Two Stage Curvature Identification with Machine Learning: Causal Inference with Possibly Invalid Instrumental Variables

Zijian Guo

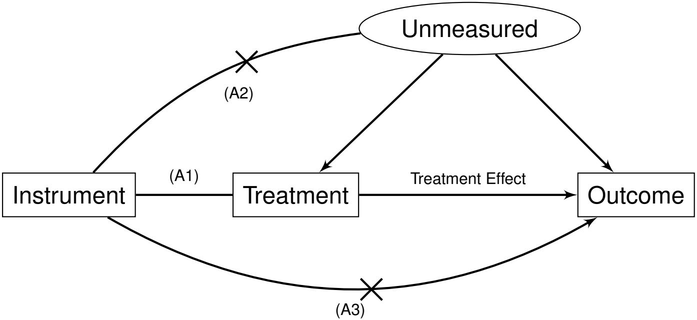
Rutgers University

Joint work with Peter Bühlmann

Overview of talk

Instrumental Variables

- Observational study: unmeasured confounders
- Educ's effect on salary (Card, 1995): family background
- Instrumental variable: proximity to college



Existing works (subset)

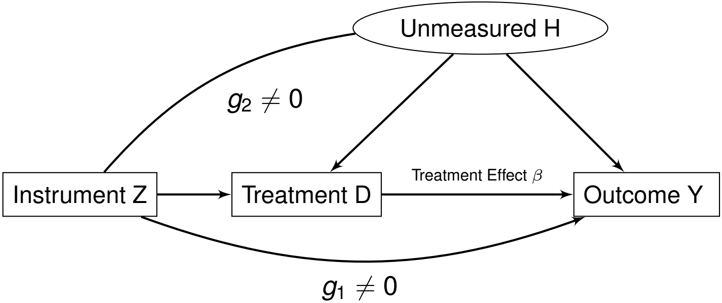
1. Weak IV: Staiger and Stock (1997); Stock, Wright, and Yogo (2002); Chao and Swanson (2005).
2. Orthogonality: Kolesár, Chetty, Friedman, Glaeser, and Imbens (2015); Bowden, Smith, and Burgess (2015).
3. Majority/plurality rule: Bowden, Smith, Haycock, and Burgess (2016); Kang, Zhang, Cai, and Small (2016); Guo, Kang, Cai, and Small (2018); Windmeijer, Farbmacher, Davies, and Smith (2019); Windmeijer, Liang, Hartwig, and Bowden (2021).
4. Heteroscedastic variance+homoscadastic correlation: Lewbel (2012); Tchetgen Tchetgen, Sun, and Walter (2021).

Robust Inference with all IVs being possibly invalid?

Outcome $Y_i \in \mathbb{R}$, treatment $D_i \in \mathbb{R}$, IV Z_i.

$$Y_i^{(z,d)} = Y_i^{(0,0)} + d\beta + g_1(z), \quad \mathbb{E}(Y_i^{(0,0)} \mid Z_i) = g_2(Z_i),$$

$$Y_i = D_i\beta + g(Z_i) + \epsilon_i \quad \text{with} \quad g = g_1 + g_2, \ \mathbb{E}(\epsilon_i \mid Z_i) = 0.$$



Invalid IV Model

$$Y_i = D_i\beta + g(Z_i, X_i) + \epsilon_i \quad \text{with} \quad \mathbb{E}(\epsilon_i \mid Z_i, X_i) = 0.$$

$$D_i = f(Z_i, X_i) + \delta_i \quad \text{with} \quad \mathbb{E}(\delta_i \mid Z_i, X_i) = 0.$$

$$h(Z_i, X_i) = g(Z_i, X_i) - \psi(X_i) \quad \text{with} \quad \psi(X_i) = \mathbb{E}[g(Z_i, X_i) \mid X_i].$$

