Identification-robust inference for the LATE with high-dimensional covariates

Yukun Ma
Vanderbilt University

January 17, 2024

Motivation: Weak IV in Empirical Practice

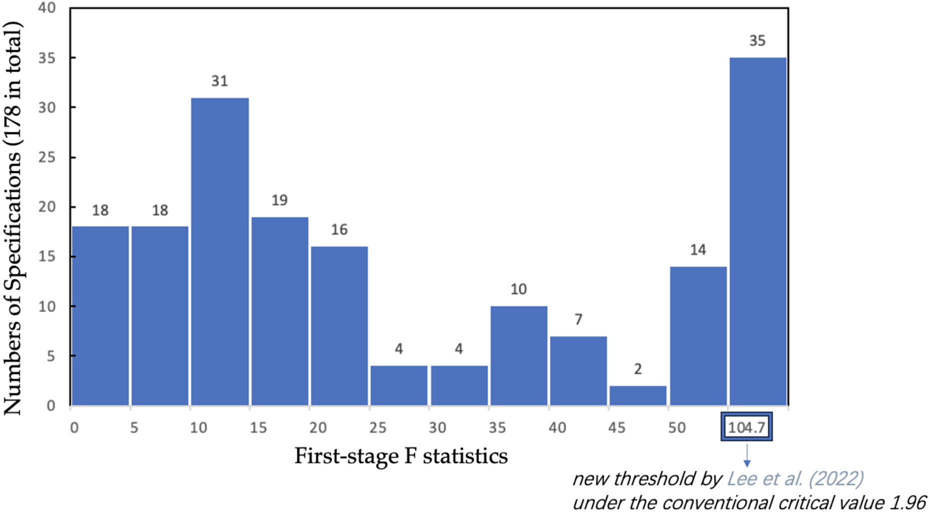


Figure: *American Economic Review* 2018-2022 ▸ heterscadesticity

Motivation: Availability of Large Datasets

- Big datasets are becoming increasingly available nowadays,
  - high data volume, 15.5 ZB in 2015, 97 ZB in 2022 (representing a 525% increase); (1 ZB = $\mathbf{10^{12}}$ GB)
  - high-dimensional controls, allowing for more flexible functional forms, e.g., polynomial terms and interaction effects.

Abstract

- New inference procedure for local average treatment effect (LATE) when
  - identification may be weak (e.g. few compliers),
  - model incorporate high-dimensional covariates (e.g. many controls).
- The proposed test statistic has uniformly correct asymptotic size,
  - inversion of the proposed test statistic for inference on LATE.
- Revisit 2 empirical studies:
  - Hornung (2015, JEEA) and Ambrus et al. (2020, AER),

  $\hookrightarrow$ in both cases, the proposed method is more efficient, yielding narrower confidence regions -whereas competitors often report larger confidence intervals.

Table of Contents

Table of Contents

LATE

- LATE: The effect of a treatment on compliers who adhere to the treatment assigned to their sample group.
- Assume we have N observations
  - Y_i : outcome of interest for unit i.
  - $D_i \in \{0, 1\}$: receipt of treatment.
  - $Z_i \in \{0, 1\}$: offer of treatment.
- Imbens and Angrist (1994) propose

$$\text{LATE} = \frac{\mathbf{E}_P[Y|Z=1] - \mathbf{E}_P[Y|Z=0]}{\mathbf{E}_P[D|Z=1] - \mathbf{E}_P[D|Z=0]} = \frac{ITT}{FS}.$$

- Incorporate covariates into LATE estimation, e.g. Abadie (2003)
  - X_i : p-dimensional covariates.
- Weak identification in LATE: $FS \to 0$.

LATE

- LATE: The effect of a treatment on compliers who adhere to the treatment assigned to their sample group.
- Assume we have N observations
  - Y_i : outcome of interest for unit i.
  - $D_i \in \{0, 1\}$: receipt of treatment.
  - $Z_i \in \{0, 1\}$: offer of treatment.
- Imbens and Angrist (1994) propose

$$\theta := \text{LATE} = \frac{\mathbf{E}_P[Y|Z=1] - \mathbf{E}_P[Y|Z=0]}{\mathbf{E}_P[D|Z=1] - \mathbf{E}_P[D|Z=0]} = \frac{ITT}{FS} := \frac{\delta}{\pi}.$$

- Incorporate covariates into LATE estimation, e.g. Abadie (2003)
  - X_i : p-dimensional covariates.
- Weak identification in LATE: $\pi \to 0$

Weak identification

- Issue: When instruments $\boldsymbol{Z}$ are weakly correlated with endogenous regressors $\boldsymbol{D}$, conventional methods for IV estimation and inference become unreliable.

$$\boldsymbol{\theta} = \frac{\boldsymbol{\delta}}{\boldsymbol{\pi}},$$

when $\widehat{\boldsymbol{\pi}}$ is close to zero, $\widehat{\boldsymbol{\theta}}$ is highly nonlinear in $\widehat{\boldsymbol{\pi}}$

- Trick: Fieller-type transformation & test inversion.
Given $\boldsymbol{H_0} : \boldsymbol{\theta} = \boldsymbol{\theta_0}$, we have $\boldsymbol{\delta} - \boldsymbol{\theta_0\pi} = \mathbf{0}$. The Anderson-Rubin (AR) test,

$$\boldsymbol{AR(\theta) = (\delta - \theta\pi)'\Omega(\theta)^{-1}(\delta - \theta\pi)},$$

follows a $\boldsymbol{\chi^2}$ distribution under $\boldsymbol{H_0}$.

Weak identification

- Issue: When instruments $\boldsymbol{Z}$ are weakly correlated with endogenous regressors $\boldsymbol{D}$, conventional methods for IV estimation and inference become unreliable.

$$\boldsymbol{\theta} = \frac{\boldsymbol{\delta}}{\boldsymbol{\pi}},$$

when $\widehat{\boldsymbol{\pi}}$ is close to zero, $\widehat{\boldsymbol{\theta}}$ is highly nonlinear in $\widehat{\boldsymbol{\pi}}$

- Trick: Fieller-type transformation & test inversion.
Given $\boldsymbol{H_0 : \theta = \theta_0}$, we have $\boldsymbol{\delta - \theta_0 \pi = 0}$. The Anderson-Rubin (AR) test,

$$\boldsymbol{AR(\theta) = (\delta - \theta\pi)'\Omega(\theta)^{-1}(\delta - \theta\pi)},$$

follows a $\boldsymbol{\chi^2}$ distribution under $\boldsymbol{H_0}$.

Relations to the Literature: Weak Identification

- Inference procedures depend on the observed process only through its value, and potentially derivative, at the point $\boldsymbol{\theta_0}$: Anderson–Rubin statistic, Stock and Wright (2000), Kleibergen (2005).
- Methods depend on the full path of the observed process: Moreira (2003) and Andrews and Mikusheva (2016).

	Low-dimensional Model
Strong Identification	z-test
Weak Identification	Stock and Wright (2000), Kleibergen (2005), Andrews and Mikusheva (2016)

Relations to the Literature: Weak Identification

- Inference procedures depend on the observed process only through its value, and potentially derivative, at the point $\boldsymbol{\theta_0}$: Anderson–Rubin statistic, Stock and Wright (2000), Kleibergen (2005).
- Methods depend on the full path of the observed process: Moreira (2003) and Andrews and Mikusheva (2016).

	Low-dimensional Model
Strong Identification	z-test
Weak Identification	Stock and Wright (2000), Kleibergen (2005), Andrews and Mikusheva (2016)

$\hookrightarrow$Limitation: None of them considers high-dimensional model (model with many covariates).

Relations to the Literature: ML Methods

In high-dimensional models, ML methods are commonly employed for model selection:

- Chernozhukov et al. (2018) introduce the double/debiased machine learning (DML) method, a combination of the Neyman orthogonality condition and cross-fitting method.
- Belloni et al. (2017) present an efficient estimator and confidence bands for the LATE with nonparametric/high-dimensional components.

	Low-dimensional Model	High-dimensional Model
Strong Identification	z-test	ML methods
Weak Identification	Stock and Wright (2000), Kleibergen (2005), Andrews and Mikusheva(2016)	

Relations to the Literature: ML Methods

In high-dimensional models, ML methods are commonly employed for model selection:

- Chernozhukov et al. (2018) introduce the double/debiased machine learning (DML) method, a combination of the Neyman orthogonality condition and cross-fitting method.
- Belloni et al. (2017) present an efficient estimator and confidence bands for the LATE with nonparametric/high-dimensional components.

