technology and new technology attracts skills $\implies$ OLS biased upwards;
  - Selection: judges are more likely to send "bad guys" to prison than EM $\implies$ OLS biased downward;

Motivation

- In many settings, researchers may have strong prior beliefs about endogeneity bias.
- Examples:
  - Omitted variables: regression of income on education with unobserved ability $\implies$ OLS biased upwards;
  - Simultaneity: preexisting skills facilitate new technology and new technology attracts skills $\implies$ OLS biased upwards;
  - Selection: judges are more likely to send "bad guys" to prison than EM $\implies$ OLS biased downward;
- Such beliefs are typically not incorporated in statistical analysis.

Motivation

- Let β and γ denote the IV and OLS estimands; $\hat{\beta}_n$ and $\hat{\gamma}_n$ denote the corresponding estimators.
- With strong IVs, for large n:

$$\begin{bmatrix} \hat{\beta}_n \\ \hat{\gamma}_n \end{bmatrix} \overset{d}{\approx} N\left(\begin{bmatrix} \beta \\ \gamma \end{bmatrix}, \begin{bmatrix} \hat{\sigma}^2_\beta & \hat{\sigma}_{\beta\gamma} \\ \hat{\sigma}_{\beta\gamma} & \hat{\sigma}^2_\gamma \end{bmatrix} \right).$$

  Additionally, know that $\beta \leqslant \gamma$.

- Typically $\hat{\sigma}^2_\beta \gg \hat{\sigma}^2_\gamma$, while $|\hat{\gamma}_n - \hat{\beta}_n|$ is often relatively small.
- As a result, $\hat{\gamma}_n$ should be informative about β.

Motivation

- Let β and γ denote the IV and OLS estimands; $\hat{\beta}_n$ and $\hat{\gamma}_n$ denote the corresponding estimators.
- With strong IVs, for large n:

$$\begin{bmatrix} \hat{\beta}_n \\ \hat{\gamma}_n \end{bmatrix} \overset{d}{\approx} N\left(\begin{bmatrix} \beta \\ \gamma \end{bmatrix}, \begin{bmatrix} \hat{\sigma}^2_\beta & \hat{\sigma}_{\beta\gamma} \\ \hat{\sigma}_{\beta\gamma} & \hat{\sigma}^2_\gamma \end{bmatrix} \right).$$

  Additionally, know that $\beta \leqslant \gamma$.

- Typically $\hat{\sigma}^2_\beta \gg \hat{\sigma}^2_\gamma$, while $|\hat{\gamma}_n - \hat{\beta}_n|$ is often relatively small.
- As a result, $\hat{\gamma}_n$ should be informative about β.

How to use this information in statistical analysis?

Related Literature: Econometrics

- Extra inequality restrictions:
  - Moon, Schorfheide (2009);
  - Andrews, Armstrong (2017);
  - Ketz, McCloskey (2022);
  - Cox (2024);
- Inference with nuisance parameters under the null:
  - Moreira (2003), Ketz (2018);
  - Romano, Shaikh, Wolf (2014);
  - Elliott, Müller, Watson (2015), Montiel Olea (2018);
- Parameters on the boundary:
  - Andrews (1999; 2001); McCloskey (2012);
  - Andrews, Guggenberger (2010); AG & Cheng (2011)
  - Fang and Santos (2019); Hong and Li (2020)

Related Literature: Statistics

- Estimation with restricted parameter spaces:
  - Lee (1981); Kelly (1989); Hwang, Peddada (1994); Peddada, Dunson, Tan (2005); Chang, Fukuda, Shinozaki (2017);
  - Marchard, Strawderman (2004); van Eden (2006);
- Inference with restricted parameter spaces:
  - Perlman, Wu (1999);
  - Silvapulle, Sen (2005);
  - Lehman, Romano (2005);

Outline

Motivation and Literature

Estimation in the Normal Experiment

Inference in the Normal Experiment

Asymptotic Results

Empirical Applications

Estimation in the Normal Experiment

- Let $X \sim N(\theta, \Sigma)$ in $\mathbb{R}^2$ with $\theta_1 \leqslant \theta_2$ and known Σ.
- The goal is to estimate $\theta_1 \in \mathbb{R}$.
- Unbiased MLE is X_1. In the constrained experiment, it is still minimum-variance unbiased and minimax.