1. Valid IV: $g(Z_i, X_i)$ does not depend on Z_i
   - Binary IV: $g(0, X_i) = g(1, X_i)$
   - $h(\cdot) = 0$ and $g(Z_i, X_i) = \psi(X_i)$
2. Invalid IV: $g(Z_i, X_i)$ depends on Z_i
   - Binary IV: $g(0, X_i) \neq g(1, X_i)$
   - $h(\cdot) \neq 0$
   - Polynomial violation: $g(Z_i, X_i) = Z_i + \cdots + Z_i^q + \psi(X_i)$
   - Interaction violation: $g(Z_i, X_i) = Z_i \cdot (X_i^{\intercal}\gamma) + \psi(X_i)$

Overview of talk

Two Stage Curvature Identification (TSCI)

Reduced form with covariates

$$Y_i = F(Z_i, X_i) + \epsilon_i + \beta\delta_i, \quad \text{with} \quad F(Z_i, X_i) = \beta f(Z_i, X_i) + g(Z_i, X_i)$$
$$D_i = f(Z_i, X_i) + \delta_i.$$

1. first stage: estimate $f(Z_i, X_i)$ by random forests.
2. second stage: assume $g(\cdot)$ is generated by $\mathcal{V}$ and adjust the violation form

$$\beta = \frac{\langle \mathcal{P}_{\mathcal{V}}^{\perp} F, \mathcal{P}_{\mathcal{V}}^{\perp} f \rangle}{\langle \mathcal{P}_{\mathcal{V}}^{\perp} f, \mathcal{P}_{\mathcal{V}}^{\perp} f \rangle} \quad \text{if} \quad f \notin \mathcal{V}$$

key assumption: $\|\mathcal{P}_{\mathcal{V}}^{\perp} f\|_2^2 > 0$.

First stage: splitting random forests (RF)

Randomly split the data into disjoint $\mathcal{A}_1$ and $\mathcal{A}_2$

1. Write $\mathcal{A}_1 = \{1, 2, \cdots, n_1\}$ and set $n_1 = \lfloor 2n/3 \rfloor$.
2. Construct the RF with $\{D_i, X_i, Z_i\}_{i \in \mathcal{A}_2}$
3. Estimate $f(z, x)$ by the RF and $\{X_j, Z_j, D_j\}_{j \in \mathcal{A}_1}$

$$\widehat{f}(z, x) = \sum_{j \in \mathcal{A}_1} \omega_j(z, x) D_j$$

- $\omega_j(z, x) \geq 0$ and $\sum_{j \in \mathcal{A}_1} \omega_j(z, x) = 1$
- Weights construction: $\{X_j, Z_j\}_{j \in \mathcal{A}_1}$ and RF.

Reference: Lin and Jeon (2006); Meinshausen (2006); Wager and Athey (2018).

First stage: a linear estimator

Estimate $f_{\mathcal{A}_1} = (f(Z_1, X_1), \cdots, f(Z_{n_1}, X_{n_1}))^\intercal$ by

$$\widehat{f}_{\mathcal{A}_1} = \Omega D_{\mathcal{A}_1} \quad \text{with} \quad \Omega_{ij} = \omega_j(Z_i, X_i) \quad \text{for} \quad i, j \in \mathcal{A}_1.$$

where $\widehat{f}_{\mathcal{A}_1} = (\widehat{f}(Z_1, X_1), \cdots, \widehat{f}(Z_{n_1}, X_{n_1}))^\intercal$.

- Ω is similar to hat matrix in linear regression
- Ω is NOT a projection matrix.

Second stage: generating $g(\cdot)$

$$g(Z_i, X_i) = h(Z_i, X_i) + \psi(X_i).$$

- Consider the violation space

$$\mathcal{V} = \mathrm{span}\{v_1(\cdot), \cdots, v_q(\cdot)\}$$

  and approximate the violation function $h(Z_i, X_i) = V_i^{\intercal}\pi$ with $V_i = (v_1(Z_i, X_i), \cdots, v_q(Z_i, X_i))^{\intercal}$
- $\psi(X_i) \approx W_i^{\intercal}\psi$ with W_i denoting the basis expansion