	Low-dimensional Model	High-dimensional Model
Strong Identification	z-test	ML methods
Weak Identification	Stock and Wright (2000), Kleibergen (2005), Andrews and Mikusheva(2016)	

Comparison of the Literature

Weak Identification lit

- identification-robust test statistics
- use traditional methods to handle $\boldsymbol{X_i}$

Drawbacks

- overfitting
- multicollinearity

Machine Learning lit

- normal z-test
- use ML to handle $\boldsymbol{X_i}$

- cannot perform well under weakly identified scenarios

$\hookrightarrow$ My proposed method takes advantage from both literature:

- identification-robust test statistics
- use ML to handle the high-dimensional $\boldsymbol{X_i}$

Comparison of the Literature

Weak Identification lit

- identification-robust test statistics
- use traditional methods to handle $\boldsymbol{X_i}$

Machine Learning lit

- normal z-test
- use ML to handle $\boldsymbol{X_i}$

Drawbacks

Weak Identification lit:

- overfitting
- multicollinearity

Machine Learning lit:

- cannot perform well under weakly identified scenarios

$\hookrightarrow$ My proposed method takes advantage from both literature:

- identification-robust test statistics
- use ML to handle the high-dimensional $\boldsymbol{X_i}$

Contributions

	Low-dimensional Model	High-dimensional Model
Strong Identification	z-test	ML methods
Weak Identification	Stock and Wright (2000), Kleibergen (2005), Andrews and Mikusheva(2016)	My proposed method

Technical Contribution: Instead of proposing a consistent estimator for $\boldsymbol{\theta}$, I present an empirical process along with its uniformly consistent estimator,

$\hookrightarrow$ the proposed test statistic is shown to be uniformly size-correct.

Table of Contents

Setup

- Model the random vector $\boldsymbol{W} = (\boldsymbol{Y}, \boldsymbol{D}, \boldsymbol{Z}, \boldsymbol{X}')'$ as follows,

$$
\begin{array}{lll}
\textit{First stage} & \boldsymbol{D} = \boldsymbol{m_0}(\boldsymbol{Z}, \boldsymbol{X}) + \boldsymbol{v}, & \mathbf{E}_{\boldsymbol{P}}[\boldsymbol{v}|\boldsymbol{Z}, \boldsymbol{X}] = \mathbf{0} \\
\textit{Reduced form} & \boldsymbol{Y} = \boldsymbol{g_0}(\boldsymbol{Z}, \boldsymbol{X}) + \boldsymbol{u}, & \mathbf{E}_{\boldsymbol{P}}[\boldsymbol{u}|\boldsymbol{Z}, \boldsymbol{X}] = \mathbf{0} \\
\textit{Propensity score} & \boldsymbol{Z} = \boldsymbol{p_0}(\boldsymbol{X}) + \boldsymbol{e}, & \mathbf{E}_{\boldsymbol{P}}[\boldsymbol{e}|\boldsymbol{X}] = \mathbf{0}
\end{array}
$$

- $\boldsymbol{m_0}$, $\boldsymbol{g_0}$, $\boldsymbol{p_0}$: no need to impose any parametric assumptions by Blandhol et al. (2022).

- The LATE framework proposed by Tan (2006) is given by

$$
\boldsymbol{\theta} = \frac{\mathbf{E}_{\boldsymbol{P}}[\boldsymbol{g}(\mathbf{1},\boldsymbol{X}) - \boldsymbol{g}(\mathbf{0},\boldsymbol{X}) + \frac{\boldsymbol{Z}}{\boldsymbol{p}(\boldsymbol{X})}(\boldsymbol{Y} - \boldsymbol{g}(\mathbf{1},\boldsymbol{X})) - \frac{\mathbf{1}-\boldsymbol{Z}}{\mathbf{1}-\boldsymbol{p}(\boldsymbol{X})}(\boldsymbol{Y} - \boldsymbol{g}(\mathbf{0},\boldsymbol{X}))]}{\mathbf{E}_{\boldsymbol{P}}[\boldsymbol{m}(\mathbf{1},\boldsymbol{X}) - \boldsymbol{m}(\mathbf{0},\boldsymbol{X}) + \frac{\boldsymbol{Z}}{\boldsymbol{p}(\boldsymbol{X})}(\boldsymbol{D} - \boldsymbol{m}(\mathbf{1},\boldsymbol{X})) - \frac{\mathbf{1}-\boldsymbol{Z}}{\mathbf{1}-\boldsymbol{p}(\boldsymbol{X})}(\boldsymbol{D} - \boldsymbol{m}(\mathbf{0},\boldsymbol{X}))]}
$$

Setup

- Model the random vector $\boldsymbol{W} = (\boldsymbol{Y}, \boldsymbol{D}, \boldsymbol{Z}, \boldsymbol{X}')'$ as follows,

$$\begin{array}{llll} \textit{First stage} & \boldsymbol{D} = \boldsymbol{m_0}(\boldsymbol{Z}, \boldsymbol{X}) + \boldsymbol{v}, & \mathbf{E}_{\boldsymbol{P}}[\boldsymbol{v}|\boldsymbol{Z}, \boldsymbol{X}] = \mathbf{0} \\ \textit{Reduced form} & \boldsymbol{Y} = \boldsymbol{g_0}(\boldsymbol{Z}, \boldsymbol{X}) + \boldsymbol{u}, & \mathbf{E}_{\boldsymbol{P}}[\boldsymbol{u}|\boldsymbol{Z}, \boldsymbol{X}] = \mathbf{0} \\ \textit{Propensity score} & \boldsymbol{Z} = \boldsymbol{p_0}(\boldsymbol{X}) + \boldsymbol{e}, & \mathbf{E}_{\boldsymbol{P}}[\boldsymbol{e}|\boldsymbol{X}] = \mathbf{0} \end{array}$$

- The LATE framework proposed by Tan (2006) is given by

$$\boldsymbol{\theta} = \frac{\mathbf{E}_{\boldsymbol{P}}[\boldsymbol{g}(\mathbf{1},\boldsymbol{X}) - \boldsymbol{g}(\mathbf{0},\boldsymbol{X}) + \frac{\boldsymbol{Z}}{\boldsymbol{p}(\boldsymbol{X})}(\boldsymbol{Y} - \boldsymbol{g}(\mathbf{1},\boldsymbol{X})) - \frac{\mathbf{1}-\boldsymbol{Z}}{\mathbf{1}-\boldsymbol{p}(\boldsymbol{X})}(\boldsymbol{Y} - \boldsymbol{g}(\mathbf{0},\boldsymbol{X}))]}{\mathbf{E}_{\boldsymbol{P}}[\boldsymbol{m}(\mathbf{1},\boldsymbol{X}) - \boldsymbol{m}(\mathbf{0},\boldsymbol{X}) + \frac{\boldsymbol{Z}}{\boldsymbol{p}(\boldsymbol{X})}(\boldsymbol{D} - \boldsymbol{m}(\mathbf{1},\boldsymbol{X})) - \frac{\mathbf{1}-\boldsymbol{Z}}{\mathbf{1}-\boldsymbol{p}(\boldsymbol{X})}(\boldsymbol{D} - \boldsymbol{m}(\mathbf{0},\boldsymbol{X}))]} = \frac{\mathbf{E}_{\boldsymbol{P}}[\boldsymbol{a}]}{\mathbf{E}_{\boldsymbol{P}}[\boldsymbol{b}]}$$

  - ▷ $\boldsymbol{a} := \boldsymbol{g}(\mathbf{1}, \boldsymbol{X}) - \boldsymbol{g}(\mathbf{0}, \boldsymbol{X}) + \dfrac{\boldsymbol{Z}}{\boldsymbol{p}(\boldsymbol{X})}(\boldsymbol{Y} - \boldsymbol{g}(\mathbf{1}, \boldsymbol{X})) - \dfrac{\mathbf{1} - \boldsymbol{Z}}{\mathbf{1} - \boldsymbol{p}(\boldsymbol{X})}(\boldsymbol{Y} - \boldsymbol{g}(\mathbf{0}, \boldsymbol{X}))$
  - ▷ $\boldsymbol{b} := \boldsymbol{m}(\mathbf{1}, \boldsymbol{X}) - \boldsymbol{m}(\mathbf{0}, \boldsymbol{X}) + \dfrac{\boldsymbol{Z}}{\boldsymbol{p}(\boldsymbol{X})}(\boldsymbol{D} - \boldsymbol{m}(\mathbf{1}, \boldsymbol{X})) - \dfrac{\mathbf{1} - \boldsymbol{Z}}{\mathbf{1} - \boldsymbol{p}(\boldsymbol{X})}(\boldsymbol{D} - \boldsymbol{m}(\mathbf{0}, \boldsymbol{X}))$

Score Function

- Consider a function

$$\psi(W;\theta,g,m,p) = \overbrace{g(1,X) - g(0,X) + \frac{Z(Y-g(1,X))}{p(X)} - \frac{(1-Z)(Y-g(0,X))}{1-p(X)}}^{a}$$
$$- \theta \times \underbrace{\left(m(1,X) - m(0,X) + \frac{Z(D-m(1,X))}{p(X)} - \frac{(1-Z)(D-m(0,X))}{1-p(X)}\right)}_{b},$$

with

- target parameter $\theta \in \Theta \subset \mathbb{R}$ is the LATE.
- nuisance parameter $\eta = (g, m, p) \in T$ for a convex[1] set T.
- ψ is a score function.