Estimation in the Normal Experiment

- Let $X \sim N(\theta, \Sigma)$ in $\mathbb{R}^2$ with $\theta_1 \leqslant \theta_2$ and known Σ.
- The goal is to estimate $\theta_1 \in \mathbb{R}$.
- Unbiased MLE is X_1. In the constrained experiment, it is still minimum-variance unbiased and minimax.
- A natural alternative is Restricted MLE, defined as:

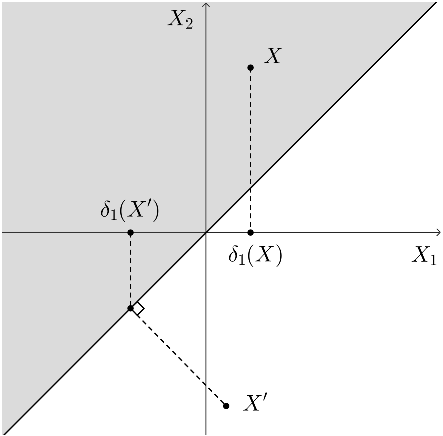
$$(\delta_1(X), \delta_2(X)) = \underset{\theta \in \mathbb{R}^2}{\operatorname{argmin}}\{(X-\theta)'\Sigma^{-1}(X-\theta) \mid \theta_1 \leqslant \theta_2\}.$$

- The solution is:

$$\delta_1(X) = X_1 \cdot \mathbf{1}(X_1 \leqslant X_2) + (w_1 X_1 + w_2 X_2) \cdot \mathbf{1}(X_1 > X_2),$$

where $w_2 = (\sigma_1^2 - \sigma_{12})/(\sigma_1^2 + \sigma_2^2 - 2\sigma_{12})$, and $w_1 = 1 - w_2$.

Illustration: RMLE



RMLE as an Estimator

Lemma 1

1. *If* $\sigma_1^2 \neq \sigma_{12}$*, then for all* $t > 0$ *and* $\theta_1 \leqslant \theta_2$*,*

$$P(|\delta_1(X) - \theta_1| \leqslant t) > P(|X_1 - \theta_1| \leqslant t).$$

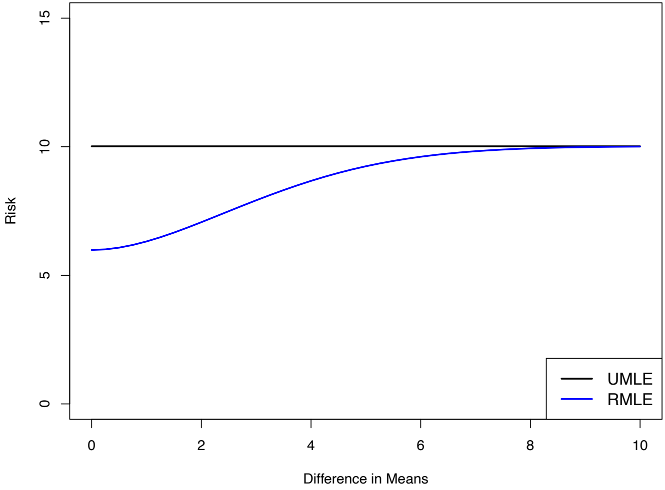
   Thus, $\delta_1(X)$ *dominates* X_1 *for any loss function* $L(x) = \rho(|x|)$*, where* $\rho : \mathbb{R}_+ \to \mathbb{R}_+$ *is non-decreasing.*

2. *If* $\sigma_1^2 > \sigma_{12}$*,* $\delta_1(X)$ *is biased downwards. For* $a < 0$*,* $b > 0$*,*

$$P(a \leqslant \delta_1(X) - \theta_1 \leqslant b) > P(a \leqslant X_1 - \theta_1 \leqslant b)$$

   holds for all $\theta_1 \leqslant \theta_2$ *if and only if* $a + b \geqslant 0$*. Thus,* $\delta_1(X)$ *dominates* X_1 *for* $L(x) = \rho_1(|x|)\mathbf{1}(x > 0) + \rho_2(|x|)\mathbf{1}(x \leqslant 0)$*, where* $\rho_j : \mathbb{R}_+ \to \mathbb{R}_+$ *are non-decreasing and* $\rho_1(t) \geqslant \rho_2(t)$*.*

Risk Profiles



Note: $\Sigma = \begin{bmatrix} 10 & 1 \\ 1 & 2 \end{bmatrix}$; quadratic loss.

Discussion

- Lemma 1 implies UMLE is inadmissible for many loss functions.
- Also implies that RMLE is minimax.
- van Eden and Zidek (2002) show RMLE itself is inadmissible under quadratic loss if Σ is diagonal, but:
  - The dominators are very complicated / unusable for inference;
  - Risk improvement over RMLE is marginal;
- Room for extensions in higher dimensions:
  - For diagonal Σ, completely solved by Kelly (1989);
  - For certain non-diagonal Σ, Hwang, Peddada (1994) show RMLE behaves unexpectedly in dimensions eight and higher;
  - For specific Σ, estimators dominating UMLE have been proposed;

Extensions

- Consider an experiment $X \sim N(\theta, \Sigma)$ in $\mathbb{R}^k$ with a target parameter:

$$\mu_1 = t'\theta$$

  and a single inequality constraint:

$$c'\theta \geqslant b.$$

- Idea: can reformulate the problem as as

$$Y \sim N(\mu, \sigma^2 I) \in \mathbb{R}^2$$

  with $\mu = (\mu_1, \mu_2)$ and $a'\mu \geqslant 0$.
- Can still show universal dominance of RMLE over MLE!

