The Synthetic Instrument

Linbo Wang
Univeristy of Toronto

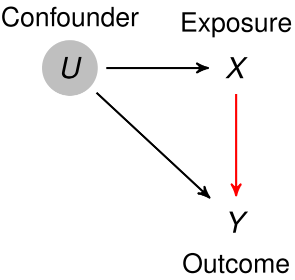
Pacific Causal Inference Conference
September 16, 2022

Acknowledgements

Dingke Tang

- Third-year PhD student in U Toronto

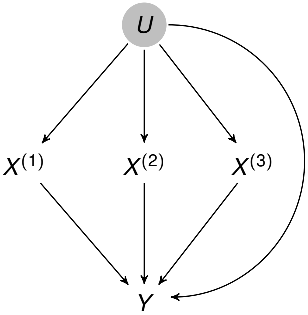
Causal inference with unmeasured confounding



- **Target:** mean potential outcome $E[Y(x)]$
- **Challenge:** often not possible to measure all the confounders

$$E[Y(x)] = E_{U} E[Y \mid X = x, U]$$

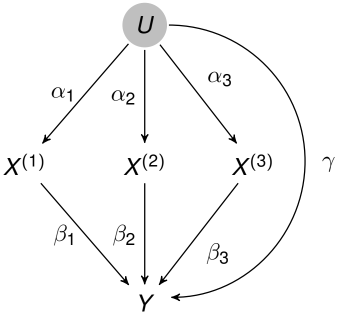
Multi-cause causal inference (Wang and Blei, 2019)



- Multiple treatments; One outcome
- Shared confounding among treatments

$$X^{(1)} \perp\!\!\!\perp X^{(2)} \perp\!\!\!\perp \ldots X^{(p)} \mid U$$

Model Setup

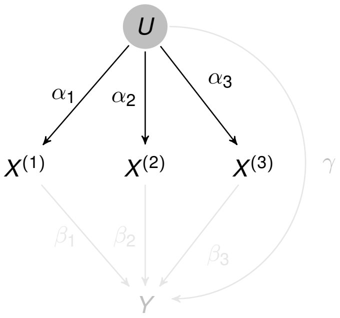


Assume linear models

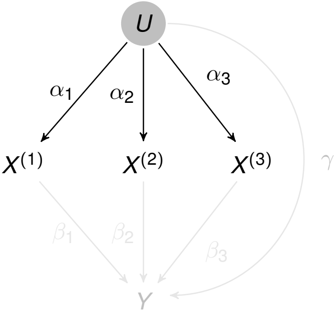
$$\boldsymbol{X} = U\boldsymbol{\alpha} + \epsilon_X;$$
$$Y = \boldsymbol{X}^T\boldsymbol{\beta} + U\gamma + \epsilon_Y.$$

Interested in estimating the causal parameters $\boldsymbol{\beta}$

Estimating α



Estimating α: Standard factor analysis ($p = 3$)

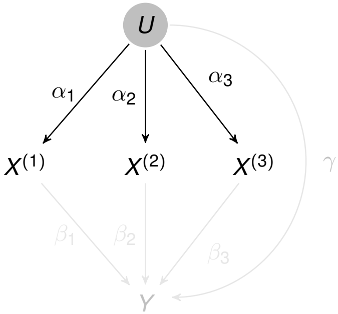


Under the linear treatment model,

$$X^{(1)} = \alpha_1 U + \epsilon_1;$$
$$X^{(2)} = \alpha_2 U + \epsilon_2;$$
$$X^{(3)} = \alpha_3 U + \epsilon_3,$$

we can identify $\alpha_1, \alpha_2, \alpha_3$ (up to sign)

Estimating α: Standard factor analysis ($p = 3$)



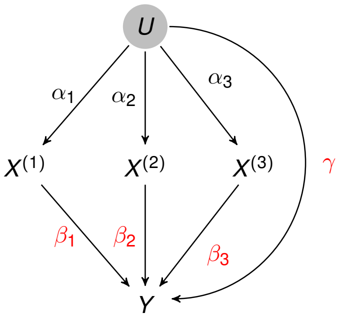
Three observed quantities:

$$Cov(X^{(1)}, X^{(2)}), Cov(X^{(1)}, X^{(3)}), Cov(X^{(2)}, X^{(3)})$$

Three unknown parameters:

$$\alpha_1, \alpha_2, \alpha_3$$

Estimating $\boldsymbol{\beta}$



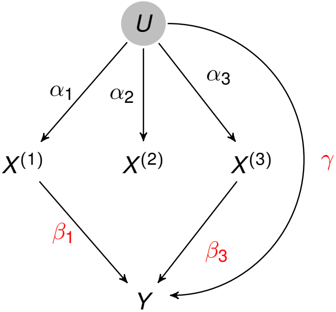
Three observed quantities:

$$Cov(X^{(1)}, Y), Cov(X^{(2)}, Y), Cov(X^{(3)}, Y)$$

Four unknown parameters:

$$\beta_1, \beta_2, \beta_3, \gamma$$

Assuming known “Negative Treatment”



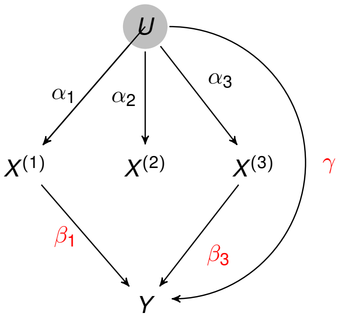
Three observed quantities:

$$Cov(X^{(1)}, Y), Cov(X^{(2)}, Y), Cov(X^{(3)}, Y)$$

Three unknown parameters ($\beta_2 = 0$):