- Two-stage procedure:
  Stage 1 estimating nuisance parameter η,
  Stage 2 making inference for the target parameter θ.

[1] To ensure that $\psi(W;\theta_0,\eta_0 + r(\eta-\eta_0))$ is well defined for all $r \in [0,1]$.

Score Function

- Consider a function

$$\psi(W;\theta,g,m,p) = \overbrace{g(1,X) - g(0,X) + \frac{Z(Y-g(1,X))}{p(X)} - \frac{(1-Z)(Y-g(0,X))}{1-p(X)}}^{a}$$
$$- \theta \times \underbrace{\left(m(1,X) - m(0,X) + \frac{Z(D-m(1,X))}{p(X)} - \frac{(1-Z)(D-m(0,X))}{1-p(X)}\right)}_{b},$$

  with
  - target parameter $\theta \in \Theta \subset \mathbb{R}$ is the LATE.
  - nuisance parameter $\eta = (g,m,p) \in T$ for a convex[1] set T.
  - ψ is a score function.

- Two-stage procedure:
  Stage 1 estimating nuisance parameter η,
  Stage 2 making inference for the target parameter θ.

[1] To ensure that $\psi(W;\theta_0,\eta_0 + r(\eta-\eta_0))$ is well defined for all $r \in [0,1]$.

Nuisance Parameters

- Specify the nuisance parameters $\boldsymbol{\eta} = (\boldsymbol{g}, \boldsymbol{m}, \boldsymbol{p})$ as follows,

$$
\begin{aligned}
&\boldsymbol{g(Z,X) = \mathbf{E}_P[Y|Z,X] = Z\beta_{21} + X'\beta_{22}} && \textit{Reduced form}\\
&\boldsymbol{m(Z,X) = \mathbf{E}_P[D|Z,X] = \Lambda(Z\beta_{11} + X'\beta_{12})} && \textit{First stage}\\
&\boldsymbol{p(X) = \mathbf{E}_P[Z|X] = \Lambda(X'\gamma)} && \textit{Propensity score}
\end{aligned}
$$

  - The logistic CDF $\boldsymbol{\Lambda(t) = \frac{\exp(t)}{1+\exp(t)}}$ for all $\boldsymbol{t} \in \mathbb{R}$

  - In this example, the nuisance parameters $\boldsymbol{\eta = (\beta_{11}, \beta_{12}, \beta_{21}, \beta_{22}, \gamma)}$.

Properties of the Score ψ

- Moment condition:

$$\mathbf{E}_P[\underbrace{\psi(W_i;\theta_0,\eta_0)}_{a-\theta_0\times b}] = 0.$$

- Neyman orthogonality condition:
  - Path-wise (or Gateaux) derivative map D_r

$$D_r[\eta-\eta_0] := \partial_r\{\mathbf{E}_P[\psi(W;\theta_0,\eta_0+r(\eta-\eta_0))]\} \text{ for } \eta \in T.$$

  - The Neyman orthogonality condition holds at (θ_0,η_0) if

$$D_0[\eta-\eta_0] = \partial_\eta \mathbf{E}_P \psi(W;\theta_0,\eta_0)[\eta-\eta_0] = 0$$

holds for all $\eta \in \mathcal{T}_N$ for a nuisance realization set $\mathcal{T}_N \subset T$.

$\hookrightarrow$ The score function ψ is an AR-type Neyman orthogonal score.

Algorithm Breakdown

- estimate nuisance parameter $\boldsymbol{\eta}$

  Step 1-2: data splitting/ cross-validation

- make an inference for the target parameter $\boldsymbol{\theta}$ based on $\widehat{\boldsymbol{\eta}}$

  Step 3-6: inversion of the condition QLR test statistics

Estimate Nuisance Parameter η

Step 1: Randomly split the sample $\{1, \cdots, N\}$ into K folds $\{I_1, \cdots, I_K\}$.

Step 2: For each $k \in \{1, \cdots, K\}$, obtain $\widehat{\eta}_k$ by ML methods using only the subsample of those observations with indices $i \in \{1, \cdots, N\} \setminus I_k$:

An illustration of K=2-fold cross-fitting.

Estimate Nuisance Parameter, Step 2

(2.1) $(\widehat{\boldsymbol{\beta}}_{21,k}, \widehat{\boldsymbol{\beta}}_{22,k})$ in reduce form: run ML (e.g., lasso) OLS regression to estimate,

$$(\widehat{\boldsymbol{\beta}}_{21,k}, \widehat{\boldsymbol{\beta}}_{22,k}) \in \arg \min_{\boldsymbol{\beta}_{21},\boldsymbol{\beta}_{22}} \mathbb{E}_{I_k^c}[(\boldsymbol{Y}_i - \boldsymbol{Z}_i\boldsymbol{\beta}_{21} - \boldsymbol{X}_i'\boldsymbol{\beta}_{22})^2] + \frac{\boldsymbol{\lambda}_3}{|\boldsymbol{I}_k^c|}\|(\boldsymbol{\beta}_{21}, \boldsymbol{\beta}_{22})\|_1,$$

(2.2) $(\widehat{\boldsymbol{\beta}}_{11,k}, \widehat{\boldsymbol{\beta}}_{12,k})$ in first stage: run ML (e.g., lasso) logistic regression to estimate,

$$(\widehat{\boldsymbol{\beta}}_{11,k}, \widehat{\boldsymbol{\beta}}_{12,k}) \in \arg \min_{\boldsymbol{\beta}_{11},\boldsymbol{\beta}_{12}} \left\{\mathbb{E}_{I_k^c}[\boldsymbol{L}_1(\boldsymbol{W}_i; \boldsymbol{\beta}_{11}, \boldsymbol{\beta}_{12})] + \frac{\boldsymbol{\lambda}_1}{|\boldsymbol{I}_k^c|}\|(\boldsymbol{\beta}_{11}, \boldsymbol{\beta}_{12})\|_1\right\},$$

(2.3) $\widehat{\boldsymbol{\gamma}}_k$ in the propensity score: run ML (e.g., lasso) logistic regression to estimate,

$$\widehat{\boldsymbol{\gamma}}_k \in \arg \min_{\boldsymbol{\gamma}} \left\{\mathbb{E}_{I_k^c}[\boldsymbol{L}_2(\boldsymbol{W}_i; \boldsymbol{\gamma})] + \frac{\boldsymbol{\lambda}_2}{|\boldsymbol{I}_k^c|}\|\boldsymbol{\gamma}\|_1\right\},$$

- $\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \boldsymbol{\lambda}_3$: penalty parameter
- $\boldsymbol{L}_1(\boldsymbol{W}_i; \boldsymbol{\beta}_{11}, \boldsymbol{\beta}_{12}) = \boldsymbol{D}_i(\boldsymbol{Z}_i\boldsymbol{\beta}_{11} + \boldsymbol{X}_i'\boldsymbol{\beta}_{12}) - \log(1 + \exp(\boldsymbol{Z}_i\boldsymbol{\beta}_{11} + \boldsymbol{X}_i'\boldsymbol{\beta}_{12})),$
- $\boldsymbol{L}_2(\boldsymbol{W}_i; \boldsymbol{\gamma}) = \boldsymbol{Z}_i\boldsymbol{X}_i'\boldsymbol{\gamma} - \log(1 + \exp(\boldsymbol{X}_i'\boldsymbol{\gamma})).$

Step 3

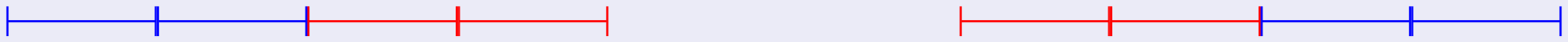
Step 3: Compute $\widehat{q}_N(\theta)$ and $\widehat{\Omega}(\theta_1, \theta_2)$ for later use,

$$\widehat{q}_N(\theta) = \frac{1}{\sqrt{N}} \sum_{k=1}^{K} \sum_{i \in I_k} \psi(W_i; \theta, \widehat{\eta}_k),$$

$$\begin{aligned}\widehat{\Omega}(\theta_1, \theta_2) &= \frac{1}{N} \sum_{k=1}^{K} \sum_{i \in I_k} \psi(W_i; \theta_1, \widehat{\eta}_k)\psi(W_i; \theta_2, \widehat{\eta}_k) \\ &\quad - \frac{1}{N^2} \sum_{k=1}^{K} \sum_{k'=1}^{K} \sum_{i \in I_k, i' \in I_{k'}} \psi(W_i; \theta_1, \widehat{\eta}_k)\psi(W_{i'}; \theta_2, \widehat{\eta}_{k'}).\end{aligned}$$

An illustration of K=2-fold cross-fitting.

