Robust Deconfounding with Instrumental Variables

Zijian Guo

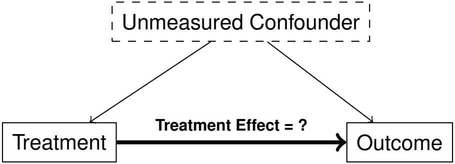
Department of Statistics, Rutgers University

Guo, Z. (2021). Causal Inference with Invalid Instruments: Post-selection Problems and A Solution Using Searching and Sampling. *arXiv preprint arXiv:2104.06911*.

Confounding and Causal Effects

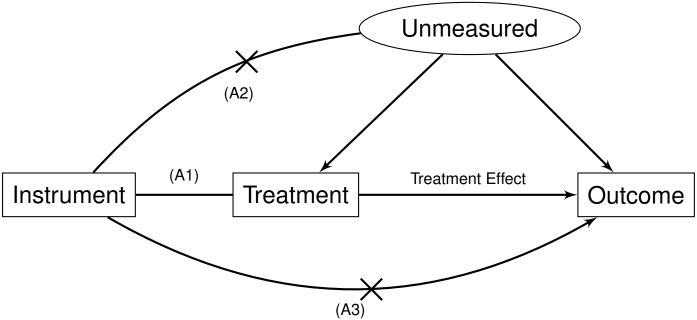
“Confounding is one of the most fundamental impediments to the elucidation of causal inferences from empirical data." (Pearl, 2009)

Causal inference with unmeasured confounders “remains a fertile field for methodological research." (Imbens, Rubin, 2015)



- Observational study: unmeasured confounders
- OLS or Lasso is biased.
- Construction of instrumental variables.

Instrumental Variables



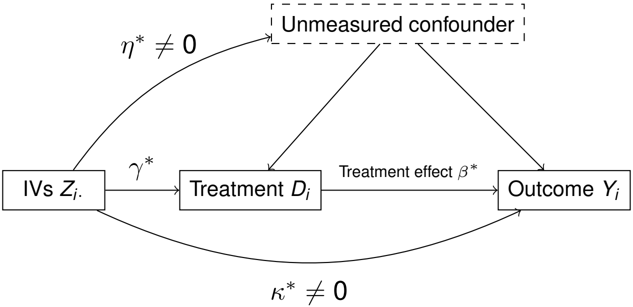
- Educ's effect on salary (Card, 1995): family background
- **Instrumental variable (IV)**: proximity to college
- Mendelian Randomization: genetic variants as IVs.

Robust Inference with Possibly Invalid IVs

A subset of existing literature:

1. Orthogonality: Kolesár, Chetty, Friedman, Glaeser, and Imbens (2015); Bowden, Smith, and Burgess (2015).
2. Majority/plurality rule: Bowden, Smith, Haycock, and Burgess (2016); Kang, Zhang, Cai, and Small (2016); G., Kang, Cai, and Small (2018); Windmeijer, Farbmacher, Davies, and Smith (2019); Windmeijer, Liang, Hartwig, and Bowden (2021).
3. Heteroscedastic variance+homoscadastic correlation: Lewbel (2012); Tchetgen Tchetgen, Sun, and Walter (2021).

Models and Research Problem



$$Y_i = D_i\beta^* + Z_{i\cdot}^{\intercal}\pi^* + X_{i\cdot}^{\intercal}\phi^* + e_i \quad \text{with} \quad \pi^* = \kappa^* + \eta^* \in \mathbb{R}^{p_z}$$

$$D_i = Z_{i\cdot}^{\intercal}\gamma^* + X_{i\cdot}^{\intercal}\psi^* + \delta_i$$

Inference for β^* when $\pi^* \neq 0$.