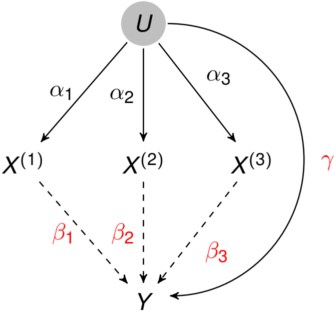
$$\beta_1, \beta_3, \gamma$$

Assuming known “Negative Treatment”



- This relates to the negative control approach in causal inference
- **Problem:** Need to know which treatment is “negative”

This talk: Assume sparse treatment effects



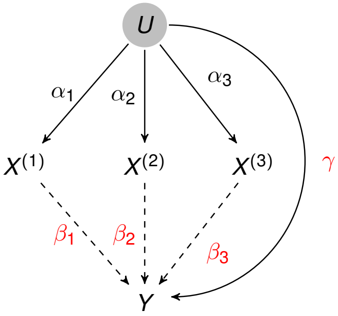
Assumption: $\|\boldsymbol{\beta}\|_0 \leq 1$

- Causal effects $\beta_1, \beta_2, \beta_3$ are identifiable
- A simple and computationally efficient algorithm to estimate the causal effect

Sparse treatment effects: Identifiability

Sparse treatment effects: Estimation

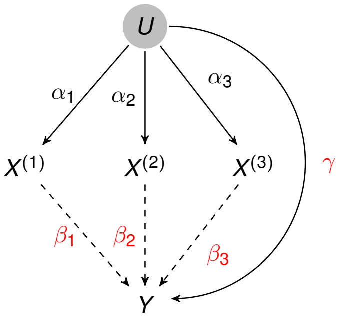
Identification under sparsity: $\|\boldsymbol{\beta}\|_0 \leq 1$



Suppose truth is $\dot{\beta}_1 = \dot{\beta}_2 = 0, \dot{\beta}_3 \neq 0$:

Voter guess	$\widehat{\beta}_1$	$\widehat{\beta}_2$	$\widehat{\beta}_3$
$\beta_1 = 0$	0	0	β_3

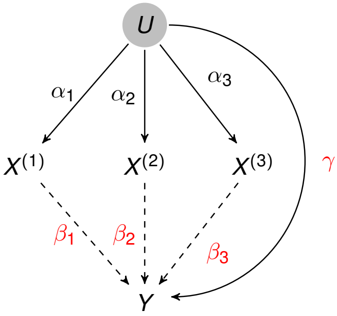
Identification under sparsity: $\|\boldsymbol{\beta}\|_0 \leq 1$



Suppose truth is $\dot{\beta}_1 = \dot{\beta}_2 = 0, \dot{\beta}_3 \neq 0$:

Voter guess	$\widehat{\beta}_1$	$\widehat{\beta}_2$	$\widehat{\beta}_3$
$\beta_1 = 0$	0	0	β_3
$\beta_2 = 0$	0	0	β_3

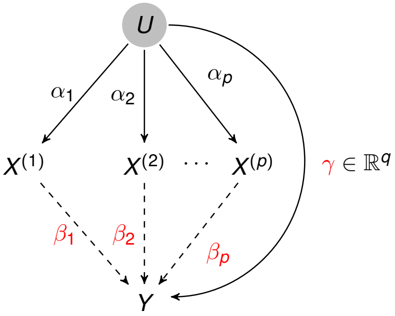
Identification under sparsity: $\|\boldsymbol{\beta}\|_0 \le 1$



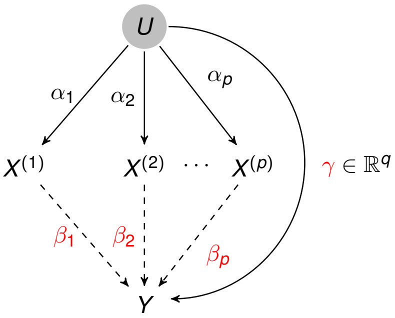
Suppose truth is $\dot\beta_1 = \dot\beta_2 = 0, \dot\beta_3 \neq 0$:

Voter guess	$\widehat{\beta}_1$	$\widehat{\beta}_2$	$\widehat{\beta}_3$
$\beta_1 = 0$	0	0	β_3
$\beta_2 = 0$	0	0	β_3
$\beta_3 = 0$	non-zero	non-zero	0

Voting in practice

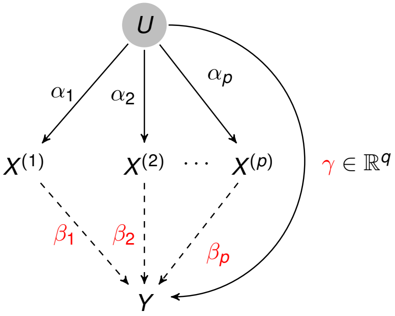


Voting in practice



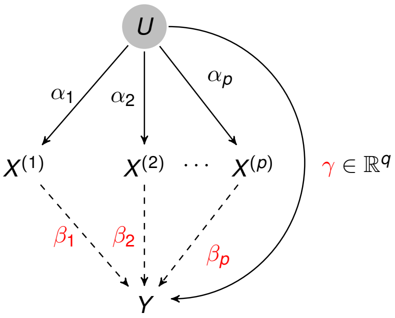
- In general, we need to compute causal effect estimates for $\binom{p}{q}$ voters

