The Effect of Varibale Selection Approaches on MSE of IPW ATE Estimator

Notes for working with QMD file

This is an example of citing a table: Table 1, using its label `tbl-main-effects`. Make sure to start *label* for a table with `tbl-`. Review chunks for formatting of labels and captions. They will automatically update labels and captions

This is an example of citing a figure: Figure 1 , using its label `fig-bias-summary`. Make sure to start *label* for a figure with `fig-`

example of citing papers via bibtex files. Citing Shortreed and Ertefaie (2017) using its tag `shortreed2017outcome` in bib file

Introduction

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8227634/#:~:text=In%20the%20Obstetrics%20and%20Perio

https://accpjournals.onlinelibrary.wiley.com/doi/abs/10.1592/phco.23.8.1037.32876?casa_token=-v6XJv5w0DkAAAAA:DJdg3MRA_a9xY093j773zpMOCKSlLYQ2jeJSjeCjR8hc7_J-WIoEeOiwNEuCvI-oUhyVgf-9bc6aNA

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3065283/

Use of random forest with a lot of variables https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5920646/

- Super high dim variable selection and ANDI data

https://onlinelibrary.wiley.com/doi/full/10.1111/biom.13625?casa_token=TPy_un3cN-EAAAAA%3Asm_EoAthJYwoV489HGuY_CnDi0BV7JIWSOvWyTcynDH6dYDq0TOQJGe6pHMzhcFI4z42w

- Some study with bias and mse

https://onlinelibrary.wiley.com/doi/full/10.1002/sim.4469?casa_token=N3XCrX6yrUsAAAAA%3AQ2uBYcmI
ikpXzZnL8_FPmqQhPGPP1No3OxG74JK0fw

