MSE-optimal adjustment sets in linear Gaussian causal models with finite sample size

Nadja Rutsch[1,2] **Sara Magliacane**[3] **Stéphanie L. van der Pas**[1,2]

[1]Amsterdam UMC location Vrije Universiteit Amsterdam, Epidemiology and Data Science, De Boelelaan 1117, Amsterdam, Netherlands
[2]Amsterdam Public Health, Methodology, Amsterdam, Netherlands
[3]AMLab, Informatics Institute, University of Amsterdam Lab 42, Science Park 900, 1012 WX Amsterdam, Netherlands

Abstract

Covariate selection for causal inference based on the causal graph commonly aims for unbiasedness and asymptotic efficiency of the causal effect estimator. When the sample size is finite, these approaches can lead to results that are suboptimal in terms of the Mean Squared Error (MSE). We aim to find the adjustment set that is optimal in terms of MSE, taking into account the joint distribution of the causal variables and the sample size. We present examples where the MSE-optimal adjustment set differs from the optimal adjustment set, depending on the sample size. To find the MSE-optimal adjustment set, we introduce a sample size criterion that compares two adjustment sets in linear Gaussian models. We develop graphical criteria to reduce the search space for this adjustment set based on the causal graph. In preliminary experiments, we show that the estimated MSE-optimal adjustment set can outperform the optimal adjustment set in finite sample size settings, and performs competitively in larger sample size settings.

1 INTRODUCTION

Causal inference from observational data is an important but challenging task. Various methods have been proposed for it, including propensity score methods [Rosenbaum and Rubin, 1983], matching [Stuart, 2010], instrumental variables [Angrist et al., 1996], regression discontinuity design [Imbens and Lemieux, 2008], and double machine learning [Chernozhukov et al., 2018]. One of the most straightforward and popular approaches is covariate adjustment. For this, one needs to select a set of covariates to adjust for. Naturally, this raises the question: which covariates should we select for the best causal effect estimate?

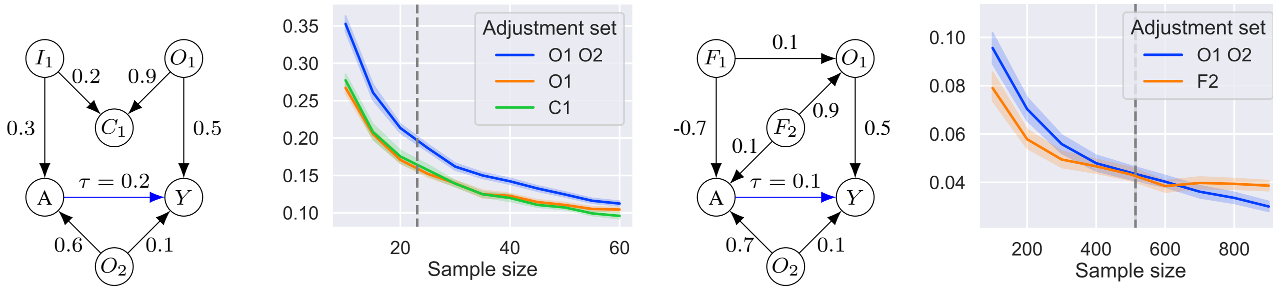
For the pre-selection of covariates, a graphical representation of the causal relations is often used [Pearl, 1993, Shpitser et al., 2010, Rotnitzky and Smucler, 2019, Henckel et al., 2022]. So far, methods based on causal graphs focus on *valid* adjustment sets. A set of covariates K is a valid adjustment set if a covariate adjustment estimator $\hat{\tau}_K$ returns an unbiased estimate of the true causal effect under correct model specification. However, in finite sample size settings, the variance may dominate the bias, such that an invalid adjustment set may be more suitable for estimation. Figure 1 shows two examples where invalid adjustment sets outperform the unbiased optimal adjustment set $\mathbf{O}$ [Henckel et al., 2022, Rotnitzky and Smucler, 2019] in terms of MSE. In the following, we describe how to find the adjustment set that is optimal in terms of MSE.