Voting in practice



- In general, we need to compare causal effect estimates for $\binom{p}{q}$ voters

Voting in practice



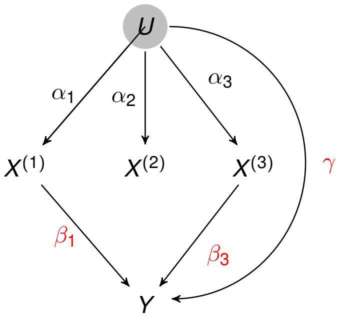
- In general, we need to compare causal effect estimates for $\binom{p}{q}$ voters

Not feasible/numerical stable if p is large!

Sparse treatment effects: Identifiability

Sparse treatment effects: Estimation

Another look from an instrumental variable (IV) perspective

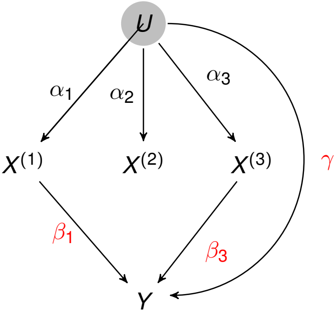


Assume Negative Treatment: $\beta_2 = 0$

Construct a **Synthetic Instrument**:

$$SIV_2^{(1)} = X^{(1)} - \frac{\alpha_1}{\alpha_2} X^{(2)}$$

The Synthetic Instrument

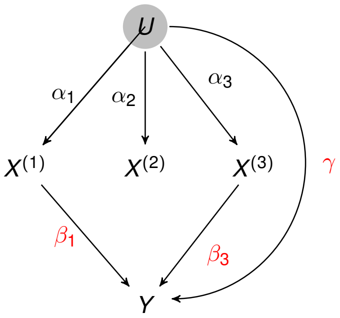


Assume Negative Treatment: $\beta_2 = 0$

Construct a **Synthetic Instrument**:

$$SIV_2^{(1)} = X^{(1)} - \frac{\alpha_1}{\alpha_2} X^{(2)} = \epsilon_1 - \frac{\alpha_1}{\alpha_2} \epsilon_2 \text{ is an IV for } X^{(1)}$$

The Synthetic Instrument



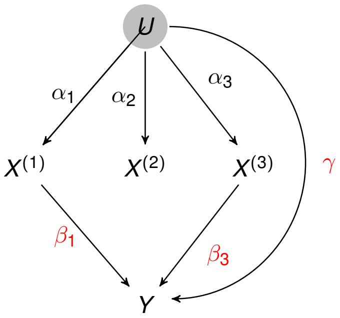
Assume Negative Treatment: $\beta_2 = 0$

Construct a **Synthetic Instrument**:

$$SIV_2^{(1)} = X^{(1)} - \frac{\alpha_1}{\alpha_2} X^{(2)} = \epsilon_1 - \frac{\alpha_1}{\alpha_2}\epsilon_2 \text{ is an IV for } X^{(1)}$$

$$SIV_2^{(3)} = X^{(3)} - \frac{\alpha_3}{\alpha_2} X^{(2)} = \epsilon_3 - \frac{\alpha_3}{\alpha_2}\epsilon_2 \text{ is an IV for } X^{(3)}$$

The Synthetic 2SLS



Two stage least squares (2SLS):

1. Regress $\boldsymbol{X} = (X^{(1)}, X^{(2)}, X^{(3)})$ on $\mathbf{SIV}_2 = (SIV_2^{(1)}, SIV_2^{(3)})$
2. Regress Y on $\widehat{\boldsymbol{X}}$ fixing $\beta_2 = 0$

Voting with Synthetic 2SLS

1. For $j = 1, 2, 3$: Regress $\boldsymbol{X}$ on $\mathbf{SIV}_j \quad \Rightarrow \quad \widehat{\boldsymbol{X}}^{(j)}$
2. For $j = 1, 2, 3$: Regress Y on $\widehat{\boldsymbol{X}}^{(j)}$ fixing $\beta_j = 0$

Voting with Synthetic 2SLS

1. For $j = 1, 2, 3$: Regress $\boldsymbol{X}$ on $\mathbf{SIV}_j \quad \Rightarrow \quad \widehat{\boldsymbol{X}}^{(j)}$

Key result 1: $\widehat{\boldsymbol{X}}^{(j)}$ does not depend on j

- $\mathbf{SIV}_1, \mathbf{SIV}_2, \mathbf{SIV}_3$ span the same linear space

Voting with Synthetic 2SLS

1. For $j = 1, 2, 3$: Regress $\boldsymbol{X}$ on $\mathbf{SIV}^{(1)}$ $\quad \Rightarrow \quad$ $\widehat{\boldsymbol{X}}^{(j)}$

Voting with Synthetic 2SLS

1. ~~For $j = 1, 2, 3$:~~ Regress $\boldsymbol{X}$ on $\mathbf{SIV}^{(1)}$ $\quad \Rightarrow \quad$ $\widehat{\boldsymbol{X}}^{\cancel{(j)}}$