Methods

Results

Simulation Results

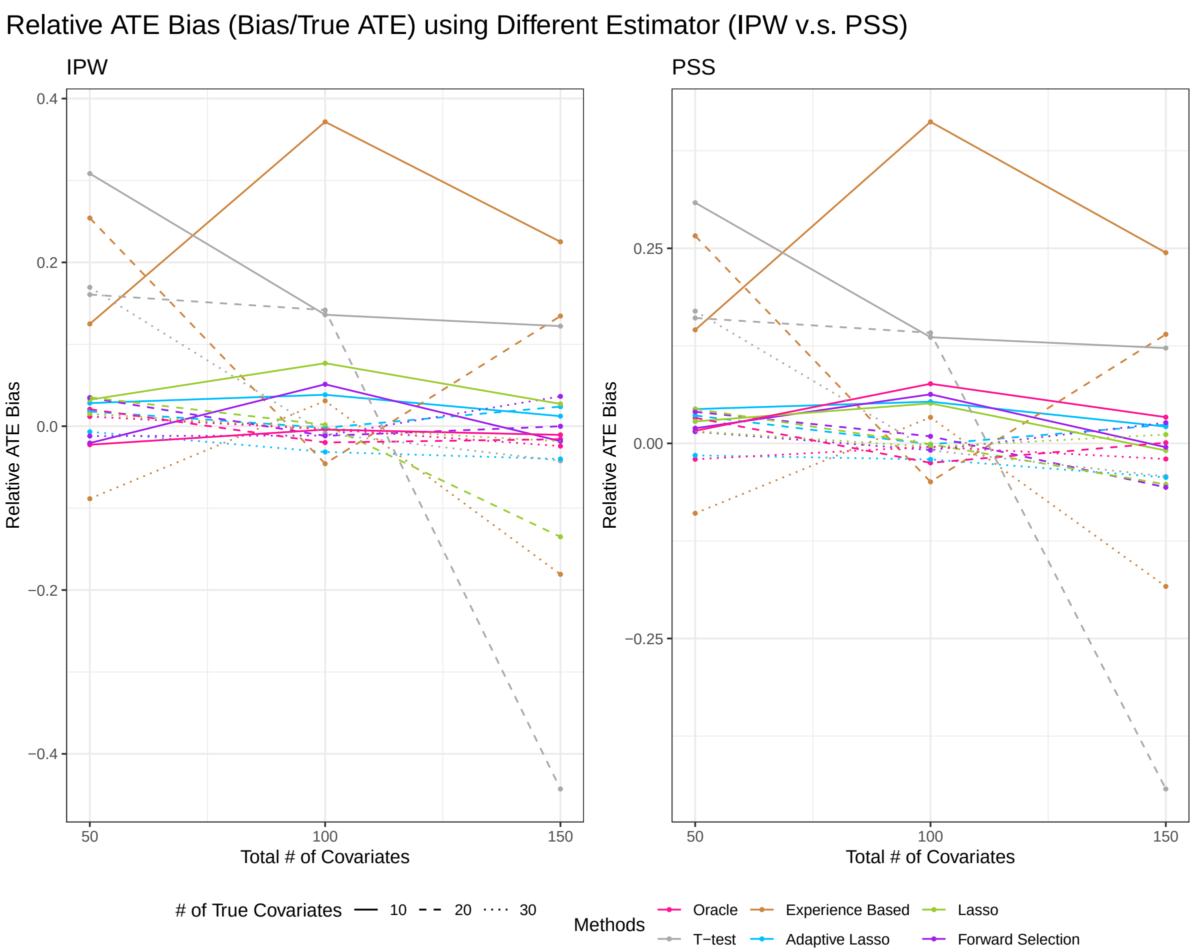


Figure 1: CAPTION FOR THIS PLOT REQURIED

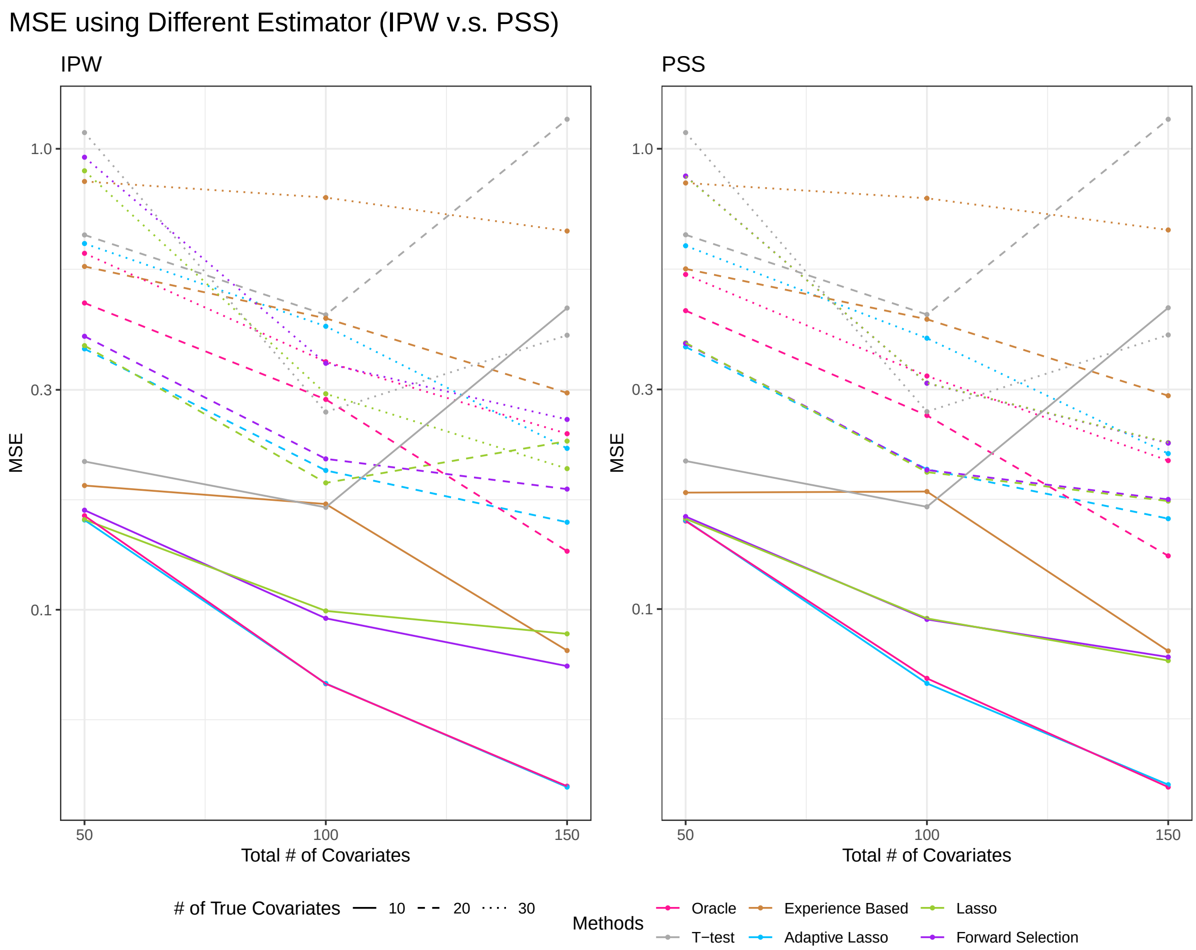


Figure 2: CAPTION FOR THIS PLOT REQURIED

