FACULTY OF PSYCHOLOGY
AND EDUCATIONAL SCIENCES

Data-driven Covariate Selection for Confounding Adjustment by Focusing on the Stability of the Effect Estimator

Wen Wei Loh, and Dongning Ren

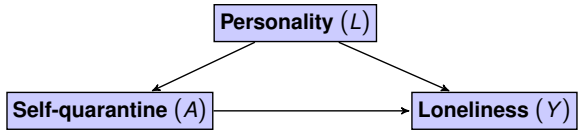
SEM conference | Tilburg University, The Netherlands | March 11, 2022
Funding Suppport: Ghent University BOF.PDO.2020.0045.01

GHENT
UNIVERSITY

Overview

1. Why select confounders (or covariates) for causal inference
2. Proposed strategy focusing on effect stability
3. Illustration with applied example

Motivating example

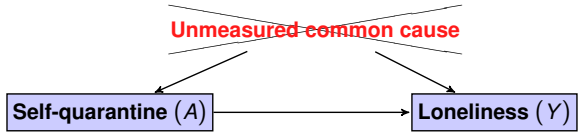


Consider an observational study (conducted May 2020) with:

- A: Self-isolation or quarantine since start of the COVID-19 outbreak;
- Y: Loneliness;
- L: Demographic and pre-pandemic information, and personality scores.

Unbiased effect estimation requires statistically *adjusting* (or controlling) for baseline common causes of A and Y.

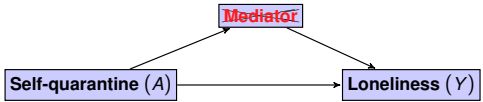
Strong ignorability

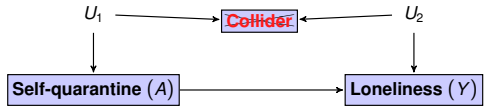


Assume that *strong ignorability* [Rosenbaum and Rubin, 1983], or *unconfoundedness*, holds; i.e., the recorded covariates are sufficient to eliminate all confounding and there are no unmeasured common causes.

Strong ignorability is guaranteed in *randomized experiments*.

Strong ignorability





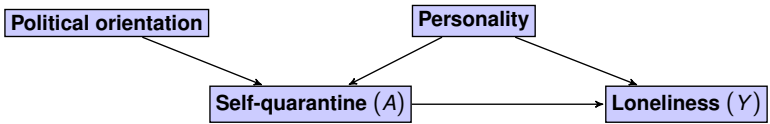
No variables:

- causally affected by treatment (mediators), and
- no (descendants of) colliders along the treatment-outcome causal path,

are included among the available covariates.

Such screening must be based on subject matter theory; readily with *causal diagrams*.

Non-confounding causes



What about adjusting for other non-confounding causes?

- Covariates that predict treatment only:
  - decrease precision of the estimator $\Rightarrow$ unstable estimates with finite sample bias;
  - do not reduce confounding bias [Brookhart et al., 2006, Vansteelandt et al., 2012].
- E.g., political orientation may affect A, but not Y directly.

Non-confounding causes

- Covariates that predict outcome only:
  - improve precision of the estimator [Brookhart et al., 2006, Little et al., 2000, Shortreed and Ertefaie, 2017].
- E.g., Pre-pandemic levels of loneliness and boredom may be unrelated to choice to self-quarantine
- But the risk of model misspecification biases increases with more covariates.

$\Rightarrow$ Select a (minimal) subset sufficient for confounding adjustment.

Change-in-estimate approach for covariate selection

In this talk: focus on the change-in-estimate approach [Mickey and Greenland, 1989].

Change-in-estimate approach for covariate selection

In this talk: focus on the change-in-estimate approach [Mickey and Greenland, 1989].

Rationale: Suppose that a subset of covariates sufficient for confounding adjustment has been selected.

Change-in-estimate approach for covariate selection

In this talk: focus on the change-in-estimate approach [Mickey and Greenland, 1989].

Rationale: Suppose that a subset of covariates sufficient for confounding adjustment has been selected.

Further adjustment for covariates associated with either treatment or outcome, *but not both*, should not systematically change the effect estimator.