Outline

Motivation and Literature

Estimation in the Normal Experiment

Inference in the Normal Experiment

Asymptotic Results

Empirical Applications

RMLE for Inference

- Testing $H_0 : \theta_1 = t$ against $H_1 : \theta_1 \neq t$.
- The standard t-test with acceptance region

$$\{(x_1, x_2) \in \mathbb{R}^2 : |x_1 - t| \leqslant z_{1-\alpha/2}\sigma_1\}$$

is still UMP unbiased but ignores the implications of Lemma 1.

RMLE for Inference

- Testing $H_0 : \theta_1 = t$ against $H_1 : \theta_1 \neq t$.
- The standard t-test with acceptance region

$$\{(x_1, x_2) \in \mathbb{R}^2 : |x_1 - t| \leqslant z_{1-\alpha/2}\sigma_1\}$$

  is still UMP unbiased but ignores the implications of Lemma 1.
- Instead, consider a test with acceptance region:

$$\{(x_1, x_2) \in \mathbb{R}^2 : |\delta_1(x) - t| \leqslant c(x)\}.$$

  By Lemma 1, $c(x) = z_{1-\alpha/2}\sigma_1$ yields a valid but conservative test.

RMLE for Inference

- Testing $H_0 : \theta_1 = t$ against $H_1 : \theta_1 \neq t$.
- The standard t-test with acceptance region

$$\{(x_1, x_2) \in \mathbb{R}^2 : |x_1 - t| \leqslant z_{1-\alpha/2}\sigma_1\}$$

  is still UMP unbiased but ignores the implications of Lemma 1.
- Instead, consider a test with acceptance region:

$$\{(x_1, x_2) \in \mathbb{R}^2 : |\delta_1(x) - t| \leqslant c(x)\}.$$

  By Lemma 1, $c(x) = z_{1-\alpha/2}\sigma_1$ yields a valid but conservative test.
- **Problem:** the distribution of $\delta_1(X) - t$ under H_0 is the same as:

$$g(Z; h) = Z_1 + w_2 \min(Z_2 - Z_1 + h, 0),$$

  where $Z \sim N(0, \Sigma)$ in $\mathbb{R}^2$, $h = \theta_2 - \theta_1$, and $w_2 = \frac{\sigma_1^2 - \sigma_{12}}{\sigma_1^2 + \sigma_2^2 - 2\sigma_{12}}$.

Nuisance Parameters Under the Null

- Well-known approaches:
  - Least-favorable tests;
  - Conditional tests: Moreira (2003); Ketz (2018);
  - Pretesting: Romano, Shaikh, Wolf (2014);
  - Max-WAP: Elliott, Müller, Watson (2015), Montiel Olea (2018);

Nuisance Parameters Under the Null

- Well-known approaches:
  - Least-favorable tests;
  - Conditional tests: Moreira (2003); Ketz (2018);
  - Pretesting: Romano, Shaikh, Wolf (2014);
  - Max-WAP: Elliott, Müller, Watson (2015), Montiel Olea (2018);
- **Proposed solution:**
  - A test with acceptance region $\{x : |\delta_1(x) - t| \leqslant c(x)\}$
  - Setting $c = Q_{1-\alpha}(|g(Z; h)|)$ would give a valid test;
  - The distribution of $|g(Z; h)|$ is stochastically increasing in h;
  - Choose data-dependent $h(X)$ to match the desired level;

The Test

- Set:

$$h_{t,\kappa}(X) = X_2 - X_1 + \tfrac{\sigma_1^2 - \sigma_{12}}{\sigma_1^2}(X_1 - t) + \kappa,$$

where $\kappa \geqslant 0$ is the smallest value such that

$$\sup_{h \geqslant 0} P\left\{|g(Z,h)| \geqslant Q_{1-\alpha}^{\tilde{Z}|Z}\left(|g(\tilde{Z}; s_\kappa(Z))|\right)\right\} \leqslant \alpha,$$

for $Z, \tilde{Z}$ independent $N(0, \Sigma)$ and

$$s_\kappa(Z) = \max(Z_2 - Z_1 + h + \tfrac{\sigma_1^2 - \sigma_{12}}{\sigma_1^2} Z_1 + \kappa, 0).$$

- Define:

$$\phi^{RMLE}(X;t) = \mathbf{1}\left\{\ |\delta_1(X) - t| > Q_{1-\alpha}^{Z|X}(|g(Z, h_{t,\kappa}(X) \vee 0)|)\ \right\}.$$

Two-Sided Tests with RMLE

Lemma 2 (Proposed Test)

Consider testing $H_0 : \theta_1 = t$ against $H_1 : \theta_1 \neq t$. Let $h_{t,\kappa}(X)$ and κ be as defined above. Then, the test $\phi^{RMLE}(X; t)$ is level-α.