2 FINDING MSE-OPTIMAL SETS

The MSE of an estimator can be decomposed into its squared bias and variance. Whether an adjustment set is valid and hence unbiased, can be determined from the causal graph alone, e.g. with the back-door criterion [Pearl, 1993], which is sufficient for unbiasedness, or with a necessary and sufficient criterion developed by Shpitser et al. [2010] and Perković et al. [2018]. Previously, optimality criteria have focused on valid adjustment sets. We consider a different notion of optimality.

2.1 DIFFERENT OPTIMALITY CRITERIA

The optimal adjustment set $\mathbf{O}$ is the adjustment set with minimal asymptotic variance among all valid adjustment sets. It was first defined for ordinary least squares (OLS) estimation in linear causal graphical models [Henckel et al., 2022] and later extended to non-parametric models [Rotnitzky and Smucler, 2019]. The optimal adjustment set $\mathbf{O}$ consists of the parents of mediators that are not themselves mediators or the treatment, where mediators are defined to also include the outcome [Guo et al., 2023].



(a) $\mathcal{M}_1$ (b) RMSE of $\hat{\tau}$ in $\mathcal{M}_1$ (c) $\mathcal{M}_2$ (d) RMSE of $\hat{\tau}$ in $\mathcal{M}_2$

Figure 1: Two toy examples of causal models $\mathcal{M}_1$ and $\mathcal{M}_2$ and the Root-Mean Squared Error (RMSE) of the OLS estimator $\hat{\tau}$ of the causal effect τ of A on Y in $\mathcal{M}_1$ and $\mathcal{M}_2$, using different adjustment sets (10000 random seeds per set and sample size). The variables in $\mathcal{M}_1$ and $\mathcal{M}_2$ are linear Gaussian distributed with a variance of 1 and a mean of 0. Depending on the sample size, a different adjustment set than $\mathbf{O} = \{O_1, O_2\}$ gives the lowest RMSE. The dashed line shows the sample size for which C_1 outperforms O_1 and $\{O_1, O_2\}$ outperforms F_2 respectively, as calculated with our sample size criterion.

While previous work is restricted to valid adjustment sets, we consider all possible adjustment sets, including invalid ones. We aim to find the adjustment set K that gives the most accurate average treatment effect estimator $\hat{\tau}_K$ in terms of MSE for a given causal model $\mathcal{M}$ and sample size n, which we call the *MSE-optimal adjustment set* $\mathbf{O}_n(\mathcal{M}, \hat{\tau})$. We focus on the setting where $\mathcal{M}$ is linear Gaussian and $\hat{\tau}$ is the OLS estimator. In relation to the optimal adjustment set $\mathbf{O}$, we conjecture that the MSE-optimal adjustment set $\mathbf{O}_n(\mathcal{M}, \hat{\tau})$ converges to the optimal adjustment set $\mathbf{O}$ as the sample size n approaches infinity, given that the graph corresponding to the model $\mathcal{M}$ is faithful.

2.2 SAMPLE SIZE CRITERION

As demonstrated in Figure 1, the adjustment set that yields the lowest MSE can depend on the sample size. We present a criterion to compare two adjustment sets K and L for treatment effect estimation given a linear Gaussian model $\mathcal{M}$ and sample size n, based on their asymptotic variances $\nu(\cdot)$ [Henckel et al., 2022], squared biases $B^2(\cdot)$ and set sizes $|\cdot|$:

$$n < \frac{\nu(L) - \left(\frac{n-|L|-3}{n-|K|-3}\right)\nu(K)}{B^2(\hat{\tau}_K) - B^2(\hat{\tau}_L)} + |L| + 3. \tag{1}$$

If the sample size criterion holds when $B^2(\hat{\tau}_K)$ is larger than $B^2(\hat{\tau}_L)$, the expected MSE of $\hat{\tau}_K$ is lower than of $\hat{\tau}_L$.