I_1 *Score* I_2 *Nuisance* | I_1 *Nuisance* I_2 *Score*



$\sum_{i \in I_1} \psi(W_i; \theta, \widehat{\eta}_1)$ | $\sum_{i \in I_2} \psi(W_i; \theta, \widehat{\eta}_2)$

Make Inference for Target Parameter θ

Step 4: Take independent draws $\xi \sim N(0, \widehat{\Omega}(\theta_0, \theta_0))$ and calculate $R = R(\xi, h_N, \widehat{\Omega})$, where null hypothesis H_0

$$R(\xi, h_N, \widehat{\Omega}) = \xi^2 \widehat{\Omega}(\theta_0, \theta_0)^{-1} - \inf_{\theta} (V(\theta)\xi + h_N)^2 \widehat{\Omega}(\theta, \theta)^{-1},$$

- $V(\theta) = \widehat{\Omega}(\theta, \theta_0)\widehat{\Omega}(\theta_0, \theta_0)^{-1}$,
- $h_N(\theta) = \widehat{q}_N(\theta) - \widehat{\Omega}(\theta, \theta_0)\widehat{\Omega}(\theta_0, \theta_0)^{-1}\widehat{q}_N(\theta_0)$.

Step 5: Calculate the conditional critical value $c_\alpha(\tilde{h})$ as

$$c_\alpha(\tilde{h}) = \min\{c : P(R(\xi, h_N, \widehat{\Omega}) > c) \leq \alpha\}.$$

Step 6: Reject the null hypothesis $H_0 : S_N \in \mathcal{S}_0$ when $R(\xi, h_N, \widehat{\Omega}) \geq c_\alpha(h_N)$, report the $(1-\alpha)$ confidence interval: $CI_\alpha = \{\theta : R(\xi, h_N, \widehat{\Omega}) \leq c_\alpha(h_N)\}$.

Table of Contents

Notations

- Let $c > 0$, $c_0 \geq 0$, $c_1 \geq 0$, $C_1 > 0$ be finite constants, and $a_N = p \vee N$.
- Let $\{\Delta_N\}_{N\geq 1}$, $\{\delta_N\}_{N\geq 1}$(estimation errors) be sequences of positive constants that converges to zero such that $\delta_N \geq N^{-1/2}$.
- Let $\|\delta\|_0$ represent the number of non-zero components of δ.
- Let $P \in \mathcal{P}_N$ be the probability law of $\{W_i\}_{i=1}^N$.
- Let $\mathcal{P}_0$ be family of distribution consistent with the null.
- We use $a \lesssim b$ to denote $a \leq cb$ for some $c > 0$ that does not depends on N.
- The sequence $\{M_N\}_{N\geq 1}$ be a set of positive constants such that $M_N \geq (\mathrm{E}_P[(Z_i \vee \|X_i\|_\infty)^{2q}])^{1/2q}$.

Assumption: Regularity Conditions for the LATE

For $\boldsymbol{P} \in \boldsymbol{\mathcal{P}_N}$, the following conditions hold.

(i) The equations are satisfied with binary variables $\boldsymbol{D}$ and $\boldsymbol{Z}$.

$$\left.\begin{array}{ll} \boldsymbol{D = m_0(Z, X) + v}, & \mathbf{E}_{\boldsymbol{P}}[\boldsymbol{v}|\boldsymbol{Z, X}] = \mathbf{0} \\ \boldsymbol{Y = g_0(Z, X) + u}, & \mathbf{E}_{\boldsymbol{P}}[\boldsymbol{u}|\boldsymbol{Z, X}] = \mathbf{0} \end{array}\right\} \rightarrow \begin{array}{c} \text{Independence.} \\ \boldsymbol{(Y, D) \perp\!\!\!\perp Z|X} \end{array}$$

$$\boldsymbol{Z = p_0(X) + e}, \quad \mathbf{E}_{\boldsymbol{P}}[\boldsymbol{e}|\boldsymbol{X}] = \mathbf{0}.$$

(ii) For some $\boldsymbol{\varepsilon > 0}$, $\boldsymbol{\varepsilon \leq P(Z = 1|X) \leq 1 - \varepsilon}$ almost surely.

(iii) $\boldsymbol{\Theta}$ is compact.

(iv) $\mathbf{E}_{\boldsymbol{P}}[\boldsymbol{D}|\boldsymbol{Z} = \mathbf{1}] \geq \mathbf{E}_{\boldsymbol{P}}[\boldsymbol{D}|\boldsymbol{Z} = \mathbf{0}]$. Assumption Comparison

(v) $\|\boldsymbol{u}\|_{\boldsymbol{P},\mathbf{2}} \geq \boldsymbol{c_0}$, and $\|\mathbf{E}_{\boldsymbol{P}}[\boldsymbol{u}^{\mathbf{2}}|\boldsymbol{X}]\|_{\boldsymbol{P},\infty} \leq \boldsymbol{c_1}$.

(vi) $\|\boldsymbol{Y}\|_{\boldsymbol{P},\boldsymbol{q}} \leq \boldsymbol{c_1}$.

Assumption: Nuisance Parameter Estimators

- Sparse eigenvalue conditions: with probability $1 - o(1)$, for some $l_N \to \infty$ slow enough, we have

$$1 \lesssim \phi_{\min}(l_N s_N) \leq \phi_{\max}(l_N s_N) \lesssim 1.$$

spare eigenvalue

- Sparsity: $\|\beta^0_{12}\|_0 + \|\beta^0_{22}\|_0 + \|\gamma^0\|_0 \leq s_N$.

- Parameters: $\|\beta^0_{12}\| + \|\beta^0_{22}\| + \|\gamma^0\| \leq C_1$.

- Covariates: for $q > 4$,
  - $\inf_{\|\xi\|=1} \mathrm{E}_P[((Z_i, X_i')\xi)^2] \geq c$.
  - $\sup_{\|\xi\|=1} \mathrm{E}_P[((Z_i, X_i')\xi)^2] \leq C_1$.
  - $N^{-1/2+2/q} M_N^2 s_N \log^2 a_N \leq \Delta_N$.

Main Result

Propose an empirical process

$$\mathbb{G}_N(\cdot) = \underbrace{\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \big(\psi(W_i; \cdot, \eta_0)}_{q_N(\cdot)} - \mathbf{E}_P[\psi(W; \cdot, \eta_0)]\big),$$

and its estimator as

$$\widehat{\mathbb{G}}_N(\theta) = \underbrace{\sqrt{N}\Big(\frac{1}{N} \sum_{k=1}^{K} \sum_{i \in I_k} \psi(W_i; \theta, \widehat{\eta}_k)}_{\widehat{q}_N(\theta)} - \mathbf{E}_P\left[\psi(W_i; \theta, \widehat{\eta}_k)\right]\Big).$$

Theorem 1

Suppose that the above assumptions are satisfied. Under the null, we have

$$\widehat{\mathbb{G}}_N(\theta) = \mathbb{G}_N(\theta) + O_P(N^{-1/2}\delta_N).$$

The process $\widehat{\mathbb{G}}_N(\cdot)$ *weakly converges to a centered Gaussian process* $\mathbb{G}(\cdot)$ *for all* $P \in \mathcal{P}_0$ *as* $N \to \infty$ *with covariance function* $\Omega(\theta_1, \theta_2) = \mathbf{E}_P[(\psi(W; \theta_1, \eta_0) - \mathbf{E}_P[\psi(W; \theta_1, \eta_0)])\,(\psi(W; \theta_2, \eta_0) - \mathbf{E}_P[\psi(W; \theta_2, \eta_0)])]$.