Example from a randomized experiment

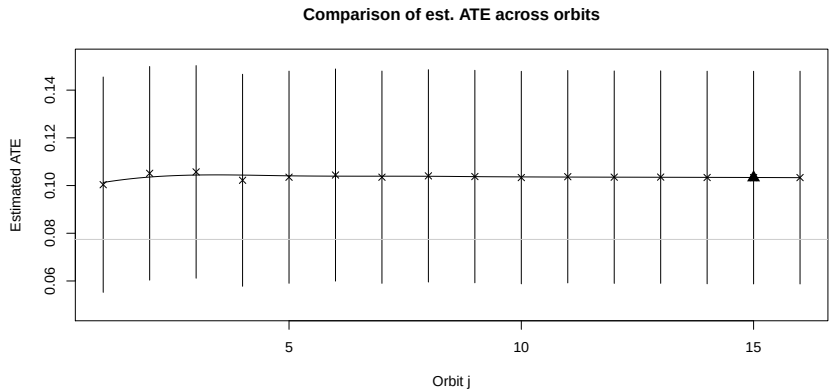


Figure 1: Trajectory of estimated average treatment effects ('est. ATE') as the number of covariates ('orbits') adjusted for changes, for a randomized experiment with no confounding. Each vertical bar represents the 95% CI for the ATE, adjusting for the covariates in that orbit. The solid black curve is a local cubic polynomial smoother.

Outline

1 WHY SELECT CONFOUNDERS (OR COVARIATES) FOR CAUSAL INFERENCE

2 PROPOSED STRATEGY FOCUSING ON EFFECT STABILITY

3 ILLUSTRATION WITH APPLIED EXAMPLE

Change-in-estimate procedure

1. Construct different (nested) covariate subsets.

Change-in-estimate procedure

1. Construct different (nested) covariate subsets.
2. Evaluate the treatment effect estimator after adjusting for each subset.

Change-in-estimate procedure

1. Construct different (nested) covariate subsets.
2. Evaluate the treatment effect estimator after adjusting for each subset.
3. Select the smallest subset whose effect estimate (or its approximate mean-squared error) remains “unchanged,” even after adjusting for additional covariates [Greenland et al., 2016, Vansteelandt et al., 2012].

Change-in-estimate procedure

1. Construct different (nested) covariate subsets.
2. Evaluate the treatment effect estimator after adjusting for each subset.
3. Select the smallest subset whose effect estimate (or its approximate mean-squared error) remains "unchanged," even after adjusting for additional covariates [Greenland et al., 2016, Vansteelandt et al., 2012].

Offers insight into stability of the effect estimator to the different covariates adjusted for.

Proposal

Step 1. How to construct nested covariates subsets?

$\Rightarrow$ Prioritize covariates for confounding adjustment using a specified criterion.

Proposal

Step 1. How to construct nested covariates subsets?

$\Rightarrow$ Prioritize covariates for confounding adjustment using a specified criterion.

Step 2. How to determine stability?

$\Rightarrow$ Directly evaluate the trajectory of the effect estimator.

Proposal

Step 1. How to construct nested covariates subsets?

$\Rightarrow$ Prioritize covariates for confounding adjustment using a specified criterion.

Step 2. How to determine stability?

$\Rightarrow$ Directly evaluate the trajectory of the effect estimator.

Focus on Step 1 in this talk.

Step 1. Covariate prioritization

Start with the empty set.

Add covariates one-at-a-time (i.e., *stepwise forward selection*).

At each time, add the candidate covariate most strongly associated (conditionally) with treatment and outcome.

Step 1. Covariate prioritization

E.g., for a candidate covariate:

1. Fit a model for treatment given (i) the candidate, and (ii) covariates already in the adjustment set.
2. Fit a model for outcome given (i) the candidate, (ii) covariates already in the adjustment set, and (iii) treatment.
3. Determine the minimum of the p-values for the coefficients of the candidate in both models.
4. Add the candidate with the smallest minimum p-value to the adjustment set.

Step 1. Covariate prioritization

- Repeating the above steps (until all covariates have been added) returns a hierarchical ordering of the covariates.
- First covariate added has the highest priority for confounding adjustment (based on the specified criterion), second has the next highest priority, and so on.
- The ordering induces a series of nested covariate subsets.
- Specified criterion inspired by *double selection* [Belloni et al., 2014] principles: consider the (partial) associations with treatment and with outcome.

Step 1. Covariate prioritization

- No pre-determined (significance-based) threshold is imposed to rule out any covariates from adjustment.
- Other orderings are possible simply by using different criteria. (*Suggestions welcome!*)

Outline

1 WHY SELECT CONFOUNDERS (OR COVARIATES) FOR CAUSAL INFERENCE

2 PROPOSED STRATEGY FOCUSING ON EFFECT STABILITY

3 ILLUSTRATION WITH APPLIED EXAMPLE

Illustration

Recall the observational study with:

- A: Time spent in self-isolation or quarantine since start of the COVID-19 outbreak;
- Y: Average (score) of the three items measuring recent loneliness;
- 34 covariates including demographic and pre-pandemic information, and personality scores;
- $N = 404$ participants.