Identification and Post-Selection

Reduced Form Parameters

The reduced form model is

$$\begin{aligned} Y_i = & \quad Z_{i\cdot}^{\intercal}\Gamma^* + X_{i\cdot}^{\intercal}\Psi^* + \epsilon_i \quad \text{with} \quad \mathbb{E}(Z_{i\cdot}\epsilon_i) = \mathbf{0}, \ \mathbb{E}(X_{i\cdot}\epsilon_i) = \mathbf{0}, \\ D_i = & \quad Z_{i\cdot}^{\intercal}\gamma^* + X_{i\cdot}^{\intercal}\psi^* + \delta_i \quad \text{with} \quad \mathbb{E}(Z_{i\cdot}\delta_i) = \mathbf{0}, \ \mathbb{E}(X_{i\cdot}\delta_i) = \mathbf{0}, \end{aligned}$$

where $\epsilon_i = \beta^*\delta_i + e_i$ and

$$\Gamma^* = \beta^*\gamma^* + \pi^* \in \mathbb{R}^{p_z}.$$

Construct unbiased estimators $\widehat{\Gamma}$ and $\widehat{\gamma}$ satisfying

$$\sqrt{n}\begin{pmatrix} \widehat{\Gamma} - \Gamma^* \\ \widehat{\gamma} - \gamma^* \end{pmatrix} \to N(\mathbf{0}, \mathrm{Cov}) \quad \text{with} \quad \mathrm{Cov} = \begin{pmatrix} \mathbf{V}^{\Gamma} & \mathbf{C} \\ \mathbf{C}^{\intercal} & \mathbf{V}^{\gamma} \end{pmatrix}$$

Reduced Form Equations

Estimate $\mathcal{S} = \{1 \leq j \leq p_z : \gamma_j^* \neq 0\}$ by hard-thresholding

$$\widehat{\mathcal{S}} = \left\{1 \leq j \leq p_z : |\widehat{\gamma}_j| \geq \sqrt{\log n} \cdot \sqrt{\widehat{\mathbf{V}}^{\gamma}_{jj}/n}\right\}$$

$$\widehat{\Gamma}_j = \beta^* \widehat{\gamma}_j + \pi_j^* \quad \text{for} \quad j \in \widehat{\mathcal{S}}.$$

Majority Rule: More than half of the relevant IVs are valid, that is,

$$|\mathcal{V}| > |\mathcal{S}|/2 \quad \text{with} \quad \mathcal{V} = \{j \in \mathcal{S} : \pi_j^* = 0\}.$$

Bowden, Smith, Haycock, and Burgess (2016); Kang, Zhang, Cai, and Small (2016)

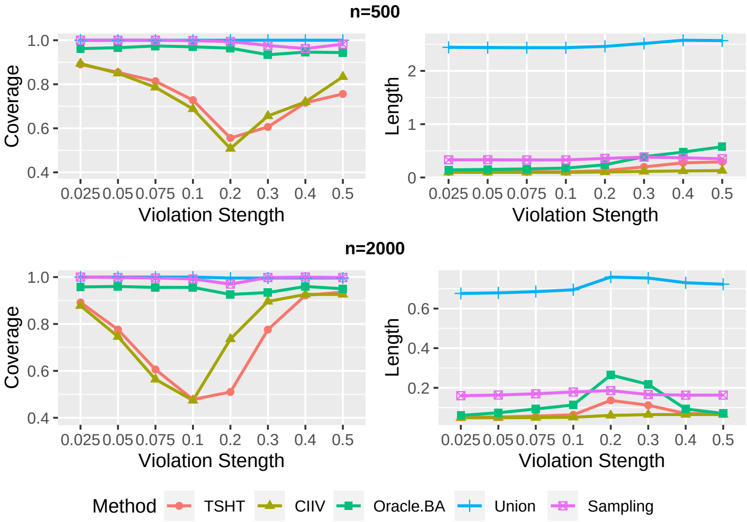
Post-selection Problem

- Both TSHT and CIIV leverage majority rule to select valid IVs.
- Both TSHT and CIIV rely on $\widehat{\mathcal{V}} = \mathcal{V}$.
- It is challenging to detect locally invalid IVs

$$\left\{j \in \mathcal{S} : 0 < \left|\pi_j^*/\gamma_j^*\right| \leq \tau_n\right\} \quad \text{with} \quad \tau_n \asymp \sqrt{\log n/n}.$$

Guo, Kang, Cai, & Small. (2018). Confidence intervals for causal effects with invalid instruments by using two-stage hard thresholding with voting. JRSSB.