Variance Estimation

Theorem 2

Under the same set of assumptions as above, the covariance function $\boldsymbol{\Omega(\theta_1, \theta_2)}$ *can be consistently estimated uniformly for all* $\boldsymbol{P \in \mathcal{P}_0}$ *by*

$$\widehat{\Omega}(\theta_1, \theta_2) = \frac{1}{N} \sum_{k=1}^{K} \sum_{i \in I_k} \psi(W_i; \theta_1, \widehat{\eta}_k) \psi(W_i; \theta_2, \widehat{\eta}_k) - \frac{1}{N^2} \sum_{k,k'=1}^{K} \sum_{i \in I_k, i' \in I_{k'}} \psi(W_i; \theta_1, \widehat{\eta}_k) \psi(W_{i'}; \theta_2, \widehat{\eta}_{k'})$$

and for any $\boldsymbol{\varepsilon > 0}$,

$$\lim_{N \to \infty} \sup_{P \in \mathcal{P}_0} P\left\{ \sup_{\theta_1, \theta_2} \|\widehat{\Omega}(\theta_1, \theta_2) - \Omega(\theta_1, \theta_2)\| > \varepsilon \right\} = 0.$$

How It Works

- Weak convergence over $\boldsymbol{\Theta_I}$:
  1. the convergence of the finite dimensional distribution of $\widehat{\mathbb{G}}_{\boldsymbol{N}}(\boldsymbol{\theta})$ for $\boldsymbol{\theta} \in \boldsymbol{\Theta_I}$.
  2. the stochastic equicontinuity of $\widehat{\mathbb{G}}_{\boldsymbol{N}}(\boldsymbol{\theta})$ over $\boldsymbol{\Theta_I}$:

$$\lim_{\delta \to 0} \limsup_{N \to \infty} \boldsymbol{P}\left(\sup_{|\theta_1 - \theta_2| \leq \delta} |\widehat{\mathbb{G}}_{\boldsymbol{N}}(\boldsymbol{\theta_1}) - \widehat{\mathbb{G}}_{\boldsymbol{N}}(\boldsymbol{\theta_2})| > \boldsymbol{\varepsilon_1} \right) = \mathbf{0},$$

     for any $\boldsymbol{\varepsilon_1} > \mathbf{0}$ and $\boldsymbol{\theta_1}, \boldsymbol{\theta_2} \in \boldsymbol{\Theta_I}$.
  3. the boundedness of $\boldsymbol{\Theta_I}$.

- The equivalence between testing $\boldsymbol{\theta} \in \boldsymbol{\Theta_I}$ and $\boldsymbol{P} \in \boldsymbol{\mathcal{P}_0}$.

$\hookrightarrow$ Uniformly consistent results for $\widehat{\mathbb{G}}_{\boldsymbol{N}}(\cdot)$.

Size Control

Theorem 3

Under the same set of assumptions above, the test that rejects the null hypothesis $\boldsymbol{H_0 : S_N \in \mathcal{S}_0}$ *when* $\boldsymbol{R(q_N(\theta_0), h_N, \Omega)}$ *exceeds the* $\boldsymbol{(1-\alpha)}$ *quantile* $\boldsymbol{c_\alpha(h_N)}$ *of its conditional distribution given* $\boldsymbol{h_N(\cdot)}$ *has uniformly correct asymptotic size. Under the null, we have*

$$\lim_{N\to\infty} \sup_{P\in\mathcal{P}_0} P(R(\widehat{q}_N(\theta_0), h_N, \widehat{\Omega}) > c_\alpha(h_N)) = \alpha.$$

Table of Contents

Simulation Setup

- Primitive random vector $\boldsymbol{X}_i'$ is constructed by

$$\boldsymbol{X_i \sim \mathcal{N}}\left(\mathbf{0}, \begin{pmatrix} \boldsymbol{U^0} & \boldsymbol{U^1} & \cdots & \boldsymbol{U^{\dim(X)-2}} & \boldsymbol{U^{\dim(X)-1}} \\ \boldsymbol{U^1} & \boldsymbol{U^0} & \cdots & \boldsymbol{U^{\dim(X)-3}} & \boldsymbol{U^{\dim(X)-2}} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \boldsymbol{U^{\dim(X)-2}} & \boldsymbol{U^{\dim(X)-3}} & \cdots & \boldsymbol{U^0} & \boldsymbol{U^1} \\ \boldsymbol{U^{\dim(X)-1}} & \boldsymbol{U^{\dim(X)-2}} & \cdots & \boldsymbol{U^1} & \boldsymbol{U^0} \end{pmatrix}\right)$$

with $\boldsymbol{U} = \mathbf{0.5}$.

- Consider $\boldsymbol{N} = \mathbf{500}$, $\mathbf{dim}(\boldsymbol{X}) = \mathbf{5}$, $\underbrace{\mathbf{200}, \mathbf{400}, \text{and } \mathbf{600}.}_{\text{high-dimensional controls}}$

Simulation Setup, Continued

- Consider the threshold crossing model:
  - The latent tendency to receive treatment $\boldsymbol{\delta_i \sim \mathcal{N}(0,1)}$.
  - The treatment assignment is given by $\boldsymbol{Z_i = \mathbb{1}\{\delta_i \geq 0\}}$.
  - The potential treatment indicators $\boldsymbol{D_i(Z_i)}$ are given by

$$\boldsymbol{D_i(0) = \mathbb{1}\{\Phi(\delta_i) < P_{AT}\}, \quad D_i(1) = \mathbb{1}\{\Phi(\delta_i) < 1 - P_{NT}\},}$$

    with $\boldsymbol{\Phi(\cdot)}$ denotes the CDF of a standard normal distribution.
  - The target parameter is set to $\boldsymbol{\theta_0 = 1}$.
  - The outcome $\boldsymbol{Y_i = D_i + X_i + \varepsilon_i}$ with $\boldsymbol{\varepsilon_i \sim \mathcal{N}(0,1)}$.

- Scenarios:
  - Strongly identified case: $\boldsymbol{(P_{AT}, P_{NT}) = (0.25, 0.25) \to P_C = 0.5}$
  - Weakly identified case: $\boldsymbol{(P_{AT}, P_{NT}) = (0.45, 0.45) \to P_C = 0.1}$

Completely unidentified

Results

I compare the proposed method HD-QLR (this paper) with

- the conditional QLR test (AM16[2]) : robust against weak identification but not against high dimensionality.
- ML methods (CCDDHNR18[3] and BCFH17[4]): robust against high dimensionality but not against weak identification.

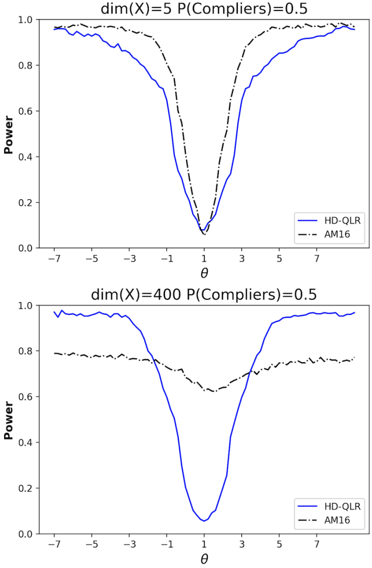
[2] Andrews and Mikusheva (2016).

[3] Chernozhukov, Chetverikov, Demirer, Duflo, Hansen, Newey, and Robins (2018).

[4] Belloni, Chernozhukov, Fernandez-Val, and Hansen (2017).

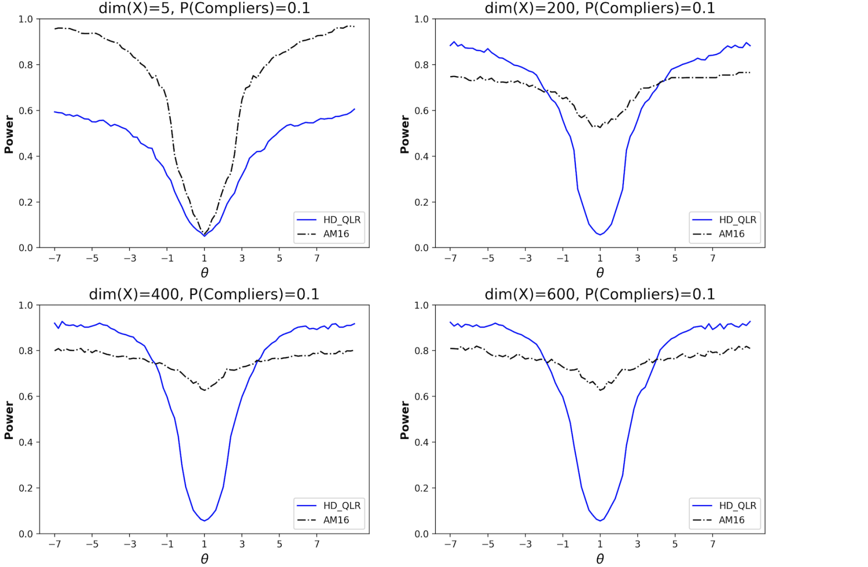
Comparison: Strong Identification ▸ Power Curve

- Power curve of nominal 5%: AM16 and HD-QLR (this paper)



Comparison: Weak Identification

- Power curve of nominal 5%: AM16 and HD-QLR (this paper)



Results

I compare the proposed method HD-QLR (this paper) with

- the conditional QLR test (AM16[2]) : robust against weak identification but not against high dimensionality.
- ML methods (CCDDHNR18[3] and BCFH17[4]): robust against high dimensionality but not against weak identification.