Constructed the nested covariate subsets, then calculated OLS estimators of the treatment effect in an outcome model with the covariates in each subset.

Illustration

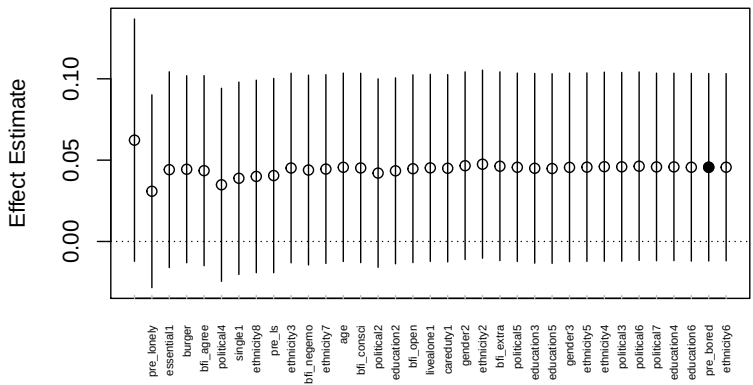


Figure 2: The OLS (regression) estimate adjusting for the covariates in each orbit, for the Loneliness data. The vertical lines indicate 95% CIs. The covariates were ordered using the double selection criterion. The most stable orbit is indicated by a filled circle.

Remarks

We exploit the causal knowledge that stability is attained once all confounders have been selected.

1. Prioritize covariates for adjustment using double selection principles.
2. Select the smallest (most parsimonious) subset that yields a stable effect estimator using a change-in-estimate approach.

Stability across different covariate subsets can be assessed visually or numerically.

Remarks

Further details in the preprint (https://psyarxiv.com/zkdqa/).

- Accounting for the sampling variability of the estimators when evaluating stability
- Simulation studies comparing the proposal with routine variable selection methods, especially recent developments for SEM using regularization (`RegSEM`) [Jacobucci et al., 2016] or penalized likelihood (`lslx`) [Huang et al., 2017]
- Illustrations using two different publicly available datasets.

R code on GitHub:
https://github.com/wwloh/covariate-selection-effect-stability

Thank you!

References

A. Belloni, V. Chernozhukov, and C. Hansen. Inference on treatment effects after selection among high-dimensional controls. *The Review of Economic Studies*, 81(2):608–650, 2014. doi: 10.1093/restud/rdt044.

M. A. Brookhart, S. Schneeweiss, K. J. Rothman, R. J. Glynn, J. Avorn, and T. Stürmer. Variable selection for propensity score models. *American Journal of Epidemiology*, 163(12): 1149–1156, 2006. doi: 10.1093/aje/kwj149.

S. Greenland, R. Daniel, and N. Pearce. Outcome modelling strategies in epidemiology: traditional methods and basic alternatives. *International Journal of Epidemiology*, 45(2): 565–575, 2016. doi: 10.1093/ije/dyw040.

D. C. Hoaglin. Misunderstandings about Q and 'Cochran's Q test' in meta-analysis. *Statistics in Medicine*, 35(4):485–495, 2016. doi: 10.1002/sim.6632.

P.-H. Huang, H. Chen, and L.-J. Weng. A penalized likelihood method for structural equation modeling. *Psychometrika*, 82(2):329–354, 2017. doi: 10.1007/s11336-017-9566-9. URL https://doi.org/10.1007/s11336-017-9566-9.

R. Jacobucci, K. J. Grimm, and J. J. McArdle. Regularized structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, 23(4):555–566, 2016. doi: 10.1080/10705511.2016.1154793. URL https://doi.org/10.1080/10705511.2016.1154793.

J. Kim and A. Z. Zambom. *SignifReg: Consistent Significance Controlled Variable Selection in Linear Regression*, 2020. URL https://CRAN.R-project.org/package=SignifReg. R package version 3.0.

R. J. Little, H. An, J. Johanns, and B. Giordani. A comparison of subset selection and analysis of covariance for the adjustment of confounders. *Psychological Methods*, 5(4): 459–476, 2000. doi: 10.1037/1082-989X.5.4.459.