Write the outcome model,

$$Y_{\mathcal{A}_1} = D_{\mathcal{A}_1}\beta + \underbrace{V_{\mathcal{A}_1}\pi + W_{\mathcal{A}_1}\psi}_{g(\cdot)} + \epsilon_{\mathcal{A}_1}.$$

Second stage: known $\mathcal{V}$

Applying Ω to the outcome model,

$$\widehat{Y}_{\mathcal{A}_1} = \widehat{f}_{\mathcal{A}_1}\beta + \widehat{V}_{\mathcal{A}_1}\pi + \widehat{W}_{\mathcal{A}_1}\psi + \widehat{\epsilon}_{\mathcal{A}_1},$$

where $\widehat{Y}_{\mathcal{A}_1} = \Omega Y_{\mathcal{A}_1}, \widehat{f}_{\mathcal{A}_1} = \Omega D_{\mathcal{A}_1}, \widehat{V}_{\mathcal{A}_1} = \Omega V_{\mathcal{A}_1}, \widehat{W}_{\mathcal{A}_1} = \Omega W_{\mathcal{A}_1}$.

$$\widehat{\beta}_{\rm init}(V) = \frac{\widehat{Y}_{\mathcal{A}_1}^{\intercal} P^{\perp}_{\widehat{V}_{\mathcal{A}_1}, \widehat{W}_{\mathcal{A}_1}} \widehat{f}_{\mathcal{A}_1}}{\widehat{f}_{\mathcal{A}_1}^{\intercal} P^{\perp}_{\widehat{V}_{\mathcal{A}_1}, \widehat{W}_{\mathcal{A}_1}} \widehat{f}_{\mathcal{A}_1}} = \frac{Y_{\mathcal{A}_1}^{\intercal} \mathbf{M}_{\rm RF}(V) D_{\mathcal{A}_1}}{D_{\mathcal{A}_1}^{\intercal} \mathbf{M}_{\rm RF}(V) D_{\mathcal{A}_1}}$$

where

$$\mathbf{M}_{\rm RF}(V) = \Omega^{\intercal} P^{\perp}_{\widehat{V}_{\mathcal{A}_1}, \widehat{W}_{\mathcal{A}_1}} \Omega.$$

Bias correction

This estimator suffers from a finite-sample bias

$$\frac{\epsilon_{\mathcal{A}_1}^{\intercal} \mathbf{M}_{\mathrm{RF}}(V) \delta_{\mathcal{A}_1}}{D_{\mathcal{A}_1}^{\intercal} \mathbf{M}_{\mathrm{RF}}(V) D_{\mathcal{A}_1}} \approx \frac{\mathrm{Cov}(\delta_i, \epsilon_i) \cdot \mathrm{Tr}[\mathbf{M}_{\mathrm{RF}}(V)]}{D_{\mathcal{A}_1}^{\intercal} \mathbf{M}_{\mathrm{RF}}(V) D_{\mathcal{A}_1}}.$$

- $\mathrm{Tr}[\mathbf{M}_{\mathrm{RF}}(V)]$: might be large for random forests.
- $D_{\mathcal{A}_1}^{\intercal} \mathbf{M}_{\mathrm{RF}}(V) D_{\mathcal{A}_1}$: IV strength.

Bias correction

This estimator suffers from a finite-sample bias

$$\frac{\epsilon_{\mathcal{A}_1}^{\intercal}\mathbf{M}_{\rm RF}(V)\delta_{\mathcal{A}_1}}{D_{\mathcal{A}_1}^{\intercal}\mathbf{M}_{\rm RF}(V)D_{\mathcal{A}_1}} \approx \frac{{\rm Cov}(\delta_i,\epsilon_i)\cdot {\rm Tr}[\mathbf{M}_{\rm RF}(V)]}{D_{\mathcal{A}_1}^{\intercal}\mathbf{M}_{\rm RF}(V)D_{\mathcal{A}_1}}.$$