Two-Sided Tests with RMLE

Lemma 2 (Proposed Test)

Consider testing $H_0 : \theta_1 = t$ *against* $H_1 : \theta_1 \neq t$. *Let* $h_{t,\kappa}(X)$ *and* κ *be as defined above. Then, the test* $\phi^{RMLE}(X;t)$ *is level-*α.

Lemma 3 (Conditional LR Test)

Consider testing $H_0 : \theta_1 = t$ *against* $H_1 : \theta_1 \neq t$. *The test:*

$$\phi^{CLR}(X) = \mathbf{1}\{|\delta_1(X) - t| > Q_{1-\alpha}^{Z|X}(|g(Z; h_{t,0}(X))|)\},$$

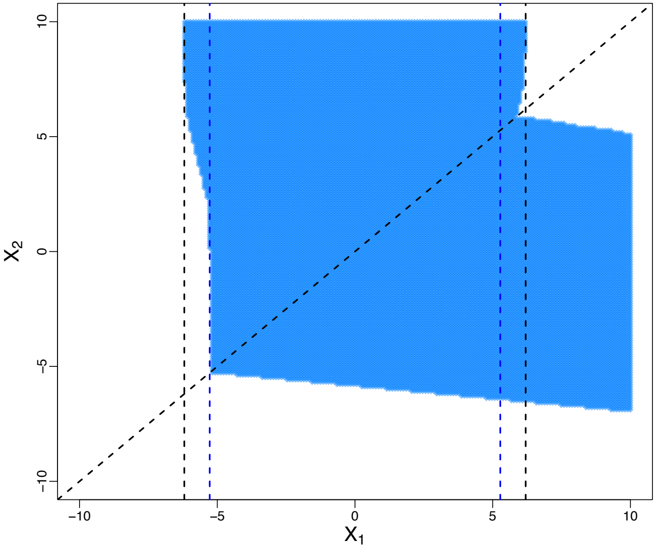
where $Z = (Z_1, \frac{\sigma_{12}-\sigma_1^2}{\sigma_1^2} Z_1)$ *with* $Z_1 \sim N(0, \sigma_1^2)$, *is level-*$\alpha$.

Discussion

- Both tests are biased w.r.t. alternatives $\theta_1 > t$.
- Both CIs centered around $\delta_1(X)$; not nested with $[X_1 \pm z_{1-\alpha/2}\sigma_1]$.
- RMLE CI is shorter than $2z_{1-\alpha/2}\sigma_1$ with probability one.
- CLR CI may not be.
- Parameter κ for RMLE is easy to compute numerically.