R. M. Mickey and S. Greenland. The impact of confounder selection criteria on effect estimation. *American journal of epidemiology*, 129(1):125–137, 1989. doi: 10.1093/oxfordjournals.aje.a115101.

P. R. Rosenbaum and D. B. Rubin. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55, 1983. doi: 10.1093/biomet/70.1.41.

S. M. Shortreed and A. Ertefaie. Outcome-adaptive lasso: Variable selection for causal inference. *Biometrics*, 73(4):1111–1122, 2017. doi: 10.1111/biom.12679.

R. Tibshirani. Regression shrinkage and selection via the LASSO: a retrospective. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(3):273–282, 2011. doi: 10.1111/j.1467-9868.2011.00771.x.

S. Vansteelandt, M. Bekaert, and G. Claeskens. On model selection and model misspecification in causal inference. *Statistical Methods in Medical Research*, 21(1):7–30, 2012.

Step 2. Assessing stability

Calculate the average treatment effect (ATE) estimator for each (nested) covariate subset.

- ATE: $\psi = \mathrm{E}(Y^1) - \mathrm{E}(Y^0)$; Y^a: *potential outcome* under treatment $A = a$.
- $\widehat{\psi}_j$: marginal effect estimator adjusting for the j-th covariate subset.

Select the smallest subset that yields a stable estimator of ψ, relative to a "benchmark" estimator.

Step 2. Assessing stability

E.g., let $\widehat{\psi}_J$ that adjusts for all J available covariates be the "benchmark" estimator.

Define the standardized difference as:

$$\frac{\widehat{\psi}_j - \widehat{\psi}_J}{\sqrt{\text{var}\left(\widehat{\psi}_j - \widehat{\psi}_J\right)}}, \quad j = 1, \ldots, J-1; \tag{1}$$

where $\text{var}(X)$ denotes the asymptotic variance of X.

Seek to select the smallest subset j that yields the most "stable" value of (1).

Step 2. Assessing stability

- Choice of benchmark based on the assumption that strong ignorability holds; $\widehat{\psi}_J$ is (asymptotically) unbiased for ψ, and the difference $\widehat{\psi}_j - \widehat{\psi}_J$ can be viewed as an approximate bias.

- Using the difference with a common benchmark may account for correlation between estimators in different subsets.

Step 3. Numerical diagnosis of stability

Alternative to visual inspection: numerical diagnostic of relative stability of (1), while accounting for its variability.

- Use an inverse variance weighted average of the differences $\widehat{\psi}_j - \widehat{\psi}_J$ within a (moving) window of consecutive nested subsets.

- Adopts the same form as “Cochran’s Q statistic” [Hoaglin, 2016] from the meta-analysis literature for assessing heterogeneity of effect-size estimates from separate studies.

Step 3. Numerical diagnosis of stability

For simplicity, we will use (symmetric) windows of width five centered around each subset $j = 3, \ldots, J-2$. The diagnostic for the j-th subset is therefore defined as:

$$Q_j = \sum_{k=j-2}^{j+2} w_k \{(\widehat{\psi}_k - \widehat{\psi}_J) - \overline{\widehat{\psi}_j}\}^2, \tag{2}$$

where the weights w_k and weighted average $\overline{\widehat{\psi}_j}$ are respectively defined as:

$$w_k = \left\{ \text{var}\left(\widehat{\psi}_k - \widehat{\psi}_J\right) \right\}^{-1}, \quad \overline{\widehat{\psi}_j} = \left(\sum_{k=j-2}^{j+2} w_k \right)^{-1} \sum_{k=j-2}^{j+2} w_k (\widehat{\psi}_k - \widehat{\psi}_J).$$

Step 3. Numerical diagnosis of stability

The smallest orbit with the most stable value of (1) therefore minimizes the Q statistic; i.e.,

$$\min_{j=3,\ldots,J-2} Q_j. \tag{3}$$

The weighted average as defined in (2) adopts the same form as ‘Cochran’s Q statistic’[Hoaglin, 2016] from the meta-analysis literature for assessing heterogeneity of effect-size estimates from separate studies.

Step 3. Numerical diagnosis of stability

Figure 3: Standardized difference (“Std. Diff.”), between the treatment effect estimator from each subset and from the largest subset, for the Loneliness data. The most stable orbit minimizing Cochran’s Q is indicated by a filled circle.