- ${\rm Tr}[\mathbf{M}_{\rm RF}(V)]$: might be large for random forests.
- $D_{\mathcal{A}_1}^{\intercal}\mathbf{M}_{\rm RF}(V)D_{\mathcal{A}_1}$: IV strength.

$$\widehat{\beta}_{\rm RF}(V) = \widehat{\beta}_{\rm init}(V) - \frac{\sum_{i=1}^{n_1}[\mathbf{M}_{\rm RF}(V)]_{ii}\widehat{\delta}_i[\widehat{\epsilon}(V)]_i}{D_{\mathcal{A}_1}^{\intercal}\mathbf{M}_{\rm RF}(V)D_{\mathcal{A}_1}}$$

where $\widehat{\delta}_{\mathcal{A}_1} = D_{\mathcal{A}_1} - \widehat{f}_{\mathcal{A}_1}$ and $\widehat{\epsilon}(V) = P_{V,W}^{\perp}[Y - D\widehat{\beta}_{\rm init}(V)]$.

First stage RF: Sample splitting+ Bias correction

Confidence interval

$$\mathrm{CI}_{\mathrm{RF}}(V) = \left(\widehat{\beta}_{\mathrm{RF}}(V) - z_{\alpha/2}\widehat{\mathrm{SE}}(V), \widehat{\beta}_{\mathrm{RF}}(V) + z_{\alpha/2}\widehat{\mathrm{SE}}(V)\right),$$

with

$$\widehat{\mathrm{SE}}(V) = \frac{\sqrt{\sum_{i=1}^{n_1} [\widehat{\epsilon}(V)]_i^2 [\mathbf{M}_{\mathrm{RF}}(V) D_{\mathcal{A}_1}]_i^2}}{D_{\mathcal{A}_1}^{\intercal} \mathbf{M}_{\mathrm{RF}}(V) D_{\mathcal{A}_1}}.$$

Heteroscadastic errors: $\mathrm{Cov}(\epsilon_i, \delta_i \mid Z_i, X_i)$, $\mathrm{Var}(\epsilon_i \mid Z_i, X_i)$, and $\mathrm{Var}(\delta_i \mid Z_i, X_i)$ are not necessarily constant.

Violation space selection

Generalized IV Strength

Generalize the concentration parameter as,

$$\mu(V) := \frac{f_{\mathcal{A}_1}^{\intercal} \mathbf{M}_{\mathrm{RF}}(V) f_{\mathcal{A}_1}}{\sum_{i \in \mathcal{A}_1} \mathrm{Var}(\delta_i \mid X_i, Z_i)/|\mathcal{A}_1|}.$$

- Test $\mu(V) \geq \max\{2\mathrm{Tr}\left[\mathbf{M}_{\mathrm{RF}}(V)\right], 10\}$
- Estimate $\mu(V)$ by $\widehat{\mu(V)} := \frac{D_{\mathcal{A}_1}^{\intercal} \mathbf{M}_{\mathrm{RF}}(V) D_{\mathcal{A}_1}}{\|D_{\mathcal{A}_1} - \widehat{f}_{\mathcal{A}_1}\|_2^2/n_1}$.

Conduct the generalized IV strength test

$$\widehat{\mu(V)} \geq \max\{2\mathrm{Tr}\left[\mathbf{M}_{\mathrm{RF}}(V)\right], 10\} + \mathcal{S}_{\alpha_0}(V),$$

where $\alpha_0 = 0.025$ and $\mathcal{S}_{\alpha_0}(V)$ controls $\widehat{\mu(V)} - \mu(V)$.

Violation space collection

Choose the best violation space among

$$\mathcal{V}_0 \subset \mathcal{V}_1 \subset \cdots \subset \mathcal{V}_Q.$$

- $\mathcal{V}_0 := \{h = 0\}$ as the null violation space
- $\mathcal{V}_q$ denotes the space spanned by a pre-specified set of basis functions $\{v_1(\cdot), v_2(\cdot), \cdots, v_q(\cdot)\}$ for $q \geq 1$.