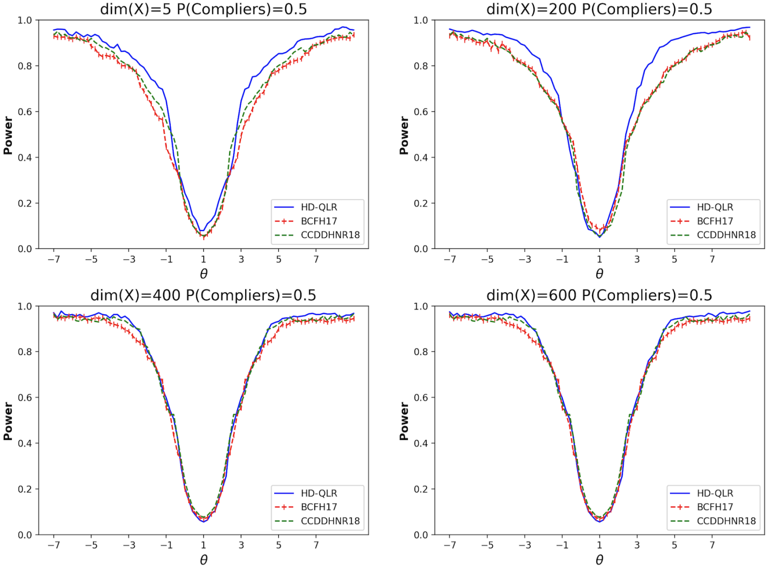
[2] Andrews and Mikusheva (2016).

[3] Chernozhukov, Chetverikov, Demirer, Duflo, Hansen, Newey, and Robins (2018).

[4] Belloni, Chernozhukov, Fernandez-Val, and Hansen (2017).

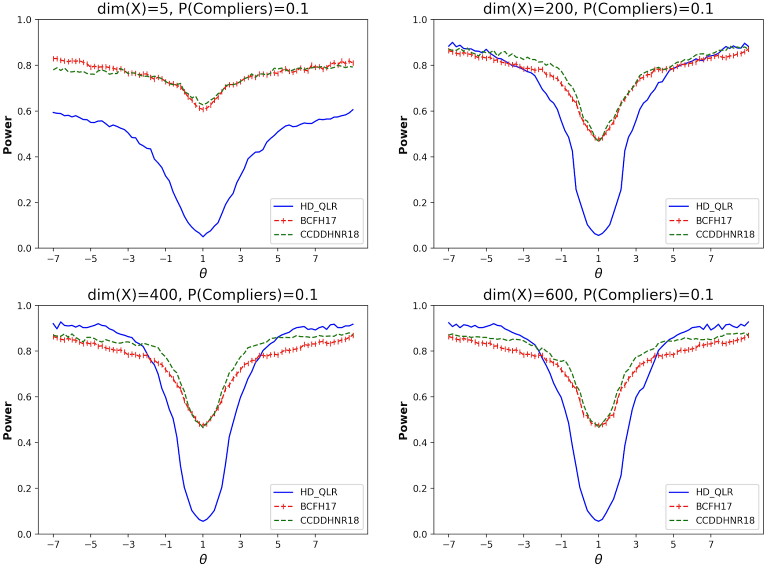
Comparison: Strong Identification

- Power curve: CCDDHNR18, BCFH17 and HD-QLR (this paper)



Comparison: Weak Identification

- Power curve: CCDDHNR18, BCFH17 and HD-QLR (this paper)



Comparison Across Four Approaches

AM16

- Neyman orthogonal score ψ
- test statistics R
- use traditional methods to handle X_i

CCDDHNR18 BCFH17

- Neyman orthogonal score ψ
- normal t-test
- use ML to handle X_i

Drawbacks

- overfitting
- multicollinearity

- cannot perform well under weakly identified scenarios

The proposed HD-QLR takes advantage from both methods:

- Neyman orthogonal score ψ
- test statistics R
- use ML to handle the high-dimensional X_i

Comparison Across Four Approaches

AM16

- Neyman orthogonal score ψ
- test statistics R
- use traditional methods to handle X_i

CCDDHNR18 BCFH17

- Neyman orthogonal score ψ
- normal t-test
- use ML to handle X_i

Drawbacks

AM16:

- overfitting
- multicollinearity

CCDDHNR18 BCFH17:

- cannot perform well under weakly identified scenarios

The proposed HD-QLR takes advantage from both methods:

- Neyman orthogonal score ψ
- test statistics R
- use ML to handle the high-dimensional X_i

Comparison Across Four Approaches

AM16

- ➤ Neyman orthogonal score ψ
- ➤ test statistics R
- ➤ use traditional methods to handle X_i

CCDDHNR18 BCFH17

- ➤ Neyman orthogonal score ψ
- ➤ normal t-test
- ➤ use ML to handle X_i

Drawbacks

- ○ overfitting
- ○ multicollinearity

- ○ cannot perform well under weakly identified scenarios

The proposed HD-QLR takes advantage from both methods:

- Neyman orthogonal score ψ
- test statistics R
- use ML to handle the high-dimensional X_i

Table of Contents

Example One: Hornung (2015) "Railroads and growth in Prussia"

- ⋆ Data: highly detailed city-level data from the historical German state of Prussia.

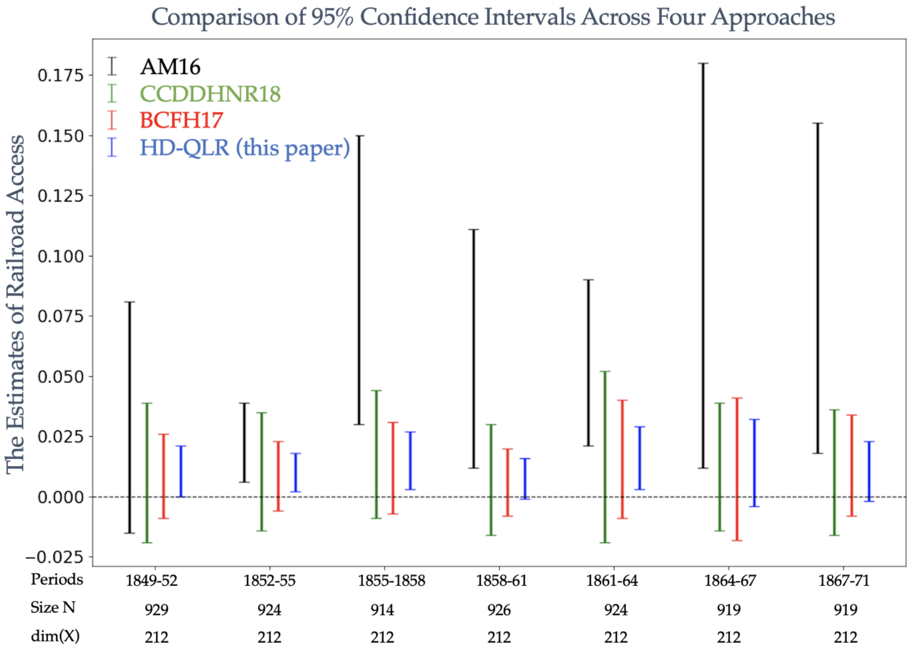
- $\boldsymbol{Y_{it}}$ **:** urban population growth rate for city $\boldsymbol{i}$ during time period $\boldsymbol{t}$.

- $\boldsymbol{D_i}$ **:** whether the city $\boldsymbol{i}$ was connected to the railroad in 1848.