2 For $j = 1, 2, 3$: Regress Y on $\widehat{\boldsymbol{X}}^{\cancel{(j)}}$ fixing $\beta_j = 0$

Voter 1	$E(Y \mid \widehat{\boldsymbol{X}}) = 0\widehat{X}_1 + \beta_2\widehat{X}_2 + \beta_3\widehat{X}_3$
Voter 2	$E(Y \mid \widehat{\boldsymbol{X}}) = \beta_1\widehat{X}_1 + 0\widehat{X}_2 + \beta_3\widehat{X}_3$
Voter 3	$E(Y \mid \widehat{\boldsymbol{X}}) = \beta_1\widehat{X}_1 + \beta_2\widehat{X}_2 + 0\widehat{X}_3$

and then compare estimates

Voting with Synthetic 2SLS

1. ~~For $j = 1, 2, 3$:~~ Regress $\boldsymbol{X}$ on $\mathbf{SIV}^{(1)}$ $\quad\Rightarrow\quad$ $\widehat{\boldsymbol{X}}^{\cancel{(j)}}$

2 For $j = 1, 2, 3$: Regress Y on $\widehat{\boldsymbol{X}}^{\cancel{(j)}}$ fixing $\beta_j = 0$

Suppose $\dot{\beta}_1 = \dot{\beta}_2 = 0, \dot{\beta}_3 \neq 0$:

Voter 1	$E(Y \mid \widehat{\boldsymbol{X}}) = 0\widehat{X}_1 + 0\widehat{X}_2 + \beta_3\widehat{X}_3$
Voter 2	$E(Y \mid \widehat{\boldsymbol{X}}) = 0\widehat{X}_1 + 0\widehat{X}_2 + \beta_3\widehat{X}_3$
Voter 3	$E(Y \mid \widehat{\boldsymbol{X}}) = \beta_1\widehat{X}_1 + \beta_2\widehat{X}_2 + 0\widehat{X}_3$

and then compare estimates

Voting with Synthetic 2SLS

1. ~~For $j = 1, 2, 3$:~~ Regress $\boldsymbol{X}$ on $\textbf{SIV}^{(1)}$ $\quad \Rightarrow \quad \widehat{\boldsymbol{X}}^{\cancel{(j)}}$

2 For $j = 1, 2, 3$: Regress Y on $\widehat{\boldsymbol{X}}^{\cancel{(j)}}$ fixing $\beta_j = 0$

Suppose $\dot{\beta}_1 = \dot{\beta}_2 = 0, \dot{\beta}_3 \neq 0$:

Voter 1	$E(Y \mid \widehat{\boldsymbol{X}}) = 0\widehat{X}_1 + 0\widehat{X}_2 + \beta_3\widehat{X}_3$
Voter 2	$E(Y \mid \widehat{\boldsymbol{X}}) = 0\widehat{X}_1 + 0\widehat{X}_2 + \beta_3\widehat{X}_3$
Voter 3	$E(Y \mid \widehat{\boldsymbol{X}}) = \beta_1\widehat{X}_1 + \beta_2\widehat{X}_2 + 0\widehat{X}_3$

and then compare estimates

Key result 2: We can directly run a penalized regression

$$Y \sim \widehat{X}_1 + \widehat{X}_2 + \widehat{X}_3.$$

subject to $\|\boldsymbol{\beta}\|_0 \leq 1$.

Synthetic 2SLS for sparse treatment effects

1. ~~For $j = 1, 2, 3$:~~ Regress $\boldsymbol{X}$ on $\mathbf{SIV}^{(1)}$ $\quad \Rightarrow \quad$ $\widehat{\boldsymbol{X}}^{\cancel{(j)}}$

2 ~~For $j = 1, 2, 3$:~~ Regress Y on $\widehat{\boldsymbol{X}}$ subject to $\|(\beta_1, \beta_2, \beta_3)\|_0 \leq 1$.

Synthetic Instrument: The general case

$$\boldsymbol{X} = \boldsymbol{U}^T A + \epsilon_X;$$
$$Y = \boldsymbol{X}^T \boldsymbol{\beta} + \boldsymbol{U}\gamma + \epsilon_Y.$$

- $\boldsymbol{X} \in \mathbb{R}^p$: p may grow with n
- $\boldsymbol{U} \in \mathbb{R}^q$: $q < p$ may also grow with n
- $\epsilon_{X_1}, \epsilon_{X_2}, \ldots, \epsilon_{X_p}, \epsilon_Y, U$ are uncorrelated
- Assume $\|\beta\|_0 \leq s$

Synthetic Instrument: The general case

$$\boldsymbol{X} = \boldsymbol{U}^T A + \epsilon_X;$$
$$Y = \boldsymbol{X}^T \boldsymbol{\beta} + \boldsymbol{U}\gamma + \epsilon_Y.$$

- $\boldsymbol{X} \in \mathbb{R}^p$: p may grow with n
- $\boldsymbol{U} \in \mathbb{R}^q$: $q < p$ may also grow with n
- $\epsilon_{X_1}, \epsilon_{X_2}, \ldots, \epsilon_{X_p}, \epsilon_Y, U$ are uncorrelated
- Assume $\|\boldsymbol{\beta}\|_0 \leq s$

Synthetic instrument (SIV): Suppose $\beta_j = 0$ for $j \in C, |C| = q$. Then the synthetic instrument is a $p - q$ dimensional vector with components

$$SIV_C^{(j)} = X^{(j)} - A_j^T A_C^{-1} X_C, \quad j \in \{1, \ldots, p\} \setminus C,$$

where A_j is the jth column of $A_{q \times p}$.