Windmeijer, Liang, Hartwig, & Bowden. (2021). The confidence interval method for selecting valid instrumental variables. JRSSB.



How to construct uniformly valid CIs?

Searching (Majority Rule)

$$\widehat{\Gamma}_j = \beta^* \widehat{\gamma}_j + \pi_j^* \quad \text{for} \quad j \in \widehat{\mathcal{S}}.$$

For any β, we estimate π_j^* by

$$\left(\widehat{\Gamma}_j - \beta \widehat{\gamma}_j\right) - \pi_j^* = \underbrace{\widehat{\Gamma}_j - \Gamma_j^* - \beta(\widehat{\gamma}_j - \gamma_j^*)}_{\text{Estimation Error}} + \underbrace{(\beta^* - \beta)\gamma_j^*}_{\text{Bias}}.$$

$$\mathbb{P}\left(\max_{j \in \widehat{\mathcal{S}}} \frac{|\widehat{\Gamma}_j - \Gamma_j^* - \beta^*(\widehat{\gamma}_j - \gamma_j^*)|}{\sqrt{(\widehat{\mathbf{V}}^{\Gamma}_{jj} + [\beta^*]^2 \widehat{\mathbf{V}}^{\gamma}_{jj} - 2\beta^* \widehat{\mathbf{C}}_{jj})/n}} \leq \Phi^{-1}\left(1 - \frac{\alpha}{2|\widehat{\mathcal{S}}|}\right)\right) \geq 1 - \alpha.$$

Threshold for the estimation error

$$\widehat{\rho}_j(\beta) = \Phi^{-1}\left(1 - \frac{\alpha}{2|\widehat{\mathcal{S}}|}\right) \cdot \sqrt{(\widehat{\mathbf{V}}^{\Gamma}_{jj} + \beta^2 \widehat{\mathbf{V}}^{\gamma}_{jj} - 2\beta \widehat{\mathbf{C}}_{jj})/n},$$

Estimate π_j^* by

$$\widehat{\pi}_j(\beta) = \left(\widehat{\Gamma}_j - \beta\widehat{\gamma}_j\right) \cdot \mathbf{1}\left(\left|\widehat{\Gamma}_j - \beta\widehat{\gamma}_j\right| \geq \widehat{\rho}_j(\beta)\right) \quad \text{for} \quad j \in \widehat{\mathcal{S}}. \tag{1}$$

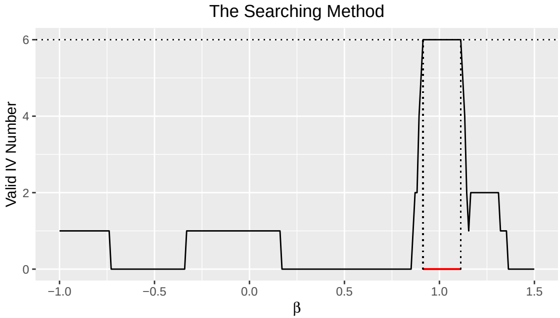
$$\mathrm{CI}^{\mathrm{search}} = \left(\min_{\{\beta \in \mathbf{R}: \|\widehat{\pi}_{\widehat{\mathcal{S}}}(\beta)\|_0 < |\widehat{\mathcal{S}}|/2\}} \beta, \max_{\{\beta \in \mathbf{R}: \|\widehat{\pi}_{\widehat{\mathcal{S}}}(\beta)\|_0 < |\widehat{\mathcal{S}}|/2\}} \beta\right). \tag{2}$$

- Accept β if the number of invalid IVs $\|\widehat{\pi}_{\widehat{\mathcal{S}}}(\beta)\|_0$ is less than $|\widehat{\mathcal{S}}|/2$

Partial checking of majority rule: if no β satisfies $\|\widehat{\pi}_{\widehat{\mathcal{S}}}(\beta)\|_0 < |\widehat{\mathcal{S}}|/2$, we set $\mathrm{CI}^{\mathrm{search}} = \varnothing$.

