

Supplementary material for 'Doubly robust nonparametric inference on the average treatment effect'

BY D. BENKESER

Department of Biostatistics and Bioinformatics, Emory University, 1518 Clifton Rd. NE,
Atlanta, Georgia 30322, U.S.A.

benkeser@emory.edu

M. CARONE

Department of Biostatistics, University of Washington, 1705 NE Pacific Street,
Box 357232, Seattle, Washington 98195, U.S.A.

mcarone@uw.edu

M. J. VAN DER LAAN

Division of Biostatistics, University of California, 108 Haviland Hall,
Berkeley, California 94720, U.S.A.

laan@berkeley.edu

AND P. B. GILBERT

Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Research Center,
1100 Fairview Avenue North, PO Box 19024, Seattle, Washington 98109, U.S.A.

pgilbert@scharp.org

ADDITIONAL SIMULATION STUDIES

Bivariate covariate

We compared the performance of the six estimators studied in § 5 of the main paper with that of other doubly robust estimators, including the modified one-step estimator of Cao et al. (2009), the calibrated likelihood estimator of Tan (2010), the parametric bias-reduced estimator of Vermeulen & Vansteelandt (2014), and the data-adaptive bias-reduced estimator of Vermeulen & Vansteelandt (2016). The former three estimators, as described in the original works, rely on parametric estimators of the outcome regression and propensity score, while the latter allows for data-adaptive estimation of the outcome regression. To highlight the potential benefits of allowing data-adaptive estimation, we used misspecified parametric models for the three parametric-based methods, employing a main-terms-only regression based on W for both the outcome regression and the propensity score. However, where methods allowed for data-adaptive nuisance parameter estimation, we used kernel regression, as in the original simulation. Thus, for the estimator of Vermeulen & Vansteelandt (2016), we estimated the outcome regression using kernel regression and the propensity score using a misspecified logistic regression model, whereas for the six estimators considered in § 5 we used kernel regression for both the outcome regression and the propensity score. We simulated 500 datasets according to the simulation set-up described in § 5, with sample sizes of 250, 1000 and 5000. We compared the Monte Carlo bias of these es-

Table S1. *Bias of estimators ($\times 10$) in the bivariate covariate setting*

n	250	1000	5000
Cao	1.135	1.182	1.188
Verm-1	1.135	1.179	1.187
Tan	0.004	0.017	0.024
Verm-2	0.216	0.115	0.060
OS	0.058	0.015	0.006
TMLE	0.019	0.005	0.002
TMLE-1	0.016	-0.001	0.003
OS-1	0.028	0.001	0.001
TMLE-2	0.001	-0.002	0.003
OS-2	0.028	0.004	0.001

Cao, the estimator of Cao et al. (2009); Verm-1, the estimator of Vermeulen & Vansteelandt (2014); Tan, the estimator of Tan (2010); Verm-2, the estimator of Vermeulen & Vansteelandt (2016); TMLE, targeted minimum loss-based estimator; TMLE-1, targeted minimum loss-based estimator with univariate correction; TMLE-2, targeted minimum loss-based estimator with bivariate correction; OS, one-step estimator; OS-1, one-step estimator with univariate correction; OS-2, one-step estimator with bivariate correction.

timators as well as the coverage probability of nominal 95% confidence intervals constructed using each of these estimators. Confidence intervals for the Tan estimator were computed using Wald intervals based on the estimated variance computed in the R package `iWeigReg` (Tan & Shu, 2013). The confidence intervals for the estimators from Vermeulen & Vansteelandt (2014) and Vermeulen & Vansteelandt (2016) were computed using the code accompanying the respective papers. The confidence intervals for the estimator of Cao et al. (2009) were constructed using a percentile-based nonparametric bootstrap based on 500 bootstrap resamples.