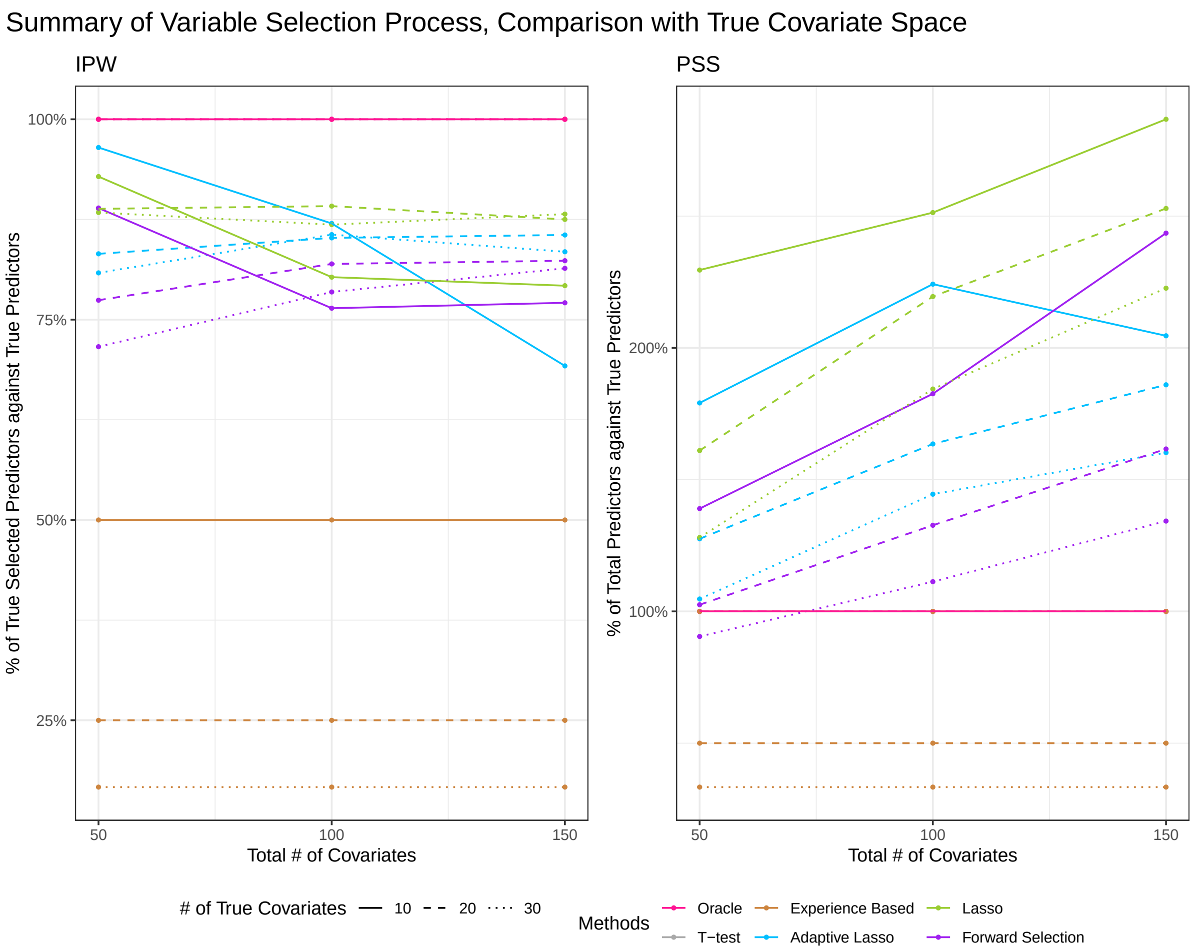


Figure 3: CAPTION FOR THIS PLOT REQURIED

Table 1: Gaussian GLM with log-transformed resposnse effect estiamtes. Coefficients are exponentiated and present the effect as % change