2.3 GRAPHICAL CRITERIA

For linear Gaussian causal models and the OLS estimator, we propose two conjectures about variables or variable combinations that can be excluded from $\mathbf{O}_n(\mathcal{M}, \hat{\tau})$, solely based on the graph $\mathcal{G}$. Our first graphical criterion concerns the exclusion of single variables. For example, in the graph

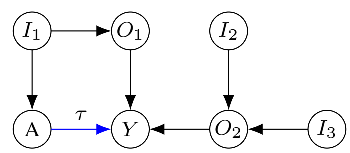


Figure 2: $\mathcal{G}_3$

$\mathcal{G}_3$ shown in Figure 2, the variables I_1, I_2, and I_3 can never be in $\mathbf{O}_n(\mathcal{M}, \hat{\tau})$, as it would always be better to adjust for O_1 or O_2 instead. Our second criterion concerns the exclusion of variable combinations. If there are two variables V_i and V_j, such that one is d-separated from the outcome Y, given the other variable and the treatment A, then at most one of them can be in $\mathbf{O}_n(\mathcal{M}, \hat{\tau})$. E.g. in the graph of $\mathcal{M}_2$ from Figure 1, F_1 is d-separated from Y given A and O_1, such that F_1 and O_1 can never both be in $\mathbf{O}_n(\mathcal{M}, \hat{\tau})$.

3 EXPERIMENTS

We estimate the expected MSE of the OLS treatment effect estimator $\hat{\tau}_K$ for each adjustment set K to find the estimated MSE-optimal adjustment set $\hat{\mathbf{O}}_n(\mathcal{M}, \hat{\tau})$. Table 1 shows that $\hat{\mathbf{O}}_n(\mathcal{M}, \hat{\tau})$ outperforms $\mathbf{O}$ in small sample sizes, and performs competitively in larger sample sizes.

Table 1: Comparison of MSE for $\mathbf{O}$ and $\hat{\mathbf{O}}_n(\mathcal{M}, \hat{\tau})$ with $\mathcal{M}_2$ from Figure 1, 10000 random seeds.

Sample Size	$\mathbf{O}$ (Mean $\pm$ SD)	$\hat{\mathbf{O}}_n$ (Mean $\pm$ SD)
10	0.2789 (0.2789)	**0.2440 (0.5172)**
50	0.0317 (0.0317)	**0.0293 (0.0441)**
100	0.0154 (0.0154)	**0.0146 (0.0208)**
500	**0.0029 (0.0029)**	0.0032 (0.0043)
1000	**0.0015 (0.0015)**	0.0016 (0.0022)

Acknowledgements

N. Rutsch and S. L. van der Pas have received funding from the European Research Council (ERC) under the European Union's Horizon Europe program under Grant agreement No. 101074082. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

References

Joshua D Angrist, Guido W Imbens, and Donald B Rubin. Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, 91 (434):444–455, 1996.

Victor Chernozhukov, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey, and James Robins. Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal*, 21(1):C1–C68, 2018.

F. Richard Guo, Emilija Perković, and Andrea Rotnitzky. Variable elimination, graph reduction and the efficient g-formula. *Biometrika*, 110(3):739–761, 2023.

Leonard Henckel, Emilija Perković, and Marloes H. Maathuis. Graphical Criteria for Efficient Total Effect Estimation Via Adjustment in Causal Linear Models. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 84(2):579–599, April 2022. ISSN 1369-7412. doi: 10.1111/rssb.12451.

Guido W Imbens and Thomas Lemieux. Regression discontinuity designs: A guide to practice. *Journal of Econometrics*, 142(2):615–635, 2008.

Judea Pearl. Comment: Graphical models, causality and intervention. *Statistical Science*, 8(3):266–269, 1993. ISSN 08834237. URL `http://www.jstor.org/stable/2245965`.

Emilija Perković, Johannes Textor, Markus Kalisch, Marloes H Maathuis, et al. Complete graphical characterization and construction of adjustment sets in markov equivalence classes of ancestral graphs. *Journal of Machine Learning Research*, 18(220):1–62, 2018.

Paul R Rosenbaum and Donald B Rubin. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55, 1983.

Andrea Rotnitzky and Ezequiel Smucler. Efficient adjustment sets for population average treatment effect estimation in non-parametric causal graphical models, December 2019.

Ilya Shpitser, Tyler VanderWeele, and James M. Robins. On the validity of covariate adjustment for estimating causal effects. In *Proceedings of the Twenty-Sixth Conference on Uncertainty in Artificial Intelligence*, UAI'10, page 527–536, Arlington, Virginia, USA, 2010. AUAI Press. ISBN 9780974903965.

Elizabeth A Stuart. Matching methods for causal inference: A review and a look forward. *Statistical science: a review journal of the Institute of Mathematical Statistics*, 25(1): 1, 2010.