The Monte Carlo bias of each of the estimators is displayed in Table S1. As anticipated, the bias of the parametric estimators does not converge to zero since both nuisance models are misspecified. The bias of the Vermeulen & Vansteelandt (2016) estimator appears to be converging to zero; however, the rate is slower than that of the estimators that allow for data-adaptive estimation of both nuisance parameters. The behaviour of the Vermeulen & Vansteelandt (2016) estimator is thus comparable to the targeted minimum loss-based and one-step estimators that used misspecified logistic regression models. As such, we do expect its bias to converge to zero, though at a rate slower than $n^{-1/2}$.

The coverage probability for nominal 95% confidence intervals is shown in Table S2. The coverage probabilities for the Cao et al. (2009) and Vermeulen & Vansteelandt (2014) estimators converge quickly to zero due to the inconsistency of these estimators. Surprisingly, and in spite of its inconsistency, the Tan (2010) estimator achieves approximately nominal coverage. In Table S1 it can be seen that the bias of this estimator does not converge to zero, and so we anticipate that this estimator will suffer from poor coverage in large enough samples. The coverage probabilities for the Vermeulen & Vansteelandt (2016) estimator do not achieve nominal coverage because this estimator is not asymptotically linear under the misspecified propensity score model. The coverage probabilities for the six estimators utilizing kernel regression for both nuisance parameters achieve their approximate nominal level.

The performance of all estimators that rely on parametric models was quite poor in this simulation study, since by design we used only misspecified models. We did so to emphasize the significant drawbacks of reliance on parametric models alone for nuisance estimation. Certainly, we expect the poor performance of these parametric model-based estimators to be much less drastic

Table S2. Nominal 95% confidence interval coverage probability in the bivariate covariate setting

n	250	1000	5000
Cao	0.275	0.000	0.000
Verm-1	0.269	0.000	0.000
Tan	0.890	0.947	0.943
Verm-2	0.817	0.872	0.840
OS	0.865	0.951	0.954
TMLE	0.871	0.957	0.954
TMLE-1	0.869	0.966	0.952
OS-1	0.880	0.961	0.962
TMLE-2	0.888	0.966	0.958
OS-2	0.871	0.961	0.962

when there is only a small amount of misspecification. However, we want to stress that, for the sake of inference, how severe a certain deviation from the model is depends entirely on the sample size. In particular, what may seem like a small deviation may in fact be quite consequential if the dataset is large. We leave to future work a comparison between the finite-sample operating characteristics of these parametric model-based estimators and those of the fully data-adaptive procedures in settings where the true nuisance functions are well-approximated by parametric models. 70

Six-dimensional covariate

To examine how the estimators perform in settings with covariates of higher dimension, we conducted an additional simulation study inspired by the study reported above. The baseline covariate vector $W = (W_1, \dots, W_6)$ had independent components, with W_1, W_3 and W_5 distributed according to a uniform distribution over the interval $(-2, +2)$, and W_2, W_4 and W_6 distributed as a binary random variable with success probability $1/2$. The conditional probability of $A = 1$ given $W = w$ was the same as in the original simulation, $g_0(w) = \text{expit}(-w_1 + 2w_1w_2)$, as was the conditional probability of $Y = 1$ given $A = a$ and $W = w$, $\bar{Q}_0(a, w) = \text{expit}(0.2a - w_1 + 2w_1w_2)$. 75

In this simulation, rather than using kernel regression to estimate nuisance parameters for the data-adaptive estimators, we used loss-based ensemble machine learning or super learning (van der Laan & Polley, 2007; Polley & van der Laan, 2013). Our ensemble estimator relied on ten-fold crossvalidation to select the convex combination of three candidate estimators that minimized crossvalidated mean squared error. The three candidate estimators were a main-terms-only logistic regression, an AIC-based forward stepwise logistic regression including all two-way interactions, and a gradient-boosted machine (Friedman, 2001) with tuning parameters also selected via ten-fold crossvalidation (Kuhn, 2016). We compared the same estimators as in the previous subsection, again using misspecified main-terms-only logistic regression models for the parametric doubly robust estimators and using super learning for estimators that allowed for data-adaptive estimation. 80