Example 1

set $\gamma^* = 0.5 \cdot \mathbf{1}_{10}$, $\pi^* = (\mathbf{0}_6, 0.05, 0.05, -0.5, -1)^\intercal$, $\beta^* = 1$ and $n = 2000$, The majority rule is satisfied. $\mathrm{CI}^{\mathrm{search}} = (0.931, 1.099)$.



The proposed searching CI is related to the Anderson-Rubin test

Sampling (Majority Rule)

Sampling Distribution

Recall that the estimators $\widehat{\Gamma}$ and $\widehat{\gamma}$ satisfy

$$\sqrt{n}\begin{pmatrix}\widehat{\Gamma}-\Gamma^{*}\\ \widehat{\gamma}-\gamma^{*}\end{pmatrix}\to N\left(\mathbf{0},\mathrm{Cov}\right)\quad\text{with}\quad \mathrm{Cov}=\begin{pmatrix}\mathbf{V}^{\Gamma} & \mathbf{C}\\ \mathbf{C}^{\intercal} & \mathbf{V}^{\gamma}\end{pmatrix}$$

Conditioning on the observed data, sample $\{\widehat{\Gamma}^{[m]},\widehat{\gamma}^{[m]}\}_{1\leq m\leq M}$ following

$$\begin{pmatrix}\widehat{\Gamma}^{[m]}\\ \widehat{\gamma}^{[m]}\end{pmatrix}\stackrel{\text{i.i.d.}}{\sim} N\left[\begin{pmatrix}\widehat{\Gamma}\\ \widehat{\gamma}\end{pmatrix},\begin{pmatrix}\widehat{\mathbf{V}}^{\Gamma}/n & \widehat{\mathbf{C}}/n\\ \widehat{\mathbf{C}}^{\intercal}/n & \widehat{\mathbf{V}}^{\gamma}/n\end{pmatrix}\right]\quad\text{for}\quad 1\leq m\leq M. \tag{3}$$

Sampling Property

There exists

$$1 \leq m^* \leq M \quad \text{and} \quad \lambda \asymp \left(\frac{\log n}{M}\right)^{\frac{1}{2|\widehat{\mathcal{S}}|}}$$

such that

$$\max_{j \in \widehat{\mathcal{S}}} \frac{|\widehat{\Gamma}_j^{[m^*]} - \Gamma_j^* - \beta(\widehat{\gamma}_j^{[m^*]} - \gamma_j^*)|}{\sqrt{(\widehat{\mathbf{V}}_{jj}^{\Gamma} + \beta^2 \widehat{\mathbf{V}}_{jj}^{\gamma} - 2\beta \widehat{\mathbf{C}}_{jj})/n}} \leq \lambda \cdot \Phi^{-1}\left(1 - \frac{\alpha}{2|\widehat{\mathcal{S}}|}\right).$$

Lower the threshold by a scale of λ!

"Sampled" Searching CI

Define the thresholding step for $1 \leq j \leq |\widehat{\mathcal{S}}|$:

$$\widehat{\pi}_j^{[m]}(\beta, \lambda) = \left(\widehat{\Gamma}_j^{[m]} - \beta\widehat{\gamma}_j^{[m]}\right) \cdot \mathbf{1}\left(\left|\widehat{\Gamma}_j^{[m]} - \beta\widehat{\gamma}_j^{[m]}\right| \geq \lambda \cdot \widehat{\rho}_j(\beta)\right),$$