Main Result 1

Theorem (Uniqueness)

$\widehat{\boldsymbol{X}} \equiv \mathbb{E}(\boldsymbol{X} \mid SIV_{\mathbf{C}})$ *does not depend on the choice of C.*

Main Result 2

Theorem (ℓ_0 optimization)

Assume that $\boldsymbol{\beta}$ is sufficiently sparse. Then under regularity conditions, $\boldsymbol{\beta}$ is identifiable via following optimization problem

$$\dot{\boldsymbol{\beta}} = \operatorname*{arg\,min}_{\boldsymbol{\beta} \in \mathbb{R}^p} \mathbb{E}(Y - \widehat{\boldsymbol{X}}^T \boldsymbol{\beta})^2,$$

subject to $||\boldsymbol{\beta}||_0 \leq (dim(\boldsymbol{X}) - dim(\boldsymbol{U}))/2$.

Main Result 2

Theorem (ℓ_0 optimization)

Assume that $\boldsymbol{\beta}$ is sufficiently sparse. Then under regularity conditions, $\boldsymbol{\beta}$ is identifiable via following optimization problem

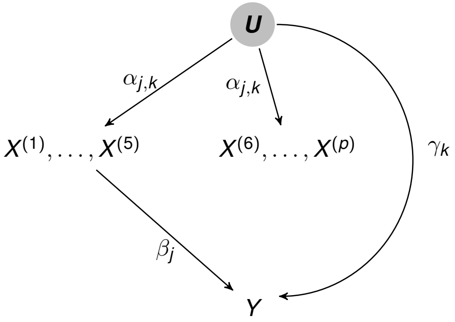
$$\dot{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta} \in \mathbb{R}^p}{\arg\min}\, \mathbb{E}(Y - \widehat{\boldsymbol{X}}^T \boldsymbol{\beta})^2,$$

subject to $||\boldsymbol{\beta}||_0 \leq (dim(\boldsymbol{X}) - dim(\boldsymbol{U}))/2.$

- Can be solved efficiently using the `L0Learn` package (Hazimeh and Mazumder, 2020)

Numerical experiments

Data generation



- $U \sim MVN(0, I_{5\times5})$
- $\alpha_{j,k} \sim Unif(-1, 1), j = 1, \ldots, p, k = 1, \ldots, 5$
- $\beta_j = 1, j = 1, \ldots, 5$
- $\gamma_k \sim Unif(-2, 2), k = 1, \ldots, 5$

Comparison Methods

- `SIV`: Synthetic 2SLS, eBIC for tunning parameter selection
- `Lasso`: Lasso, eBIC for tunning parameter selection
- `Null`: Miao et al. (2021)'s method
  - A robust linear regression based approach
  - No variable selection: all $\widehat{\beta}_j, j = 1, \ldots, p$ are non-zero
  - Only considered the low-dimensional settings (we tried an extension to high-dimensional settings)
- `Trim`: Ćevid et al. (2020) and Guo et al. (2021)'s method

  Coef($Y \sim \boldsymbol{X}$) $=$ sparse coefficient$+$non-sparse confounding bias

  - Assume the confounding bias is asymptotically negligible
  - Only consistent in high-dimensional settings

Settings

- Low-dimensional case: $p = 100, n = 200, \ldots, 5000$
- High-dimensional case: $n = 500, p = 500, \ldots, 3000$

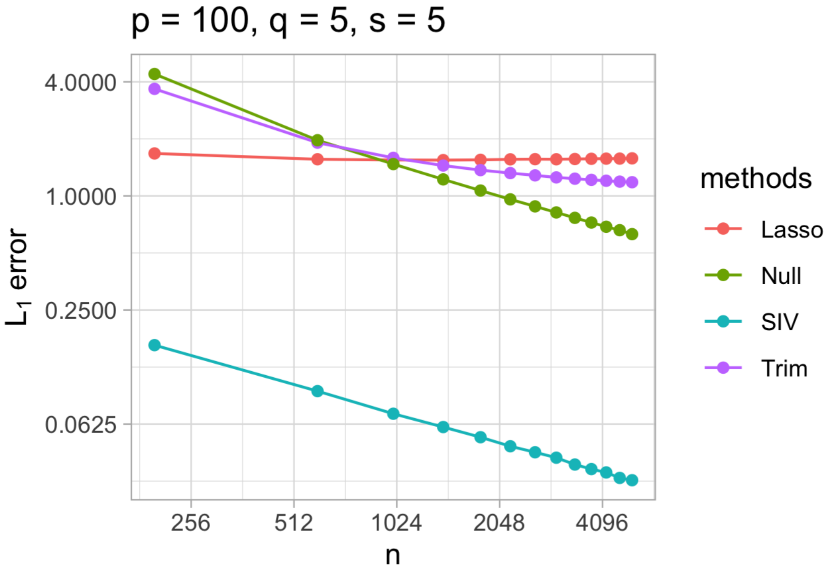
Measures of performance:

- Estimation error: ℓ_1 error

$$\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\|_1$$

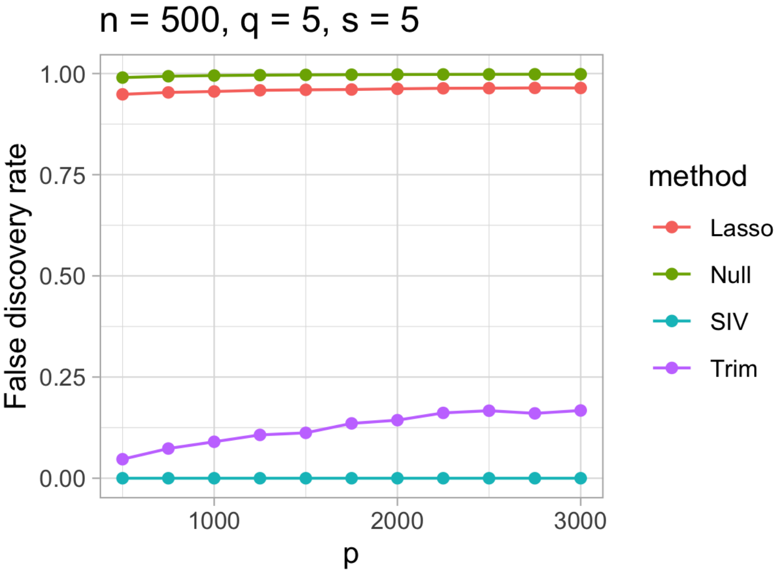
- Variable selection: false discovery rate

Low dimensional case: Estimation error



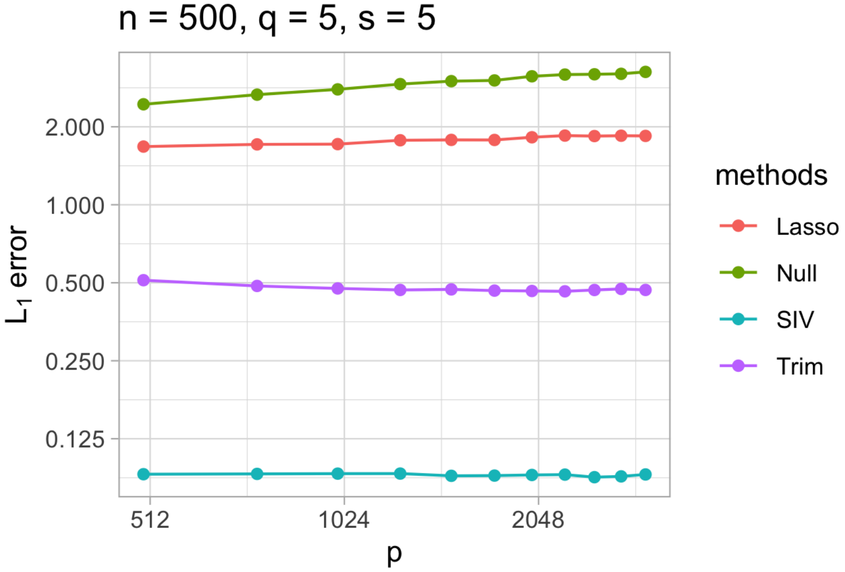
Estimation error $||\widehat{\beta} - \beta||_1$ for various methods

Low dimensional case: Variable selection



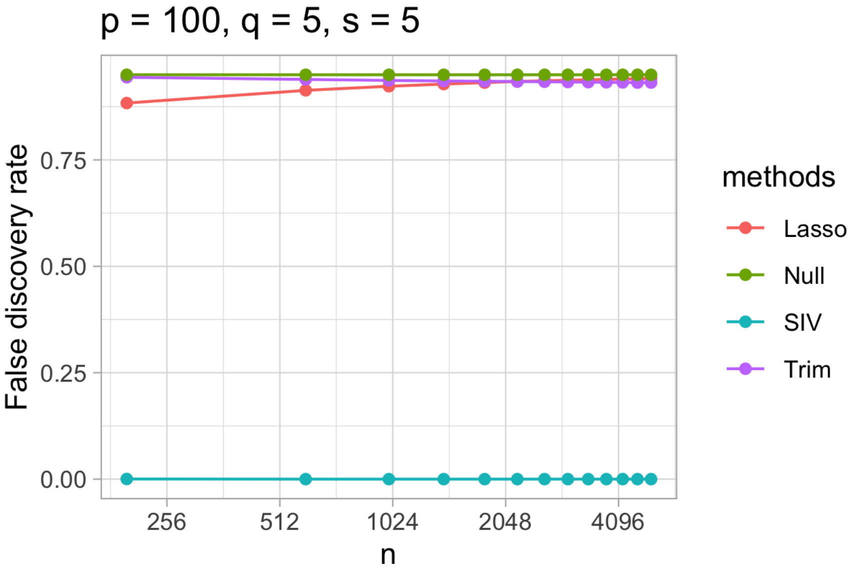
All methods correctly identify $X_1, \ldots, X_5$ as causes of Y

High dimensional case: Estimation error



Estimation error $||\widehat{\beta} - \beta||_1$ for various methods

High dimensional case: Variable selection



All methods correctly identify $X_1, \ldots, X_5$ as causes of Y

Summary

- Causal inference is possible with a high-dimensional exposure and sparse treatment effects
- Synthetic instrument is a powerful tool for causal effect estimation under linear models with multiple causes
    - Easy to implement
    - Computationally efficient
    - Outperform the state-of-art method in various settings
        - Causal effect estimation
        - Selection of true causes

References I

Ćevid, D., Bühlmann, P., and Meinshausen, N. (2020). Spectral deconfounding via perturbed sparse linear models. Journal of Machine Learning Research, 21:232.

