Supplementary Material for manuscript: Valid causal inference with unobserved confounding in high-dimensional settings

Niloofar Moosavi, Tetiana Gorbach, Xavier de Luna

Department of Statistics, USBE, Umeå University, Umeå, Sweden

email: name.surname@umu.se

We consider three low-dimensional settings. The distribution of covariates and error terms, number of replication and other details not mentioned here can be found in the paper. The motivation for considering low-dimensional scenarios is that they are, needless to say, more convenient to work with and interpret, and they are, nevertheless, no exception to the issues that can arise for post-selection estimators [see, e.g., Leeb and Pötscher, 2005, Moosavi et al., 2023].

Scenario one:

$$Y(1) = 2 + 0.5X_1 + \alpha X_2 - \rho\lambda(X_2) + \xi,$$

$$T^* = X_2 + \eta.$$

Web Table 1. The value of α in Scenarios one and two and (α, α') in Scenario three for each considered sample size n

n	1000	5000	20000
$Scenario1$	-0.07	-0.035	-0.015
$Scenario2$	0.12	0.075	0.06
$Scenario3$	$(-0.07, 0.12)$	$(-0.035, 0.075)$	$(-0.015, 0.06)$

Scenario two:

$$Y(1) = 2 - 0.5X_2 - \rho\lambda(X_1 + \alpha X_2) + \xi,$$

$$T^* = X_1 + \alpha X_2 + \eta.$$

Scenario three:

$$Y(1) = 2 - 0.5X_1 + \alpha X_2 - \rho\lambda(X_1 + \alpha' X_2) + \xi,$$

$$T^* = X_1 + \alpha' X_2 + \eta.$$

In Scenario one, covariate X_2 has a weaker association with the outcome and, therefore, is sometimes omitted in the variable selection step. The same occurs with Scenario two, where covariate X_2 has a weaker relationship with the treatment assignment, and with Scenario three, where X_2 is weakly associated with both outcome and treatment. If α was constant with increasing sample size, X_2 would eventually be selected by lasso for large sample sizes. In order to consider the cases where lasso can omit X_2 even for bigger sample sizes, we let the value of α depend on the sample size (see Web Table 1).

Web Table 2. Scenario one: Empirical coverages of 95% confidence intervals for τ

Estimator	$\hat{\tau}_{AIPW}^{refit} - b^*$						$\hat{\tau}_{AIPW}^{refit} - \hat{b}^{refit}$						$\hat{\tau}_{AIPW}^{refit} - \hat{b}_c^{refit}$					
$n \backslash \rho$	0.8	0.7	0.6	0.5	0.4	0.3	0.8	0.7	0.6	0.5	0.4	0.3	0.8	0.7	0.6	0.5	0.4	0.3
1000	0.94	0.95	0.94	0.95	0.94	0.94	0.37	0.71	0.88	0.92	0.92	0.95	0.88	0.91	0.93	0.93	0.92	0.93
5000	0.97	0.93	0.95	0.97	0.95	0.96	0.02	0.20	0.58	0.83	0.92	0.95	0.90	0.93	0.92	0.95	0.94	0.95
20000	0.94	0.97	0.96	0.94	0.94	0.94	0.00	0.01	0.11	0.54	0.84	0.94	0.90	0.90	0.95	0.95	0.97	0.95

Web Table 3. Scenario two: Empirical coverages of 95% confidence intervals for τ

Estimator	$\hat{\tau}_{AIPW}^{refit} - b^*$						$\hat{\tau}_{AIPW}^{refit} - \hat{b}^{refit}$						$\hat{\tau}_{AIPW}^{refit} - \hat{b}_c^{refit}$					
$n \setminus \rho$	0.8	0.7	0.6	0.5	0.4	0.3	0.8	0.7	0.6	0.5	0.4	0.3	0.8	0.7	0.6	0.5	0.4	0.3
1000	0.95	0.95	0.96	0.95	0.95	0.94	0.38	0.65	0.84	0.88	0.93	0.94	0.89	0.90	0.91	0.94	0.95	0.94
5000	0.96	0.95	0.97	0.94	0.95	0.94	0.03	0.18	0.52	0.80	0.93	0.93	0.87	0.92	0.94	0.93	0.97	0.95
20000	0.93	0.97	0.96	0.97	0.94	0.96	0.00	0.02	0.11	0.53	0.83	0.94	0.92	0.93	0.95	0.93	0.93	0.96

Web Table 4. Scenario three: Empirical coverages of 95% confidence intervals for τ

Estimator	$\hat{\tau}_{AIPW}^{refit} - b^*$						$\hat{\tau}_{AIPW}^{refit} - \hat{b}^{refit}$						$\hat{\tau}_{AIPW}^{refit} - \hat{b}_c^{refit}$					
$n \setminus \rho$	0.8	0.7	0.6	0.5	0.4	0.3	0.8	0.7	0.6	0.5	0.4	0.3	0.8	0.7	0.6	0.5	0.4	0.3
1000	0.94	0.95	0.94	0.93	0.94	0.96	0.41	0.65	0.86	0.90	0.95	0.95	0.91	0.91	0.90	0.94	0.96	0.93
5000	0.97	0.95	0.95	0.94	0.94	0.94	0.03	0.18	0.55	0.81	0.93	0.93	0.89	0.91	0.95	0.95	0.97	0.94
20000	0.96	0.97	0.95	0.95	0.93	0.97	0.00	0.02	0.12	0.55	0.83	0.95	0.94	0.93	0.95	0.94	0.93	0.94

References

H. Leeb and B. M. Pötscher. Model selection and inference: Facts and fiction. *Econometric Theory*, 21(1):21–59, 2005.

N. Moosavi, J. Häggström, and X. de Luna. The Costs and Benefits of Uniformly Valid Causal Inference with High-Dimensional Nuisance Parameters. *Statistical Science*, 38(1):1 – 12, 2023. doi: 10.1214/21-STS843. URL `https://doi.org/10.1214/21-STS843`.