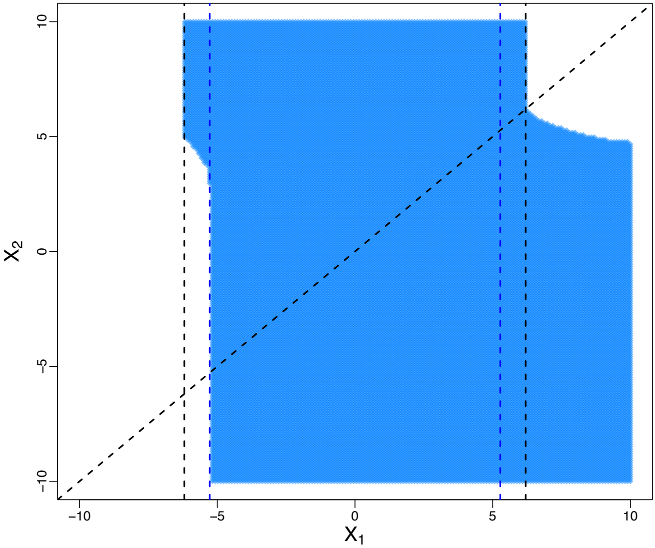
RMLE Acceptance Region

Testing $H_0 : \theta_1 = 0$ vs $H_1 : \theta_1 \neq 0$

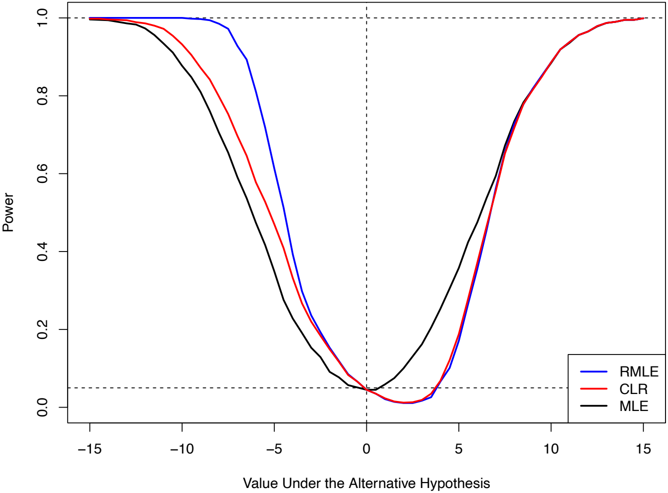


CLR Acceptance Region

Testing $H_0 : \theta_1 = 0$ vs $H_1 : \theta_1 \neq 0$

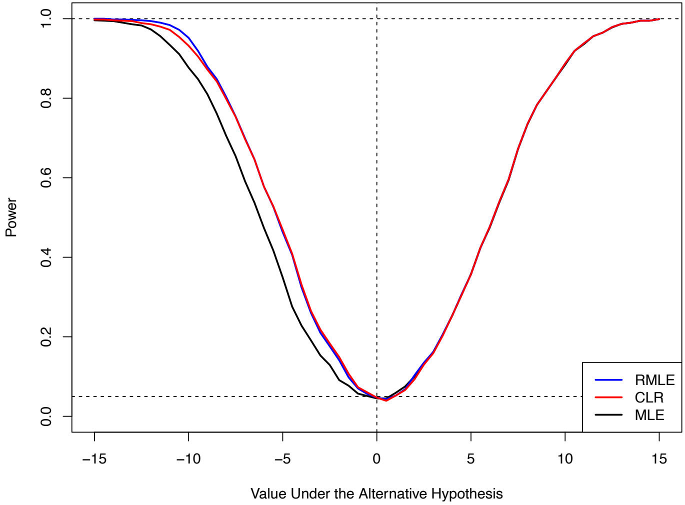


Power Comparisons



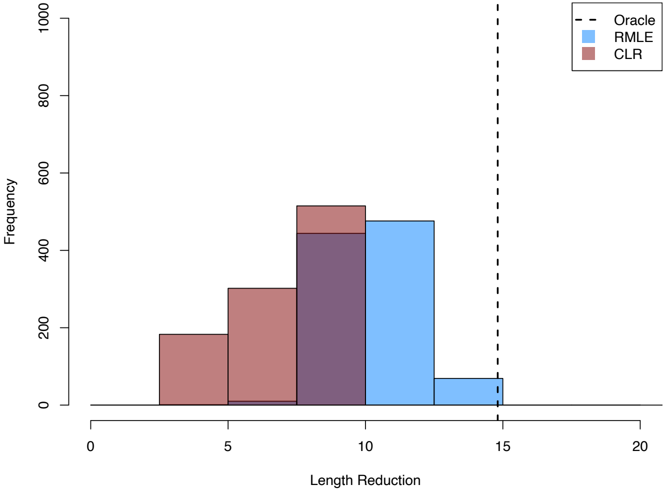
Note: $\Sigma = \begin{bmatrix} 10 & 1 \\ 1 & 2 \end{bmatrix}; \quad h = 0.$

Power Comparisons



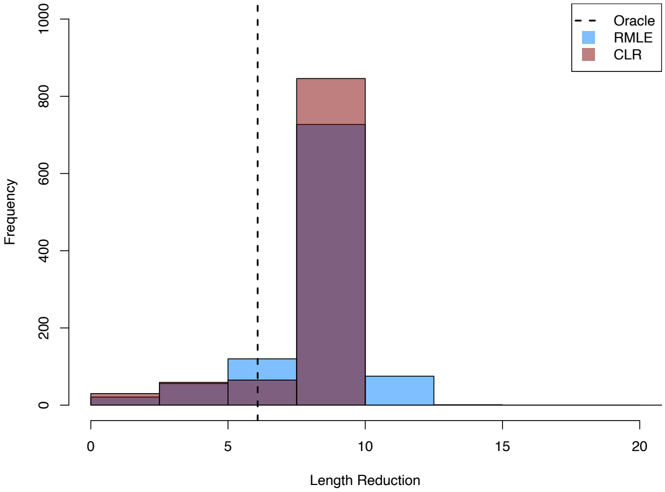
Note: $\Sigma = \begin{bmatrix} 10 & 1 \\ 1 & 2 \end{bmatrix}; \quad h = 5.$

% Reduction of CI Length



Note: Σ as above, $h = 0$. Median for RMLE is 11%, max is 14%.

% Reduction of CI Length



Note: Σ as above, h $=$ 5. Median for RMLE is 8%, max is 13%.