Guo, Z., Ćevid, D., and Bühlmann, P. L. (2021). Doubly debiased lasso: High-dimensional inference under hidden confounding. The Annals of Statistics.

Hazimeh, H. and Mazumder, R. (2020). Fast best subset selection: Coordinate descent and local combinatorial optimization algorithms. Operations Research, 68(5):1517–1537.

Miao, W., Hu, W., Ogburn, E. L., and Zhou, X. (2021). Identifying effects of multiple treatments in the presence of unmeasured confounding. Journal of the American Statistical Association, (just-accepted):1–36.

Wang, Y. and Blei, D. M. (2019). The blessings of multiple causes. Journal of the American Statistical Association, 114(528):1574–1596.

Assumptions

B1 The eigenvalues of $A^{\mathrm{T}}A/p$ and D are bounded away from 0 and infinity. $||\gamma||_2 \leq \infty$.

B2 ϵ_y is independent of (X, U). Further more, assume $\epsilon_{y,i}$, $X_{i,j}$ are i.i.d sub-gaussian random variables such that $||\epsilon_{y,i}||_{\psi 2} = \sigma_y^2$, $\max_{1\leq j\leq p} ||X_{i,j}||_{\psi 2} = \sigma_x^2$. The parameters satisfies $\sigma_y^2 \leq C_4$, $C_5 \leq \sigma_x^2 \leq C_6$ for some constant $C_4, C_5, C_6 > 0$.

B3 (Restrict sparse eigenvalue condition) with probability $1 - \exp(cn)$ for some positive constant c, there exist a constant π_0 such that

$$||\widehat{X}\theta||_2 \geq \pi_0\sqrt{n}||\theta||_2, \forall ||\theta||_0 \leq 2s.$$

Miao et al. (2021)'s method

(Their δ is our γ, corresponding to the edge $U \to Y$)

1. Standard factor analysis to get $\boldsymbol{\alpha}$, and $\boldsymbol{\gamma} = \Sigma_X^{-1}\boldsymbol{\alpha}$
2. Regress Y on $\boldsymbol{X}$ to get ξ_j as the coefficient for $\boldsymbol{X}^{(j)}$
3. Since $\xi_j = \beta_j + \gamma_j \delta$, and $\|\boldsymbol{\beta}\|_0 \leq (p-q)/2$, they let

$$\widehat{\delta} = \arg\min_{\delta} \text{median}\{(\widehat{\xi}_j - \widehat{\gamma}_j \delta)^2\}$$

4. $\widehat{\boldsymbol{\beta}} = \widehat{\boldsymbol{\xi}} - \widehat{\boldsymbol{\gamma}}\widehat{\delta}$

Ćevid et al. (2020)'s method

Assume X and U are jointly Gaussian.

Assume confounding is negligible:

$$\|b\|_2^2 = O(\frac{s\sigma^2 \log p}{p}).$$

It is important that the effect of the latent variables is spread out over many predictors

Synthetic 2SLS: Stage I

Stage I: Regress $\boldsymbol{X}$ on $B_{A^\perp}\boldsymbol{X}$:

$$E(\boldsymbol{X} \mid B_{A^\perp}\boldsymbol{X}) = DB_{A^\perp}(B_{A^\perp}^{\mathrm{T}} DB_{A^\perp})^{-1} B_{A^\perp}^T \boldsymbol{X},$$

where $D = Cov(\epsilon_X)$ is a diagonal matrix.

1. Estimate $A_{q\times p}$ and $B_{A^\perp}$
2. Estimate $D_{p\times p}$
3. Plug in

Synthetic 2SLS: Stage I

Stage I: Regress $\boldsymbol{X}$ on $B_{A^\perp}\boldsymbol{X}$:

$$E(\boldsymbol{X} \mid B_{A^\perp}\boldsymbol{X}) = DB_{A^\perp}(B_{A^\perp}^{\mathrm{T}} DB_{A^\perp})^{-1} B_{A^\perp}^{T}\boldsymbol{X},$$

where $D = Cov(\epsilon_X)$ is a diagonal matrix.

1. Estimate $A_{q\times p}$ and $B_{A^\perp}$: standard factor analysis
2. Estimate $D_{p\times p}$
3. Plug in

Synthetic 2SLS: Stage I

Stage I: Regress $\boldsymbol{X}$ on $B_{A^\perp}\boldsymbol{X}$:

$$E(\boldsymbol{X} \mid B_{A^\perp}\boldsymbol{X}) = DB_{A^\perp}(B_{A^\perp}^{\mathrm{T}} DB_{A^\perp})^{-1} B_{A^\perp}^T \boldsymbol{X},$$

where $D = Cov(\epsilon_X)$ is a diagonal matrix.

1. Estimate $A_{q\times p}$ and $B_{A^\perp}$: standard factor analysis
2. Estimate $D_{p\times p}$
$$\widehat{D} = \widehat{Var}(\boldsymbol{X}) - \widehat{A}^T\widehat{A}$$
3. Plug in

Synthetic 2SLS: Stage I

Stage I: Regress $\boldsymbol{X}$ on $B_{A^\perp}\boldsymbol{X}$:

$$E(\boldsymbol{X} \mid B_{A^\perp}\boldsymbol{X}) = DB_{A^\perp}(B_{A^\perp}^{\mathrm{T}} DB_{A^\perp})^{-1}B_{A^\perp}^{T}\boldsymbol{X},$$

where $D = Cov(\epsilon_X)$ is a diagonal matrix.