“Sampled” Searching CI

Define the thresholding step for $1 \leq j \leq |\widehat{\mathcal{S}}|$:

$$\widehat{\pi}_j^{[m]}(\beta, \lambda) = \left(\widehat{\Gamma}_j^{[m]} - \beta\widehat{\gamma}_j^{[m]}\right) \cdot \mathbf{1}\left(\left|\widehat{\Gamma}_j^{[m]} - \beta\widehat{\gamma}_j^{[m]}\right| \geq \lambda \cdot \widehat{\rho}_j(\beta)\right),$$

For $1 \leq m \leq M$, construct “sampled" searching CI $(\beta_{\min}^{[m]}(\lambda), \beta_{\max}^{[m]}(\lambda))$

$$\beta_{\min}^{[m]}(\lambda) = \min_{\left\{\beta \in \mathcal{B} : \|\widehat{\pi}_{\widehat{\mathcal{S}}}^{[m]}(\beta,\lambda)\|_0 < |\widehat{\mathcal{S}}|/2\right\}} \beta, \quad \beta_{\max}^{[m]}(\lambda) = \max_{\left\{\beta \in \mathcal{B} : \|\widehat{\pi}_{\widehat{\mathcal{S}}}^{[m]}(\beta,\lambda)\|_0 < |\widehat{\mathcal{S}}|/2\right\}} \beta.$$

If no β satisfies $\|\widehat{\pi}^{[m]}(\beta, \lambda)\|_0 < |\widehat{\mathcal{S}}|/2$, we set $(\beta_{\min}^{[m]}(\lambda), \beta_{\max}^{[m]}(\lambda)) = \varnothing$.

Sampling CI

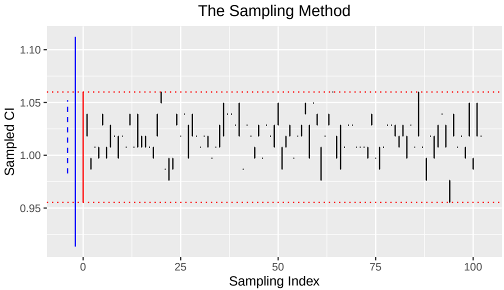


Figure: 102 of $M = 1000$ CIs are non-empty.

Define $\mathcal{M} = \{1 \leq m \leq M : (\beta_{\min}^{[m]}(\lambda), \beta_{\max}^{[m]}(\lambda)) \neq \varnothing\}$ and take a union,

$$\mathrm{CI}^{\mathrm{sample}} = \left(\min_{m \in \mathcal{M}} \beta_{\min}^{[m]}(\lambda), \max_{m \in \mathcal{M}} \beta_{\max}^{[m]}(\lambda) \right).$$

Plurality Rule (G., Kang, Cai, and Small, 2018)

$$|\mathcal{V}| > \max_{\nu \neq 0} |\mathcal{I}_\nu| \quad \text{with} \quad \mathcal{I}_\nu = \left\{ j \in \mathcal{S} : \pi_j^*/\gamma_j^* = \nu \right\}.$$

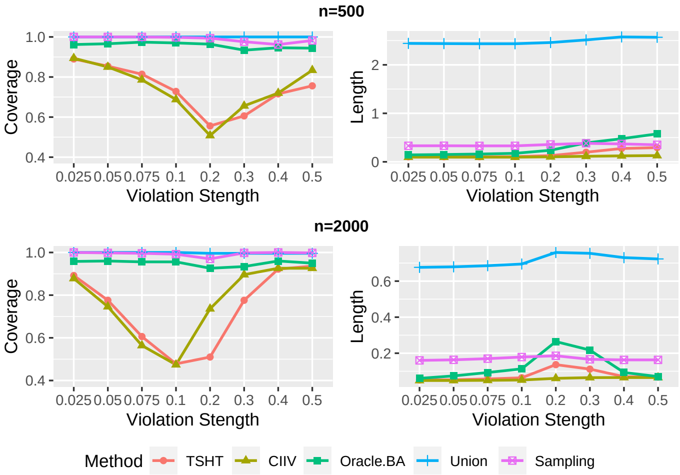
- Uniformly valid confidence intervals

Theoretical justification

- Uniform coverage, robust to locally invalid IVs.
- Parametric length

Numerical Analysis

Set $\gamma^* = 0.5 \cdot \mathbf{1}_{10}$ and $\pi^* = (\mathbf{0}_4, \tau/2, \tau/2, -\frac{1}{3}, -\frac{2}{3}, -1, -\frac{4}{3})^\intercal$;