Outline

Motivation and Literature

Estimation in the Normal Experiment

Inference in the Normal Experiment

Asymptotic Results

Empirical Applications

Setup

Assumption 1

The distribution P of the data belongs to the set:

$$\mathbf{P} = \{P \in \mathcal{M} : \beta(P) \leqslant \gamma(P)\},$$

where $\mathcal{M}$ is defined by suitable moment and support restrictions.

Assumption 2

There is a sequence of regular estimators $(\hat{\beta}_n, \hat{\gamma}_n)$ satisfying

$$\begin{bmatrix} \sqrt{n}(\hat{\beta}_n - \beta) \\ \sqrt{n}(\hat{\gamma}_n - \gamma) \end{bmatrix} \overset{d}{\to} N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_\beta^2 & \sigma_{\beta\gamma} \\ \sigma_{\beta\gamma} & \sigma_\gamma^2 \end{bmatrix} \right), \tag{1}$$

with non-degenerate covariance matrix Σ, and a sequence of estimators $\hat{\Sigma}_n$ satisfying $\hat{\Sigma}_n \to_p \Sigma$, for each $P \in \mathbf{P}$.

Setup

- The estimator:

$$
\begin{aligned}
\hat{\delta}_n &= \hat{\beta}_n \mathbf{1}(\hat{\beta}_n \leqslant \hat{\gamma}_n) + (\hat{w}_{1,n}\hat{\beta}_n + \hat{w}_{2,n}\hat{\gamma}_n)\mathbf{1}(\hat{\beta}_n > \hat{\gamma}_n) \\
&= \hat{\beta}_n + \hat{w}_{2,n} \min(\hat{\gamma}_n - \hat{\beta}_n, 0),
\end{aligned}
$$

  where $\hat{w}_{2,n} = (\hat{\sigma}^2_{\beta,n} - \hat{\sigma}_{\beta\gamma,n})/(\hat{\sigma}^2_{\beta,n} + \hat{\sigma}^2_{\gamma,n} - 2\hat{\sigma}_{\beta\gamma,n})$.

- Local asymptotics: let $P_{n,h} \in \mathbf{P}$ be a sequence of distributions local to P such that $\sqrt{n}(\gamma(P_{n,h}) - \beta(P_{n,h})) \to h$ for some $h \geqslant 0$.

Asymptotic Properties: Estimation

Theorem 1

Let Assumptions 1 and 2 hold and $P_{n,h}$ as above. Then:

1. *Along $P_{n,h}$:*

$$\sqrt{n}(\hat{\delta}_n - \beta_{n,h}) \quad \to_d \quad Z_1 + w_2 \min(Z_1 + Z_2 + h, 0),$$

   where $Z = (Z_1, Z_2) \sim N(0, \Sigma)$.

2. *For any $h > 0, t > 0$, there exists $N(t,h) \in \mathbb{N}$ such that:*

$$P_{n,h}(|\sqrt{n}(\hat{\delta}_n - \beta_{n,h})| \leqslant t) > P_{n,h}(|\sqrt{n}(\hat{\beta}_n - \beta_n)| \leqslant t)$$

   for all $n \geqslant N(t,h)$.

Asymptotic Properties: Inference

Theorem 2

Let Assumptions 1 and 2 hold, $H_0 : \beta(P) = \beta_0$ and $H_1 : \beta(P) \neq \beta_0$. Define:

$$\bar{h}_n = \sqrt{n}(\hat{\gamma}_n - \hat{\beta}_n) + \frac{\hat{\sigma}^2_{\beta,n} - \hat{\sigma}_{\beta\gamma,n}}{\hat{\sigma}^2_{\beta,n}}\sqrt{n}(\hat{\beta}_n - \beta_0) + \hat{\kappa}_n,$$

where $\hat{\kappa}_n \geqslant 0$ is the smallest value such that:

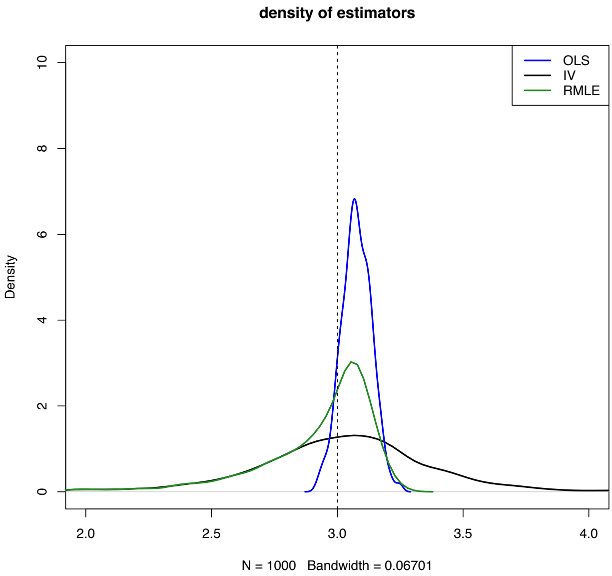
$$\sup_{h \geqslant 0} P\left(|g(Z,h)| \geqslant Q^{\tilde{Z}|Z}_{1-\alpha}(|g(\tilde{Z}; s_\kappa(Z))|)\right) \leqslant \alpha,$$