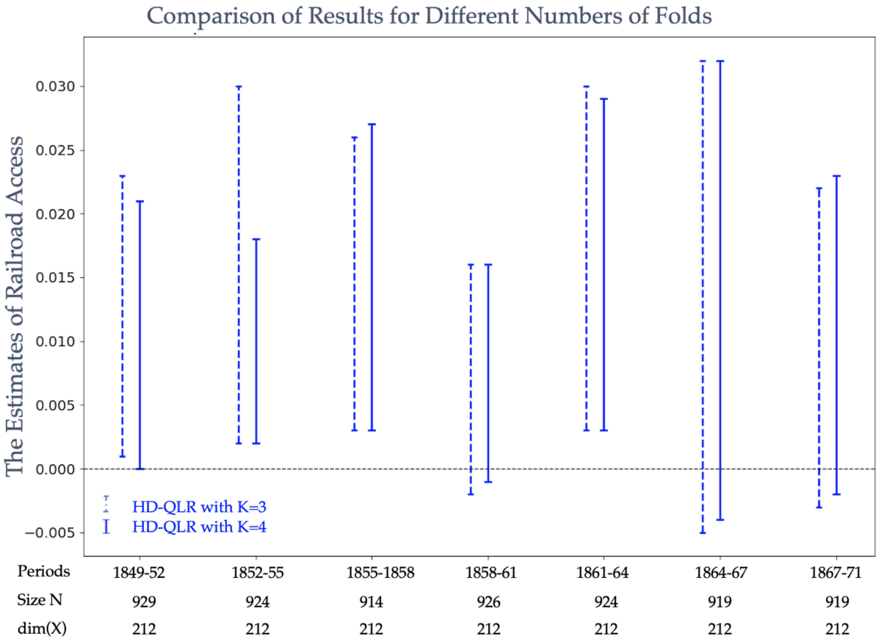
- $\boldsymbol{Z_i}$ **:** whether the city $\boldsymbol{i}$ was located within a straight-line corridor between two important cities.

- $\boldsymbol{X_i}$ **:** whether the city had access to the main roads, whether the city had waterway access, military population, age composition, school enrollment rate, etc.

Results


Comparison of 95% Confidence Intervals Across Four Approaches
AM16
CCDDHNR18
BCFH17
HD-QLR (this paper)
The Estimates of Railroad Access
0.175
0.150
0.125
0.100
0.075
0.050
0.025
0.000
−0.025
Periods
1849-52
1852-55
1855-1858
1858-61
1861-64
1864-67
1867-71
Size N
929
924
914
926
924
919
919
dim(X)
212
212
212
212
212
212
212

Results, Continued



Example Two: Ambrus, Field, and Gonzalez (2020) "The Impact on Housing Prices of A Cholera Epidemic"

- ⋆ "*In August 1854, St. James experienced a sudden outbreak of cholera when one of the 13 shallow wells that serviced the parish, the Broad Street pump, became contaminated with cholera bacteria.*"

- $\boldsymbol{Y_i}$ **:** the log rental price of house $\boldsymbol{i}$ in 1864.

- $\boldsymbol{D_i}$ **:** whether house $\boldsymbol{i}$ had at least one cholera death.

- $\boldsymbol{Z_i}$ **:** whether house $\boldsymbol{i}$ fell inside the contaminated areas.

- $\boldsymbol{X_i}$ **:** distance to the closest pump, distance to the fire station, distance to the urinal, sewer access, among a total of 23 variables.

Results

	AM16	CCDDHNR18	BCFH17	HD-QLR (this paper)
95% CI	[-2.160, -0.670]	[-1.132, 0.406]	[-1.291, 0.576]	[-1.080, 0.035]
length of CI	1.490	1.538	1.866	1.115

Table: Displayed are the CIs and the length of CI. Inference results are based on 10 iterations of resampled cross fitting with $\boldsymbol{K = 4}$ folds for cross fitting. The number of observations $\boldsymbol{N = 467}$.

Table of Contents

Takeaways

- I develop an inference method for the high-dimensional LATE, independent of the strength of identification.

	Low-dimensional Model	High-dimensional Model
Strong Identification	z-test	CCDDHNR18, BCFH17
Weak Identification	AM16	HD-QLR (my paper)

- The proposed method has uniformly correct asymptotic size.
- The proposed test is robust against weak identification and high dimensionality, outperforming other conventional methods.
- The proposed method yields narrower confidence intervals than conventional methods, as demonstrated in applications.

Takeaways

- I develop an inference method for the high-dimensional LATE, independent of the strength of identification.

	Low-dimensional Model	High-dimensional Model
Strong Identification	z-test	CCDDHNR18, BCFH17
Weak Identification	AM16	HD-QLR (my paper)

- The proposed method has uniformly correct asymptotic size.
- The proposed test is robust against weak identification and high dimensionality, outperforming other conventional methods.
- The proposed method yields narrower confidence intervals than conventional methods, as demonstrated in applications.

Thank you!

feel free to email me any comments
yukun.ma@vanderbilt.edu

Motivation: Lee et al. (2022)

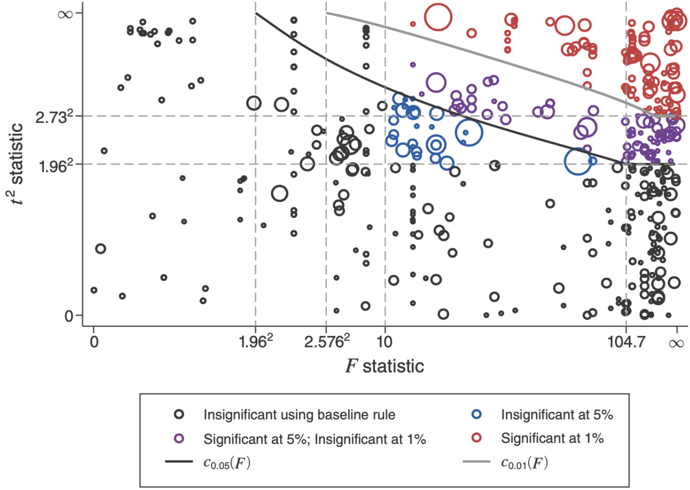


Figure: *American Economic Review* 2013-2019

back

Tuning Parameters

Lemma (Convergence rate for Lasso with logistic model)

Suppose some regularity assumptions hold. In addition, suppose that the penalty choice $\boldsymbol{\lambda_1 = K_1\sqrt{N\log(pN)}}$ *and* $\boldsymbol{\lambda_2 = K_2\sqrt{N\log(pN)}}$ *for* $\boldsymbol{K_1, K_2 > 0}$*. Then with probability* $\boldsymbol{1 - o(1)}$*,*

$$\|(\widehat{\beta}_{11}, \widehat{\beta}_{12}) - (\beta_{11}^0, \beta_{12}^0)\| \vee \|\widehat{\gamma} - \gamma^0\| \lesssim \sqrt{\frac{s_N \log(pN)}{N}}.$$

Lemma (Convergence rate for Lasso with OLS)

Suppose some regularity assumptions hold. Moreover, suppose that the penalty choice $\boldsymbol{\lambda_3 = K_3\sqrt{N\log(pN)}}$ *for* $\boldsymbol{K_3 > 0}$*. Then with probability* $\boldsymbol{1 - o(1)}$*,*

$$\|(\widehat{\beta}_{21}, \widehat{\beta}_{22}) - (\beta_{21}^0, \beta_{22}^0)\| \lesssim \sqrt{\frac{s_N \log(pN)}{N}}.$$

back

Null Hypothesis

- Define $\boldsymbol{S_N(\cdot) = \mathbf{E}_P[N^{-1/2}\sum_{i=1}^{N}\psi(W_i;\cdot,\eta_0)]}$.

  Case 1: $\boldsymbol{H_0 : \theta = \theta_0}$ with assuming $\boldsymbol{\theta}$ is point-identified
  $\hookrightarrow$ $\boldsymbol{S_N(\theta_0) = 0}$.

  Case 2: $\boldsymbol{H_0 : \theta \in \Theta_I}$ with the identified set $\boldsymbol{\Theta_I \subset \Theta}$ when point identification fails
  $\hookrightarrow$ $\boldsymbol{S_N(\theta) = 0}$ for $\boldsymbol{\forall\theta \in \Theta_I}$.