Kang, H., Lee, Y., Cai, T. T., & Small, D. S. (2020). Two robust tools for inference about causal effects with invalid instruments. Biometrics.

The effect of education on earnings

China Family Panel Studies with 3758 observations.
Baseline covariates: gender, urban, hukou.
Include 9 IVs:

- the father's education level, the mother's education level, the spouse's education level, the family size, and the log transformation of the education expenditure in the past 12 months;
- the group-level years of education;
- the statement on the fair competition, the statement on the talent pay-off, and whether the subject reads some books or not in the past 12 months.

Method	CI	Method	CI
OLS	(0.0305, 0.0503)	Searching CI	(0.0409, 0.1698)
TSLS	(0.0959, 0.1190)	Sampling CI	(0.0552, 0.1268)
TSHT	(0.0946, 0.1178)	Union ($\bar{s} = p_z - 1$)	(-0.4915, 1.6043)
CIIV	(0.0948, 0.1175)	Union($\bar{s} = \lceil p_z/2 \rceil$)	(0.0409, 0.1342)

Table: Confidence intervals for the effect of education on earnings.

The family size is detected as the invalid IV.

Other direction 1: All Possibly Invalid IVs?

TSCI with Machine Learning

1. Fit the treatment model with machine learning
   - Random forests
   - Boosting
   - DNN
2. In the outcome model, use “Curvature" to test against violation.

Guo, Z. & Bühlmann, P. (2022). Two Stage Curvature Identification with Machine Learning: Causal Inference with Possibly Invalid Instrumental Variables. *arXiv preprint arXiv:2203.12808*.

Other direction 2: Many Treatments?

Instrumental variable methods

- Many treatments requires many IVs
- validity?
- efficiency?

Spectral deconfounding

- Dense confounding.
- Trimming the top singular values of many treatments.
- Efficient inference.

Guo, Z., Ćevid , D. & Bühlmann, P. (2022) Doubly Debiased Lasso: High-dimensional Inference under Hidden Confounding. The Annals of Statistics 50(3): 1320-1347.

Take Home Message

1. Robust deconfounding
   - IV selection under majority/plurality rule
   - Curvature identification (Guo, Bühlmann, 2022)
   - Spectral deconfounding (Ćevid, Meinshausen, Bühlmann, 2021; Guo, Ćevid, Bühlmann, 2022)
   - ...
2. Address the post-selection problems: Searching and Sampling

Robust Causal Inference with Hidden Confounding

Reference and Acknowledgement

Guo, Z. (2021). Post-selection Problems for Causal Inference with Invalid Instruments: A Solution Using Searching and Sampling. *arXiv preprint arXiv:2104.06911*.

Code is available at `https://github.com/zijguo/RobustIV`

Guo, Z. & Bühlmann, P. (2022). Two Stage Curvature Identification with Machine Learning: Causal Inference with Possibly Invalid Instrumental Variables. *arXiv preprint arXiv:2203.12808*.

Code is available at `https://github.com/zijguo/TSCI-Replication`

Guo, Z., Ćevid , D. & Bühlmann, P. (2022) Doubly Debiased Lasso: High-dimensional Inference under Hidden Confounding. The Annals of Statistics 50(3): 1320-1347.

Code is available at `https://github.com/zijguo/Doubly-Debiased-Lasso`

Acknowledgement to NSF and NIH for fundings.

Thank you!