Simulation study: Data-generating process

We partitioned the $J = 20$ covariates into four subsets (based on their indices):

- $\mathcal{S}_1 = \{1, 2\}$: confounders simultaneously affected treatment and outcome;
- $\mathcal{S}_2 = \{3, 4\}$: covariates affected outcome only;
- $\mathcal{S}_3 = \{5, 6\}$: instruments associated with treatment only;
- $\mathcal{S}_4 = \{7, \ldots, J\}$: unassociated with either treatment or outcome.

Simulation study: Data-generating process

Datasets with $N = 400$ were generated under the null $H_0 : Y^1 = Y^0$. For each individual i:

1. Draw the covariates as $L_{is} \sim \mathcal{N}(0, 1)$, $s = 1, \ldots, J$; denote all covariates by L_i.
2. Determine the underlying treatment as $A_i^* = \sum_{s=1}^{p} \gamma_s L_{is}$.
   Set $\gamma_s = 1.0$ if $s \in \mathcal{S}_1$ (a confounder), or $\gamma_s = 0.8$ if $s \in \mathcal{S}_3$ (an instrument), or 0 otherwise.

Simulation study: Data-generating process

3. Randomly draw the observed treatment as $A_i \sim \mathcal{N}\left(A_i^*, b_i^2\right)$, where $b_i = \sqrt{\frac{|A_i^*|}{\max_i |A_i^*|}} \in (0, 1]$. Standardize to have zero mean and unit variance.
4. Determine the underlying outcome as $Y_i^* = \sum_{s=1}^{p} \beta_s L_{is}$, where $\beta_s = 0.8$ if $s \in \mathcal{S}_1 \cup \mathcal{S}_2$ (a confounder or an outcome-only predictor), or 0 otherwise.
5. Randomly draw the observed outcome as $Y_i \sim \mathcal{N}\left(Y_i^*, \sigma^2\right)$, where $\sigma = \max_i |Y_i^*|$.

Other covariate selection methods

- `LASSO`: Regularized, or penalized, linear regression using the least absolute shrinkage and selection operator (LASSO; Tibshirani, 2011); or

Other covariate selection methods

- `LASSO`: Regularized, or penalized, linear regression using the least absolute shrinkage and selection operator (LASSO; Tibshirani, 2011); or

- `RegSEM`: Regularized SEM [Jacobucci et al., 2016]); or

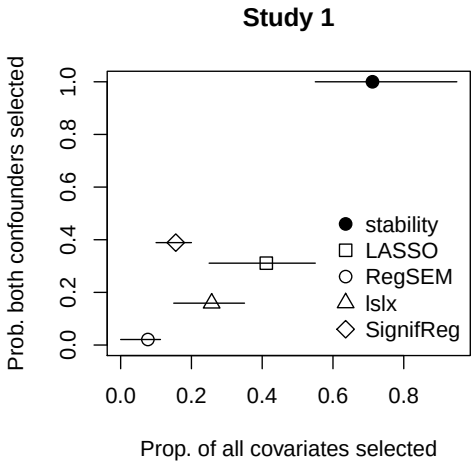
Other covariate selection methods

- `LASSO`: Regularized, or penalized, linear regression using the least absolute shrinkage and selection operator (LASSO; Tibshirani, 2011); or
- `RegSEM`: Regularized SEM [Jacobucci et al., 2016]); or
- `lslx`: Semi-confirmatory SEM ("SC-SEM") via penalized likelihood [Huang et al., 2017]; or

Other covariate selection methods

- `LASSO`: Regularized, or penalized, linear regression using the least absolute shrinkage and selection operator (LASSO; Tibshirani, 2011); or
- `RegSEM`: Regularized SEM [Jacobucci et al., 2016]); or
- `lslx`: Semi-confirmatory SEM ("SC-SEM") via penalized likelihood [Huang et al., 2017]; or
- `SignifReg`: forward selection for linear regression models [Kim and Zambom, 2020].

Results: selection



Results: estimation

Table 1: Empirical summaries of estimates of the treatment effect and its standard error ("SE") either following the use of each covariate selection method (S1 – S5), or directly applying each estimation method (M1 – M5). The true value of the treatment effect was zero.

	Method	Double Selection	Bias	ESE	RMSE	ASE	Type I
S1	Stability	TRUE	0.04	1.26	1.26	1.17	0.07
S2	LASSO	FALSE	0.78	0.79	1.11	0.50	0.55
S3	RegSEM	FALSE	1.05	0.45	1.14	0.34	0.84
S4	SC-SEM	FALSE	0.94	0.67	1.15	0.42	0.67
S5	SignifReg	FALSE	0.49	0.77	0.91	0.40	0.38