Define $Q_{\max}$ as,

$$Q_{\max} = \arg\max_{q \geq 0} \left\{ \widehat{\mu(V_q)} \geq \max\{2\mathrm{Tr}\left[\mathbf{M}_{\mathrm{RF}}(V_q)\right], 10\} + \mathcal{S}_{\alpha_0}(V_q) \right\}.$$

Violation space selection

Choose the best violation space among

$$\mathcal{V}_0 \subset \mathcal{V}_1 \subset \cdots \subset \mathcal{V}_{Q_{\max}}.$$

Choose the smallest $0 \leq q \leq Q_{\max}$ such that

$$\widehat{\beta}_{\rm RF}(V_q) \text{ is not much different from } \{\widehat{\beta}_{\rm RF}(V_{q'})\}_{q+1 \leq q' \leq Q_{\max}}$$

Violation space selection

Choose the best violation space among

$$\mathcal{V}_0 \subset \mathcal{V}_1 \subset \cdots \subset \mathcal{V}_{Q_{\max}}.$$

Choose the smallest $0 \leq q \leq Q_{\max}$ such that

$$\widehat{\beta}_{\rm RF}(V_q) \text{ is not much different from } \{\widehat{\beta}_{\rm RF}(V_{q'})\}_{q+1 \leq q' \leq Q_{\max}}$$

Example for $Q_{\max} = 4$

	$\widehat{\beta}_{\rm RF}(V_1)$	$\widehat{\beta}_{\rm RF}(V_2)$	$\widehat{\beta}_{\rm RF}(V_3)$	$\widehat{\beta}_{\rm RF}(V_4)$
$\widehat{\beta}_{\rm RF}(V_0)$	✓	✓	✓	✓

Violation space selection

Choose the best violation space among

$$\mathcal{V}_0 \subset \mathcal{V}_1 \subset \cdots \subset \mathcal{V}_{Q_{\max}}.$$

Choose the smallest $0 \leq q \leq Q_{\max}$ such that

$$\widehat{\beta}_{\rm RF}(V_q) \text{ is not much different from } \{\widehat{\beta}_{\rm RF}(V_{q'})\}_{q+1 \leq q' \leq Q_{\max}}$$

Example for $Q_{\max} = 4$

	$\widehat{\beta}_{\rm RF}(V_1)$	$\widehat{\beta}_{\rm RF}(V_2)$	$\widehat{\beta}_{\rm RF}(V_3)$	$\widehat{\beta}_{\rm RF}(V_4)$
$\widehat{\beta}_{\rm RF}(V_0)$	✓	✓	✓	✓

	$\widehat{\beta}_{\rm RF}(V_1)$	$\widehat{\beta}_{\rm RF}(V_2)$	$\widehat{\beta}_{\rm RF}(V_3)$	$\widehat{\beta}_{\rm RF}(V_4)$
$\widehat{\beta}_{\rm RF}(V_0)$	X	X	X	✓
$\widehat{\beta}_{\rm RF}(V_1)$		✓	✓	✓

TSCI with Machine Learning

Write the first stage machine learning estimator as,

$$\widehat{f}_{\mathcal{A}_1} = \Omega D_{\mathcal{A}_1} \quad \text{for some matrix} \quad \Omega \in \mathbb{R}^{n_1 \times n_1}.$$

Define a generalized transformation matrix

$$\mathbf{M}(V) = \Omega^{\intercal} P^{\perp}_{\widehat{V}_{\mathcal{A}_1}, \widehat{W}_{\mathcal{A}_1}} \Omega, \quad \text{with} \quad \widehat{V}_{\mathcal{A}_1} = \Omega V_{\mathcal{A}_1}, \quad \widehat{W}_{\mathcal{A}_1} = \Omega W_{\mathcal{A}_1},$$

and the TSCI with machine learning

$$\widehat{\beta}(V) = \frac{Y^{\intercal}_{\mathcal{A}_1} \mathbf{M}(V) D_{\mathcal{A}_1}}{D^{\intercal}_{\mathcal{A}_1} \mathbf{M}(V) D_{\mathcal{A}_1}} - \frac{\sum_{i=1}^{n_1} [\mathbf{M}(V)]_{ii} \widehat{\delta}_i [\widehat{\epsilon}(V)]_i}{D^{\intercal}_{\mathcal{A}_1} \mathbf{M}(V) D_{\mathcal{A}_1}}.$$

- Ω for Basis approximation, Boosting, DNN.