Predictor	Estimate	95% CI	P-value	Significance
m	-1%	(-1.07 % , -0.89 %)	0.0000	*
n	0%	(0 % , 0 %)	0.0485	*
s	4.6%	(-0.32 % , 9.73 %)	0.0672	
methodExperience Based	8.6%	(-39.87 % , 96.04 %)	0.7850	
methodAdaptive Lasso	-1.3%	(-29.44 % , 38.09 %)	0.9395	
methodLasso	72.3%	(19.7 % , 148.02 %)	0.0034	*
methodForward Selection	52.6%	(7.36 % , 116.96 %)	0.0185	*
p_true_selected	-34.9%	(-71.91 % , 51.03 %)	0.3177	
p_total_selected	-5.1%	(-13.59 % , 4.28 %)	0.2772	
s:methodExperience Based	3.4%	(-0.43 % , 7.42 %)	0.0822	
s:methodAdaptive Lasso	0%	(-1.53 % , 1.5 %)	0.9697	
s:methodLasso	-1.8%	(-3.27 % , -0.35 %)	0.0152	*
s:methodForward Selection	-1.2%	(-2.85 % , 0.5 %)	0.1655	
s:p_true_selected	3.3%	(-1.44 % , 8.31 %)	0.1751	

[a] Regression model explains 10.26% of variation in Squared Errors of IPW estimator