with $Z, \tilde{Z}$ i.i.d. $N(0, \hat{\Sigma}_n)$, conditional on the data. The test:

$$\hat{\phi}_n(t) = \mathbf{1}(\sqrt{n}|\hat{\delta}_n - \beta_0| \geqslant Q^{Z|W_1^n}_{1-\alpha}(|g(Z; \bar{h}_n \vee 0)|)),$$

is locally asymptotically level-α.

A Snapshot from Simulations



Outline

Motivation and Literature

Estimation in the Normal Experiment

Inference in the Normal Experiment

Asymptotic Results

Empirical Applications

Criminal Recidivism and Electronic Monitoring

Di Tella and Schargorodsky (2013)

- Argue that using EM instead of prison leads to lower recidivism.
- Using individual-level data, estimate:

$$R_i = \alpha + \beta D_i + X_i'\eta + \varepsilon_i$$

  - R_i — indicator for recidivism;
  - D_i — indicator for being assigned an EM;
  - X_i — {type of crime, age, nationality, time dummies, ...}

Criminal Recidivism and Electronic Monitoring

Di Tella and Schargorodsky (2013)

- Argue that using EM instead of prison leads to lower recidivism.
- Using individual-level data, estimate:

$$R_i = \alpha + \beta D_i + X_i'\eta + \varepsilon_i$$

  - R_i — indicator for recidivism;
  - D_i — indicator for being assigned an EM;
  - X_i — {type of crime, age, nationality, time dummies, ...}

- **Endogeneity bias:** the "bad guys" are less likely to get out with an EM $\implies$ OLS is biased upwards.
- **Instrument:** judge characteristics.

Criminal Recidivism and Electronic Monitoring

Results

- Re-running the main specification yields:

$$\begin{bmatrix} \hat{\beta}_n^{IV} \\ \hat{\gamma}_n^{OLS} \end{bmatrix} = \begin{bmatrix} -0.137 \\ -0.090 \end{bmatrix}; \;\; \hat{\Sigma}_n = \begin{bmatrix} 4.965 & 0.597 \\ 0.597 & 0.410 \end{bmatrix},$$

  with $n = 1503$. Expect $\beta \leqslant \gamma$ and get $\hat{\beta}_n^{IV} \leqslant \hat{\gamma}_n^{OLS}$.
- The implied 95% IV confidence interval is $[-0.245, -0.020]$.

Criminal Recidivism and Electronic Monitoring

- Re-running the main specification yields:

$$\begin{bmatrix} \hat{\beta}_n^{IV} \\ \hat{\gamma}_n^{OLS} \end{bmatrix} = \begin{bmatrix} -0.137 \\ -0.090 \end{bmatrix}; \quad \hat{\Sigma}_n = \begin{bmatrix} 4.965 & 0.597 \\ 0.597 & 0.410 \end{bmatrix},$$

  with $n = 1503$. Expect $\beta \leqslant \gamma$ and get $\hat{\beta}_n^{IV} \leqslant \hat{\gamma}_n^{OLS}$.
- The implied 95% IV confidence interval is $[-0.245, -0.020]$.
- $\hat{\delta}_n^{RMLE} = -0.137$ produces $[-0.245, -0.040]$, which is 9% shorter.
- Despite the large sample size, good IVs, and the direction of the bias as expected, still see an improvement.

Mechanics of Industrial Revolution

Kelly, Mokyr, Ó Gráda (2023)

- Argue that different patterns of growth across counties of England in 1760s–1830s can be explained by pre-existing supply of mechanical skills.
- Using county-level data, estimate:

$$Y_i = \alpha + D_i\beta + X_i'\eta + \varepsilon_i$$

  - Y_i — share of men working in textiles in 1831 (or 1851);
  - D_i — existing supply of mechanical skills based on 1790s;
  - X_i — average wages in 1760s and market potential in 1750s;

Mechanics of Industrial Revolution

Kelly, Mokyr, Ó Gráda (2023)

- Argue that different patterns of growth across counties of England in 1760s–1830s can be explained by pre-existing supply of mechanical skills.
- Using county-level data, estimate:

$$Y_i = \alpha + D_i\beta + X_i'\eta + \varepsilon_i$$

  - Y_i — share of men working in textiles in 1831 (or 1851);
  - D_i — existing supply of mechanical skills based on 1790s;
  - X_i — average wages in 1760s and market potential in 1750s;
- **Endogeneity bias:** new textile industries may have attracted skilled workers, not vice versa $\Longrightarrow$ OLS is biased upwards.
- **Instrument:** apprentice fees.