The bias of each estimator considered is displayed in Table S3. The results are similar to those obtained in the bivariate setting. The bias of the data-adaptive estimators converges to zero, while the bias of the parametric estimators does not. The bias of the Vermeulen & Vansteelandt (2016) estimator again converges more slowly than the estimators that allow for data-adaptive estimation of both nuisance parameters. For the smallest sample size considered, the proposed univariate-corrected targeted minimum loss-based estimator performed best, with lower bias than both the uncorrected and the bivariate-corrected targeted minimum loss-based estimator of van der Laan 85

100

Table S3. *Bias of estimators ($\times 10$) in the six-dimensional covariate setting*

n	250	1000	5000
Cao	1.200	1.172	1.180
Verm-1	1.181	1.167	1.179
Tan	1.051	1.034	1.038
Verm-2	0.371	0.171	0.133
OS	0.095	0.001	-0.002
TMLE	0.050	-0.001	-0.002
TMLE-1	0.012	-0.009	-0.006
OS-1	0.069	-0.007	-0.004
TMLE-2	0.027	-0.008	-0.006
OS-2	0.228	0.017	-0.003

Table S4. *Nominal 95% confidence interval coverage probability in the six-dimensional covariate setting*

n	250	1000	5000
Cao	0.214	0.000	0.000
Verm-1	0.204	0.000	0.000
Tan	0.330	0.034	0.002
Verm-2	0.758	0.774	0.664
OS	0.896	0.936	0.948
TMLE	0.896	0.938	0.948
TMLE-1	0.896	0.940	0.944
OS-1	0.872	0.946	0.948
TMLE-2	0.878	0.932	0.938
OS-2	0.808	0.916	0.942

(2014). In Table S4, the coverage probability for the nominal 95% confidence interval about each estimator is provided. Similarly to above, because of strong model misspecification, the parametric estimators do not achieve proper coverage. Furthermore, the surprisingly good nominal coverage achieved by the Tan (2010) estimator in the bivariate scenario is not seen in this case. The coverage probability of the interval based on the Vermeulen & Vansteelandt (2016) estimator again decreases with sample size. The coverage probability of the intervals based on the six estimators utilizing super learning for both nuisance functions is approximately nominal for moderate sample sizes.

REFERENCES

- 110 CAO, W., TSIATIS, A. A. & DAVIDIAN, M. (2009). Improving efficiency and robustness of the doubly robust estimator for a population mean with incomplete data. *Biometrika* **96**, 723–34.
- FRIEDMAN, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Ann. Statist.* **29**, 1189–232.
- KUHN, M. (2016). *caret: Classification and Regression Training*. R package version 6.0-71.
- POLLEY, E. & VAN DER LAAN, M. (2013). *SuperLearner: Super Learner Prediction*. R package version 2.0-10.
- 115 TAN, Z. (2010). Bounded, efficient and doubly robust estimation with inverse weighting. *Biometrika* **97**, 661–82.
- TAN, Z. & SHU, H. (2013). *iWeigReg: Improved Methods for Causal Inference and Missing Data Problems*. R package version 1.0.
- VAN DER LAAN, M. J. (2014). Targeted estimation of nuisance parameters to obtain valid statistical inference. *Int. J. Biostatist.* **10**, 29–57.
- 120 VAN DER LAAN, M. J. & POLLEY, E. C. (2007). Super learner. *Statist. Appl. Genet. Molec. Biol.* **6**, 1–23.
- VERMEULEN, K. & VANSTEELENDT, S. (2014). Bias-reduced doubly robust estimation. *J. Am. Statist. Assoc.* **110**, 1024–36.

VERMEULEN, K. & VANSTEELANDT, S. (2016). Data-adaptive bias-reduced doubly robust estimation. *Int. J. Biostatist.* **12**, 253–82.