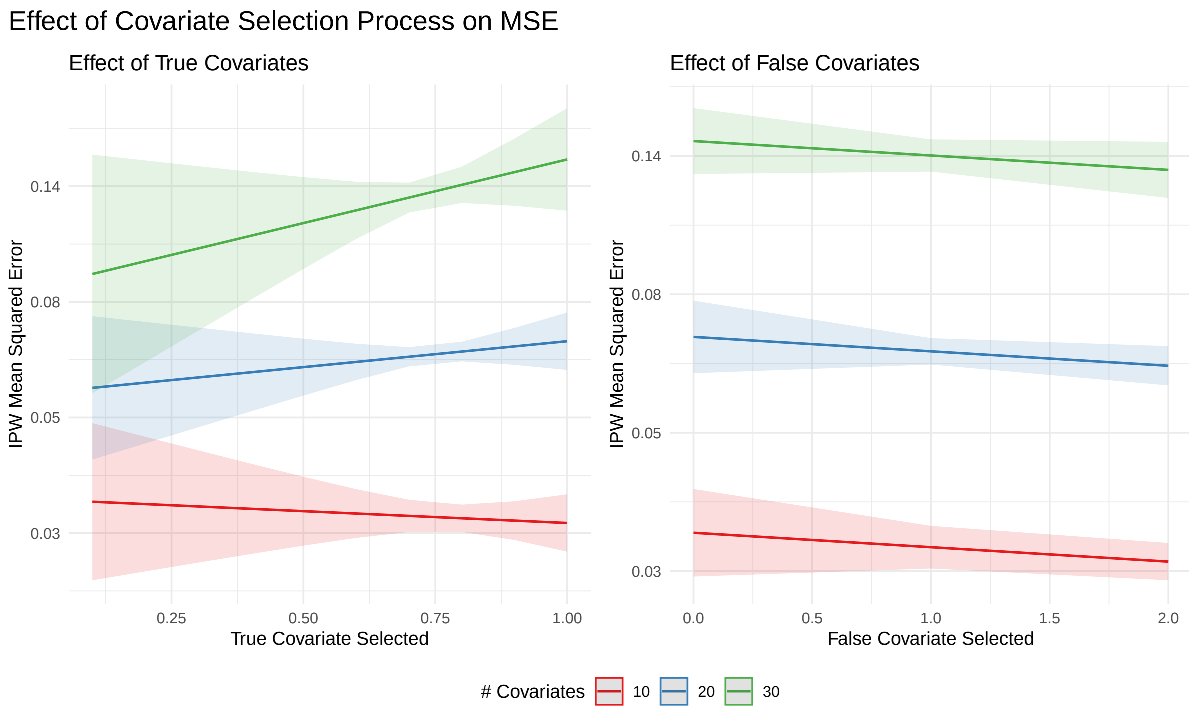


Figure 4: Estimated Marginal Effects IPW Estimator FILL IN THE REST

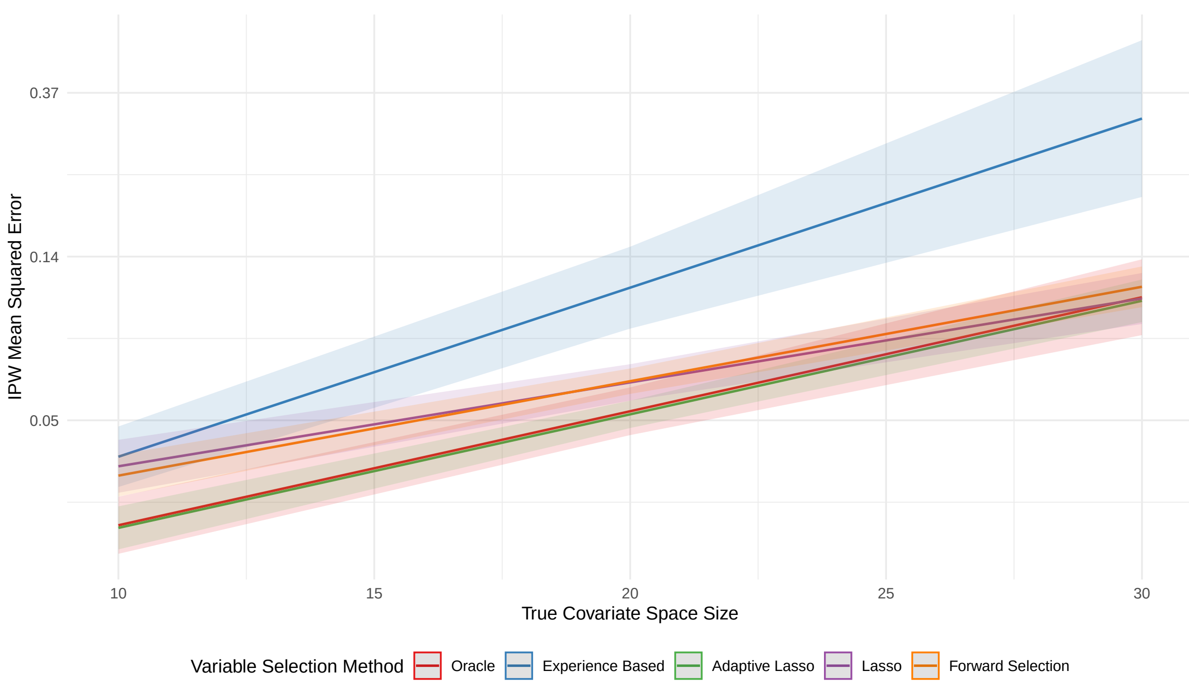
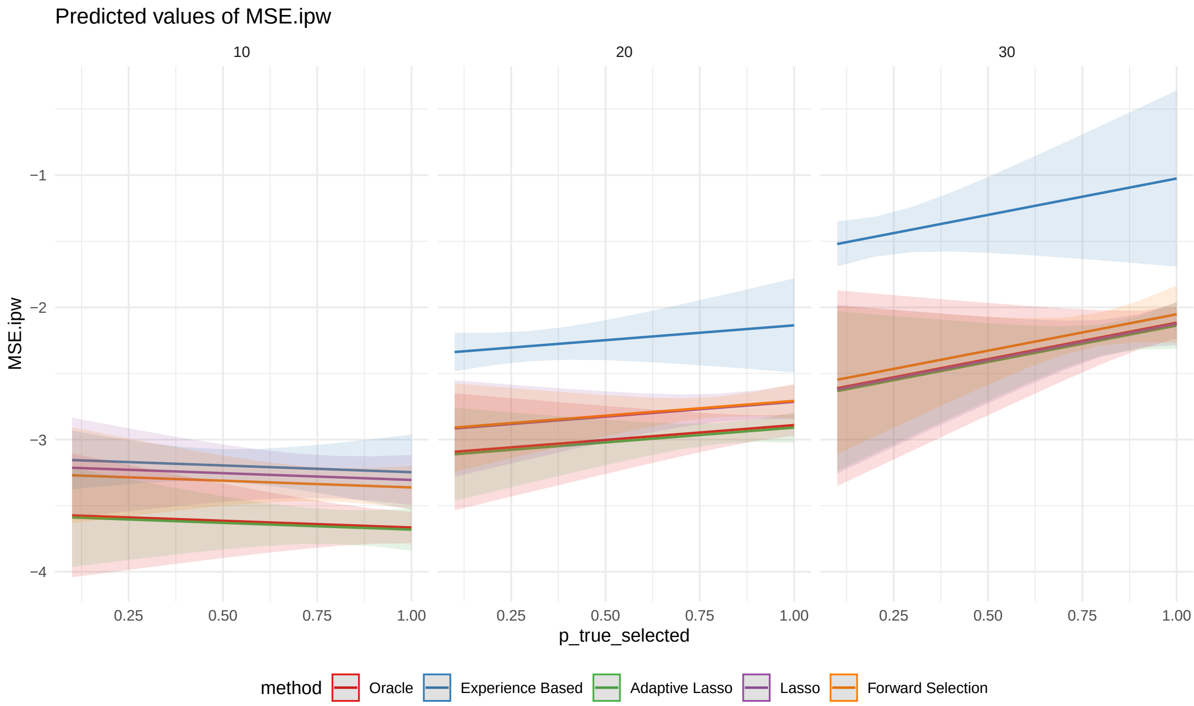
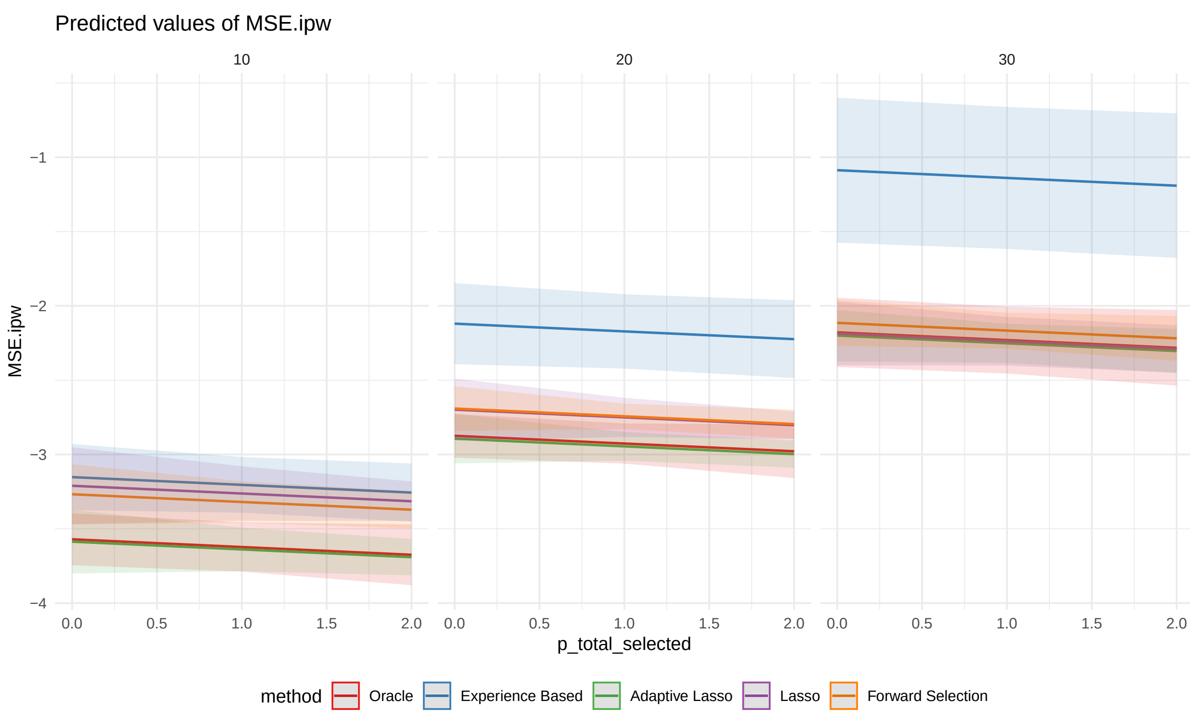


Figure 5: Effect of Variable Selection Method on IPW MSE

Main Effects





Mains

1. As the number of true predictors increases, models get more complicated, and we have on average higeher MSE
2. However, in a situatin where true predictor space is samll, adaptive lasso performs almost as good as the best case scenario of knowing the true predicotrs
3. As we go from simpler to more complicated space, all methods are good. Conclusion: use adaptive lasso

Discussion

Conclusion

References

Shortreed, Susan M, and Ashkan Ertefaie. 2017. "Outcome-Adaptive Lasso: Variable Selection for Causal Inference." *Biometrics* 73 (4): 1111–22. https://doi.org/10.1111/biom.12679.

Appendix