Inference guarantee

$$\text{(R2-Inf)} \quad f_{\mathcal{A}_1}^{\intercal} [\mathbf{M}(V)]^2 f_{\mathcal{A}_1} \gg \max\{(\mathrm{Tr}[\mathbf{M}(V)])^2, 1\}.$$

Theorem 1 (G. and Bühlmann, 2022).

Under Condition (R2-Inf) *and other regularity conditions, we have*

$$\frac{1}{\mathrm{SE}(V)} \left(\widehat{\beta}(V) - \beta\right) \xrightarrow{d} N(0, 1).$$

In the paper, we also establish

- Improvement of bias correction.
- Consistency of violation space selection.

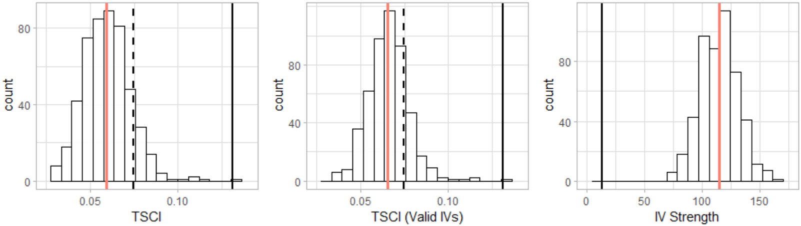
Overview of talk

Effect of education on income

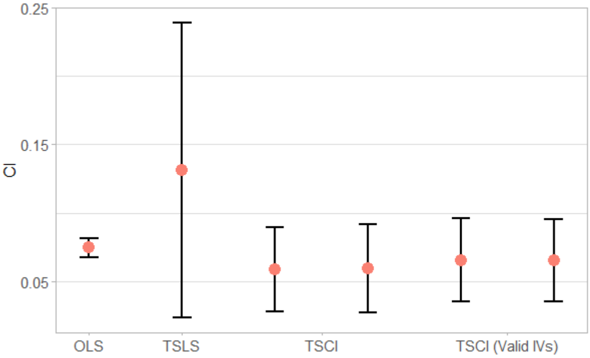
Analyze the same data set as Card (1995).

- $n = 3010$: National Longitudinal Survey of Young Men
- log wages (outcome), years of schooling (treatment).
- IV: an indicator for a nearby 4-year college in 1966.
- Baseline covariates: a quadratic function of potential experience, a race indicator, and dummy variables for residence in a metropolitan area and the south in 1976.

We test against the violation form generated by $\mathcal{V}_1 = \{\texttt{nearc4}, \texttt{nearc4} \cdot \texttt{exper}, \texttt{nearc4} \cdot \texttt{exper}^2, \texttt{nearc4} \cdot \texttt{race}, \texttt{nearc4} \cdot \texttt{sama}, \texttt{nearc4} \cdot \texttt{south}\}$.

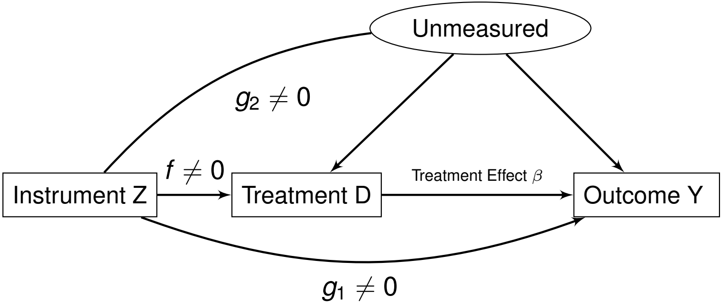


- Report 500 TSCI estimates due to 501 different splitting.
- Correction of the positive “ability bias" of OLS.