Mechanics of Industrial Revolution

Results

- Re-running the main specification yields:

$$\begin{bmatrix} \hat{\beta}_n^{IV} \\ \hat{\gamma}_n^{OLS} \end{bmatrix} = \begin{bmatrix} 3.178 \\ 2.125 \end{bmatrix}; \;\; \hat{\Sigma}_n = \begin{bmatrix} 72.762 & 8.448 \\ 8.448 & 7.577 \end{bmatrix},$$

  with $n = 41$. Expect $\beta \leqslant \gamma$ but get $\hat{\beta}_n^{IV} > \hat{\gamma}_n^{OLS}$.

- The implied 95% IV confidence interval is $[0.566, 5.789]$.

Mechanics of Industrial Revolution

Results

- Re-running the main specification yields:

$$\begin{bmatrix} \hat{\beta}_n^{IV} \\ \hat{\gamma}_n^{OLS} \end{bmatrix} = \begin{bmatrix} 3.178 \\ 2.125 \end{bmatrix}; \;\; \hat{\Sigma}_n = \begin{bmatrix} 72.762 & 8.448 \\ 8.448 & 7.577 \end{bmatrix},$$

  with $n = 41$. Expect $\beta \leqslant \gamma$ but get $\hat{\beta}_n^{IV} > \hat{\gamma}_n^{OLS}$.

- The implied 95% IV confidence interval is $[0.566, 5.789]$.
- $\hat{\delta}_n^{RMLE} = 2.11$ yields $[-0.309, 4.299]$, which is 12% shorter and includes zero.
- Similar results for 1851 specification.

Learning by Doing

Levitt, List, Syverson (2013)

- Document learning by doing at an automobile assembly plant.
- Using daily and weekly data, estimate:

$$\log S_t = \alpha + \beta \log E_t + \varepsilon_t$$

  - S_t — quality of produced cars;
  - E_t — production experience at time t.

Learning by Doing

Levitt, List, Syverson (2013)

- Document learning by doing at an automobile assembly plant.
- Using daily and weekly data, estimate:

$$\log S_t = \alpha + \beta \log E_t + \varepsilon_t$$

  - S_t — quality of produced cars;
  - E_t — production experience at time t.

- **Endogeneity bias:** manager steers production towards the higher-quality shift $\implies$ OLS is biased downwards.
- **Instrument:** experience of the other shift.

Learning by Doing

Results

- Re-running the main specification using weekly data yields:

$$\begin{bmatrix} \hat{\beta}_n^{IV} \\ \hat{\gamma}_n^{OLS} \end{bmatrix} = \begin{bmatrix} -0.282 \\ -0.267 \end{bmatrix}; \;\; \hat{\Sigma}_n = \begin{bmatrix} 0.065 & 0.055 \\ 0.055 & 0.047 \end{bmatrix},$$

  with $n = 39$. Expect $\beta \geqslant \gamma$ but get $\hat{\beta}_n^{IV} < \hat{\gamma}_n^{OLS}$.
- The implied 95% IV confidence interval is $[-0.363, -0.203]$.

Learning by Doing

Results

- Re-running the main specification using weekly data yields:

$$\begin{bmatrix} \hat{\beta}_n^{IV} \\ \hat{\gamma}_n^{OLS} \end{bmatrix} = \begin{bmatrix} -0.282 \\ -0.267 \end{bmatrix}; \quad \hat{\Sigma}_n = \begin{bmatrix} 0.065 & 0.055 \\ 0.055 & 0.047 \end{bmatrix},$$

  with $n = 39$. Expect $\beta \geqslant \gamma$ but get $\hat{\beta}_n^{IV} < \hat{\gamma}_n^{OLS}$.

- The implied 95% IV confidence interval is $[-0.363, -0.203]$.
- $\hat{\delta}_n^{RMLE} = -0.227$ produces $[-0.295, -0.156]$, which is 13% shorter, and the difference is economically significant:
  - $\hat{\beta}_n^{IV}$ implies the rate of defects is halved for every 12-fold increase in production;
  - $\hat{\delta}_n^{RMLE}$ implies that for 21-fold;
- Similar results using daily data with $n = 190$ observations.

Summary

- Knowing the direction of endogeneity bias can meaningfully improve statistical analysis.
- This paper proposed a simple way to construct more precise estimates and shorter confidence intervals.
- Obtained 9-13% shorter CIs in three empirical applications.
- Further research: minimum expected length CI? extensions to multiple inequalities? applications to partial identification?