1. Estimate $A_{q\times p}$ and $B_{A^\perp}$: standard factor analysis
2. Estimate $D_{p\times p}$

   $$\widehat{D} = diag(\widehat{Var}(X^{(j)}) - \widehat{A}_j^T\widehat{A}_j),\ \text{here } \widehat{A}_j \text{ is j'th row of } \widehat{A}$$

3. Plug in

Synthetic 2SLS: Stage I

Stage I: Regress $\boldsymbol{X}$ on $B_{A^\perp}\boldsymbol{X}$:

$$E(\boldsymbol{X} \mid B_{A^\perp}\boldsymbol{X}) = DB_{A^\perp}(B_{A^\perp}^{\mathrm{T}} DB_{A^\perp})^{-1} B_{A^\perp}^{T}\boldsymbol{X},$$

where $D = Cov(\epsilon_X)$ is a diagonal matrix.

1. Estimate $A_{q\times p}$ and $B_{A^\perp}$: standard factor analysis
2. Estimate $D_{p\times p}$

   $$\widehat{D} = diag(\widehat{Var}(X^{(j)}) - \widehat{A}_j^T \widehat{A}_j), \text{ here } \widehat{A}_j \text{ is j'th row of } \widehat{A}$$

3. Plug in: Need to invert $(B_{A^\perp}^{\mathrm{T}} DB_{A^\perp})_{(p-q)\times(p-q)}$

Synthetic 2SLS: Stage I

Stage I: Regress $\boldsymbol{X}$ on $B_{A^\perp}\boldsymbol{X}$:

$$E(\boldsymbol{X} \mid B_{A^\perp}\boldsymbol{X}) = DB_{A^\perp}(B_{A^\perp}^{\mathrm{T}} DB_{A^\perp})^{-1} B_{A^\perp}^T \boldsymbol{X},$$

where $D = Cov(\epsilon_X)$ is a diagonal matrix.

1. Estimate $A_{q\times p}$ and $B_{A^\perp}$: standard factor analysis
2. Estimate $D_{p\times p}$

   $$\widehat{D} = diag(\widehat{Var}(X^{(j)}) - \widehat{A}_j^T\widehat{A}_j), \text{ here } \widehat{A}_j \text{ is j'th row of } \widehat{A}$$

3. Plug in: Need to invert $(B_{A^\perp}^{\mathrm{T}} DB_{A^\perp})_{(p-q)\times(p-q)}$
   Let $\widetilde{\boldsymbol{X}} = \boldsymbol{X}\widehat{D}^{-1/2}$ so that

   $$E(\widetilde{\boldsymbol{X}} \mid B_{A^\perp}\widetilde{\boldsymbol{X}}) = \widehat{D}^{1/2}\widetilde{B}_{A^\perp}(\widetilde{B}_{A^\perp}^{\mathrm{T}}\widetilde{B}_{A^\perp})^{-1}\widetilde{B}_{A^\perp}^T\widehat{D}^{-1/2}\widetilde{\boldsymbol{X}},$$

Synthetic 2SLS: Stage I

Stage I: Regress $\boldsymbol{X}$ on $B_{A^\perp}\boldsymbol{X}$:

$$E(\boldsymbol{X} \mid B_{A^\perp}\boldsymbol{X}) = DB_{A^\perp}(B_{A^\perp}^{\mathrm{T}} DB_{A^\perp})^{-1} B_{A^\perp}^{T}\boldsymbol{X},$$

where $D = Cov(\epsilon_X)$ is a diagonal matrix.

1. Estimate $A_{q\times p}$ and $B_{A^\perp}$: standard factor analysis
2. Estimate $D_{p\times p}$

$$\widehat{D} = diag(\widehat{Var}(X^{(j)}) - \widehat{A}_j^T\widehat{A}_j), \text{ here } \widehat{A}_j \text{ is j'th row of } \widehat{A}$$

3. Plug in: Need to invert $(B_{A^\perp}^{\mathrm{T}} DB_{A^\perp})_{(p-q)\times(p-q)}$
   Let $\widetilde{\boldsymbol{X}} = \boldsymbol{X}\widehat{D}^{-1/2}$ so that

$$E(\widetilde{\boldsymbol{X}} \mid B_{A^\perp}\widetilde{\boldsymbol{X}}) = \widehat{D}^{1/2}\widetilde{B}_{A^\perp}\cancel{(\widetilde{B}_{A^\perp}^{\mathrm{T}}\widetilde{B}_{A^\perp})^{-1}}\widetilde{B}_{A^\perp}^{T}\widehat{D}^{-1/2}\widetilde{\boldsymbol{X}},$$

Synthetic 2SLS in high dimensions: Stage II

Need to estimate

$$\dot{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta} \in \mathbb{R}^p} \mathbb{E}(Y - \widehat{X}^T \boldsymbol{\beta})^2,$$

subject to $||\boldsymbol{\beta}||_0 \leq (dim(X) - dim(U))/2$

- Can be solved efficiently using the `L0Learn` package (Hazimeh and Mazumder, 2020)

Theoretical results: Error bound

Theorem

Under the same assumptions as before, and standard regularity conditions, we have

$$||\widehat{\beta} - \beta||_1 = O_p(s\sqrt{\frac{\log(p)}{n}})$$

- $s = \|\beta\|_0$
- $p = dim(X)$
- $n = sample\ size$