- Proximity to the college is not valid: out of the 500 splits, 302 splits report an invalid IV.
- Multi-splitting CI is $(0.0282, 0.0898)$

Conclusion and Discussion



- ML+ self-checking weak and invalid IVs.
- Correction of the overfitting bias.
- Non-constant effect + post-selection inference.

Reference and Acknowledgement

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`

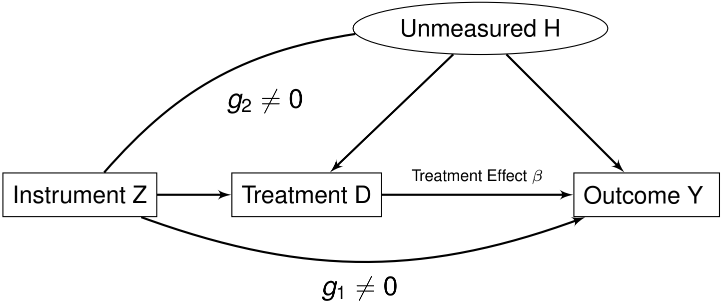
Acknowledgement to NIH and NSF for support.

Thank You!

SEM with hidden confounders

$$Y_i \leftarrow D_i\beta + g_1(Z_i) + \nu_1(H_i) + \epsilon_i^0, \quad D_i \leftarrow f_1(Z_i) + \nu_2(H_i) + \delta_i^0.$$

Define $g_2(Z_i) = \mathbb{E}\left(\nu_1(H_i) \mid Z_i\right)$ and $f_2(Z_i) = \mathbb{E}\left(\nu_2(H_i) \mid Z_i\right)$.



Adjustment of finite-sample randomness by splitting

Consider S random sample splitting and denote the TSCI point and standard error estimators as

$$\{\widehat{\beta}^{s}\}_{1\leq s\leq S} \quad \text{and} \quad \{\widehat{\mathrm{SE}}^{s}\}_{1\leq s\leq S}$$

For any $\beta_0 \in \mathbb{R}$, we construct S p values as

$$p^{s}(\beta_0) = 2(1-\psi(|\widehat{\beta}^{s}-\beta_0|/\widehat{\mathrm{SE}}^{s})), \quad \text{for} \quad 1 \leq s \leq S.$$

We define the multi-splitting confidence interval as

$$\{\beta_0 \in \mathbb{R} : 2 \cdot \mathrm{median}\{p^{s}(\beta_0)\}_{s=1}^{S} \leq \alpha\}.$$

Violation space selection

Choose the smallest $0 \leq q \leq Q_{\max}$ such that

$\widehat{\beta}_{\rm RF}(V_q)$ is not much different from $\{\widehat{\beta}_{\rm RF}(V_{q'})\}_{q+1 \leq q' \leq Q_{\max}}$

For $0 \leq q \leq Q_{\max} - 1$, define

$$\mathcal{C}^{\rm RF}(V_q) = \begin{cases} 0 & \text{if } \max_{q+1 \leq q' \leq Q_{\max}} \frac{\left|\widehat{\beta}_{\rm RF}(V_q) - \widehat{\beta}_{\rm RF}(V_{q'})\right|}{\sqrt{\widehat{H}(V_q, V_{q'})}} \leq \widehat{\rho} \\ 1 & \text{otherwise} \end{cases},$$

where $\widehat{\rho} > 0$ is a positive threshold.

$$\widehat{q}_c = \arg\min_{0 \leq q \leq Q_{\max}} \left\{\mathcal{C}^{\rm RF}(V_q) = 0\right\}.$$

IV Strength Assumption

$$\text{(R2)} f_{\mathcal{A}_1}^{\intercal} \mathbf{M}(V) f_{\mathcal{A}_1} \gg \max\{\mathrm{Tr}[\mathbf{M}(V)], \|R(V)\|_2^2, 1\}.$$