- Let $\boldsymbol{\mathcal{S}_0}$ be the collection of function $\boldsymbol{S_N(\cdot)}$ satisfying $\boldsymbol{S_N(\theta) = 0}$.
  $\hookrightarrow$ $\boldsymbol{H_0' : S_N(\cdot) \in \mathcal{S}_0}$.

back

Spare eigenvalue

For any $\boldsymbol{T} \subset [\boldsymbol{p}+\mathbf{1}]$, $\boldsymbol{\delta} = (\boldsymbol{\delta}_{\mathbf{1}}, \cdots, \boldsymbol{\delta}_{\boldsymbol{p}+\mathbf{1}})' \in \mathbb{R}^{\boldsymbol{p}+\mathbf{1}}$ with $\boldsymbol{\delta}_{\boldsymbol{T},\boldsymbol{j}} = \boldsymbol{\delta}_{\boldsymbol{j}}$ if $\boldsymbol{j} \in \boldsymbol{T}$ and $\boldsymbol{\delta}_{\boldsymbol{T},\boldsymbol{j}} = \mathbf{0}$ if $\boldsymbol{j} \notin \boldsymbol{T}$. Define the minimum and maximum sparse eigenvalue by

$$\phi_{\min}(m) = \inf_{\|\delta\|_0 \leq m} \frac{\|(Z_i, X_i')\delta\|_{2,N}}{\|\delta_T\|_1}$$

$$\phi_{\max}(m) = \sup_{\|\delta\|_0 \leq m} \frac{\|(Z_i, X_i')\delta\|_{2,N}}{\|\delta_T\|_1}.$$

back

Identification Assumption Comparison

- In my paper:

$$\mathbf{E}_{\boldsymbol{P}}[\boldsymbol{D}|\boldsymbol{Z} = \mathbf{1}] \geq \mathbf{E}_{\boldsymbol{P}}[\boldsymbol{D}|\boldsymbol{Z} = \mathbf{0}].$$

- In weak identification literature:

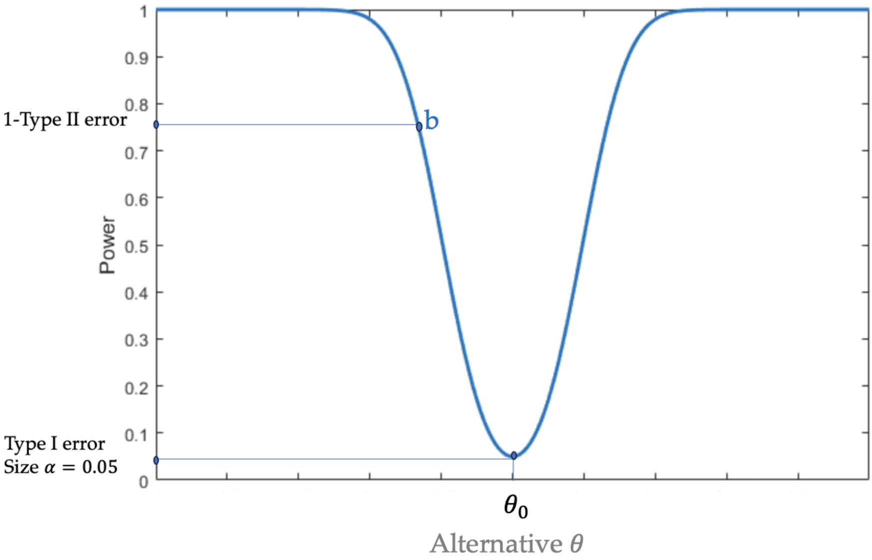
$$\mathbf{E}_{\boldsymbol{P}}[\boldsymbol{D}|\boldsymbol{Z} = \mathbf{1}] - \mathbf{E}_{\boldsymbol{P}}[\boldsymbol{D}|\boldsymbol{Z} = \mathbf{0}] = \frac{\boldsymbol{C}_{\mathbf{1}}}{\sqrt{\boldsymbol{N}}} \quad \text{with } \boldsymbol{C}_{\mathbf{1}} > \mathbf{0}.$$

- In ML literature:

$$\mathbf{E}_{\boldsymbol{P}}[\boldsymbol{D}|\boldsymbol{Z} = \mathbf{1}] - \mathbf{E}_{\boldsymbol{P}}[\boldsymbol{D}|\boldsymbol{Z} = \mathbf{0}] \geq \boldsymbol{C}_{\mathbf{2}} \quad \text{with } \boldsymbol{C}_{\mathbf{2}} > \mathbf{0}.$$

back

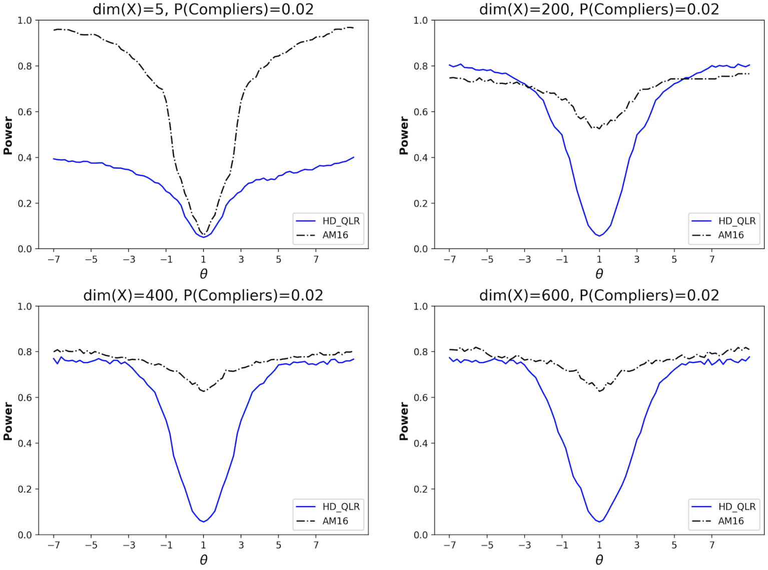
Power Curve



back

Comparisons: Unidentified case

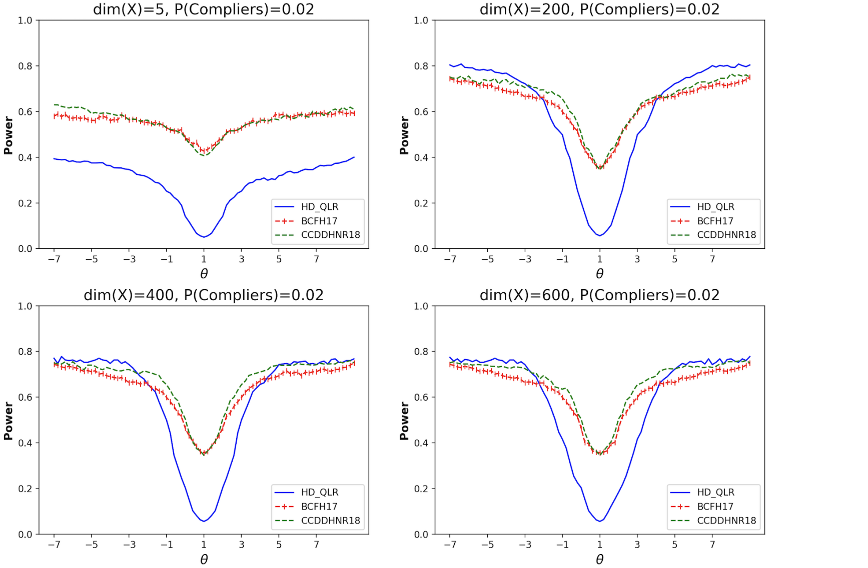
- Power curve: AM16 with HD-QLR (this paper)



Comparisons: Unidentified Case

back

- Power curve: CCDDHNR18, BCFH17 with HD-QLR (this paper)



Belloni, A., V. Chernozhukov, and K. Kato (2015): "Uniform post-selection inference for least absolute deviation regression and other Z-estimation problems," *Biometrika*, 102, 77–94.

Blandhol, C., J. Bonney, M. Mogstad, and A. Torgovitsky (2022): "When is TSLS actually late?" Tech. rep., National Bureau of Economic Research.

Chernozhukov, V., D. Chetverikov, and K. Kato (2013): "Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors," *The Annals of Statistics*, 41, 2786–2819.

——— (2016): "Empirical and multiplier bootstraps for suprema of empirical processes of increasing complexity, and related Gaussian couplings," *Stochastic Processes and their Applications*, 126, 3632–3651.

——— (2017): "Central limit theorems and bootstrap in high dimensions," *The Annals of Probability*, 45, 2309–2352.

Kleibergen, F. (2005): "Testing parameters in GMM without assuming that they are identified," *Econometrica*, 73, 1103–1123.

Moreira, M. J. (2003): "A conditional likelihood ratio test for structural models," *Econometrica*, 71, 1027–1048.

Stock, J. H. and J. H. Wright (2000): "GMM with weak identification," *Econometrica*, 68, 1055–1096.

Tan, Z. (2006): "Regression and weighting methods for causal inference using instrumental variables," *Journal of the American Statistical Association*, 101, 1607–1618.

van der Vaart, A. and J. A. Wellner (1996): "WEAK CONVERGENCE AND EMPIRICAL PROCESSES," .