- $\mu(V) = f_{\mathcal{A}_1}^{\intercal} \mathbf{M}(V) f_{\mathcal{A}_1} / \sigma_{\delta}^2$
- Test $\mu(V) \geq \max\{2\mathrm{Tr}[\mathbf{M}(V)], 10\}$.
- Small approximation error $R(V) = h - V\pi$.
- Strong IV: $f_{\mathcal{A}_1}^{\intercal} \mathbf{M}(V) f_{\mathcal{A}_1} \asymp n$.
- Enough association: not requiring $\widehat{f}_{\mathcal{A}_1}$ to be consistent.

$$\text{(R2-Inf)} \quad f_{\mathcal{A}_1}^{\intercal} [\mathbf{M}(V)]^2 f_{\mathcal{A}_1} \gg \max\{(\mathrm{Tr}[\mathbf{M}(V)])^2, \|R(V)\|_2^2, 1\}.$$

- RF: $f_{\mathcal{A}_1}^{\intercal} [\mathbf{M}(V)]^2 f_{\mathcal{A}_1} \leq f_{\mathcal{A}_1}^{\intercal} \mathbf{M}(V) f_{\mathcal{A}_1}$
- Basis/DNN: $f_{\mathcal{A}_1}^{\intercal} [\mathbf{M}(V)]^2 f_{\mathcal{A}_1} = f_{\mathcal{A}_1}^{\intercal} \mathbf{M}(V) f_{\mathcal{A}_1}$

Simulation setting

$$Y_i = D_i + h(Z_i) + 0.2 \cdot \sum_{j=1}^{20} X_{ij} + \epsilon_i, \quad D_i = f(Z_i, X_i) + \delta_i, \quad \text{for } 1 \leq i \leq n.$$

$$f(Z_i, X_i) = -\frac{25}{12} + Z_i + Z_i^2 + \frac{1}{8} Z_i^4 + Z_i \cdot (a \cdot \sum_{j=1}^{5} X_{ij}) - \sum_{j=1}^{p} 0.3 X_{ij}.$$

Generate Z_i as uniform $(-2, 2)$.

- Linear violation (vio= 1): $h(Z_i) = Z_i$;
- Quadratic violation (vio= 2): $h(Z_i) = Z_i + Z_i^2 - 1$.

Empirical Coverage and Validity Test

			TSCI-RF				TSLS	Other RF(oracle)		
vio	a	n	Oracle	Comp	Robust	Invalidity		Init	Plug	Full
1	0.0	1000	0.93	0.93	0.93	1.00	0	0.93	0.02	0.79
		3000	0.95	0.95	0.95	1.00	0	0.94	0.00	0.74
		5000	0.96	0.95	0.95	1.00	0	0.96	0.00	0.66
	0.5	1000	0.96	0.96	0.96	1.00	0	0.96	0.03	0.80
		3000	0.95	0.93	0.92	1.00	0	0.95	0.00	0.68
		5000	0.95	0.93	0.93	1.00	0	0.95	0.00	0.60
	1.0	1000	0.94	0.93	0.90	1.00	0	0.93	0.02	0.79
		3000	0.94	0.94	0.93	1.00	0	0.94	0.00	0.60
		5000	0.94	0.93	0.93	1.00	0	0.94	0.00	0.47
2	0.0	1000	0.48	0.00	0.00	1.00	0	0.24	0.00	0.00
		3000	0.78	0.00	0.00	1.00	0	0.58	0.00	0.00
		5000	0.85	0.02	0.02	1.00	0	0.66	0.00	0.00
	0.5	1000	0.84	0.00	0.00	0.95	0	0.64	0.07	0.01
		3000	0.87	0.87	0.87	1.00	0	0.77	0.63	0.00
		5000	0.93	0.93	0.93	1.00	0	0.83	0.39	0.00
	1.0	1000	0.91	0.90	0.90	1.00	0	0.89	0.34	0.21
		3000	0.93	0.93	0.92	1.00	0	0.90	0.00	0.02
		5000	0.93	0.93	0.93	1.00	0	0.92	0.00	0.01