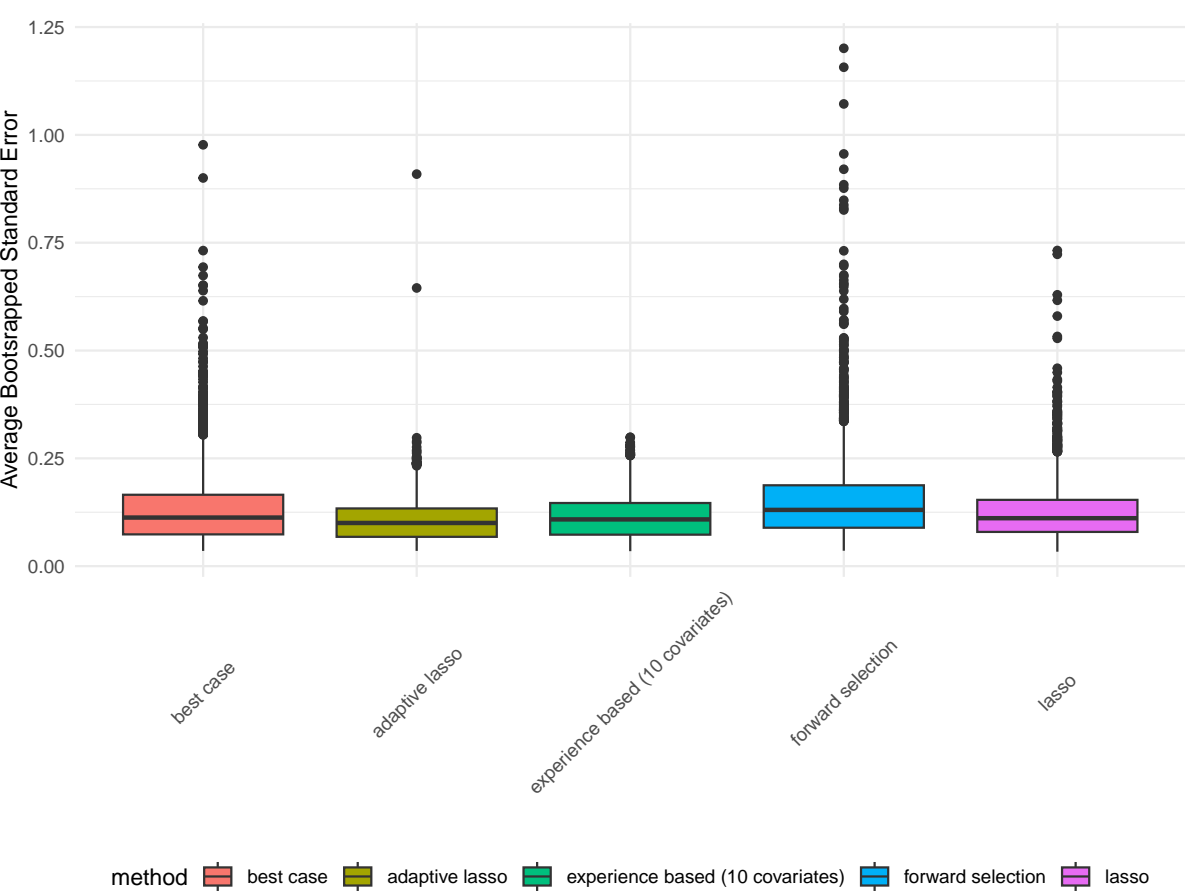


Simulation Project - Step 3 - Results

Results

Introduction



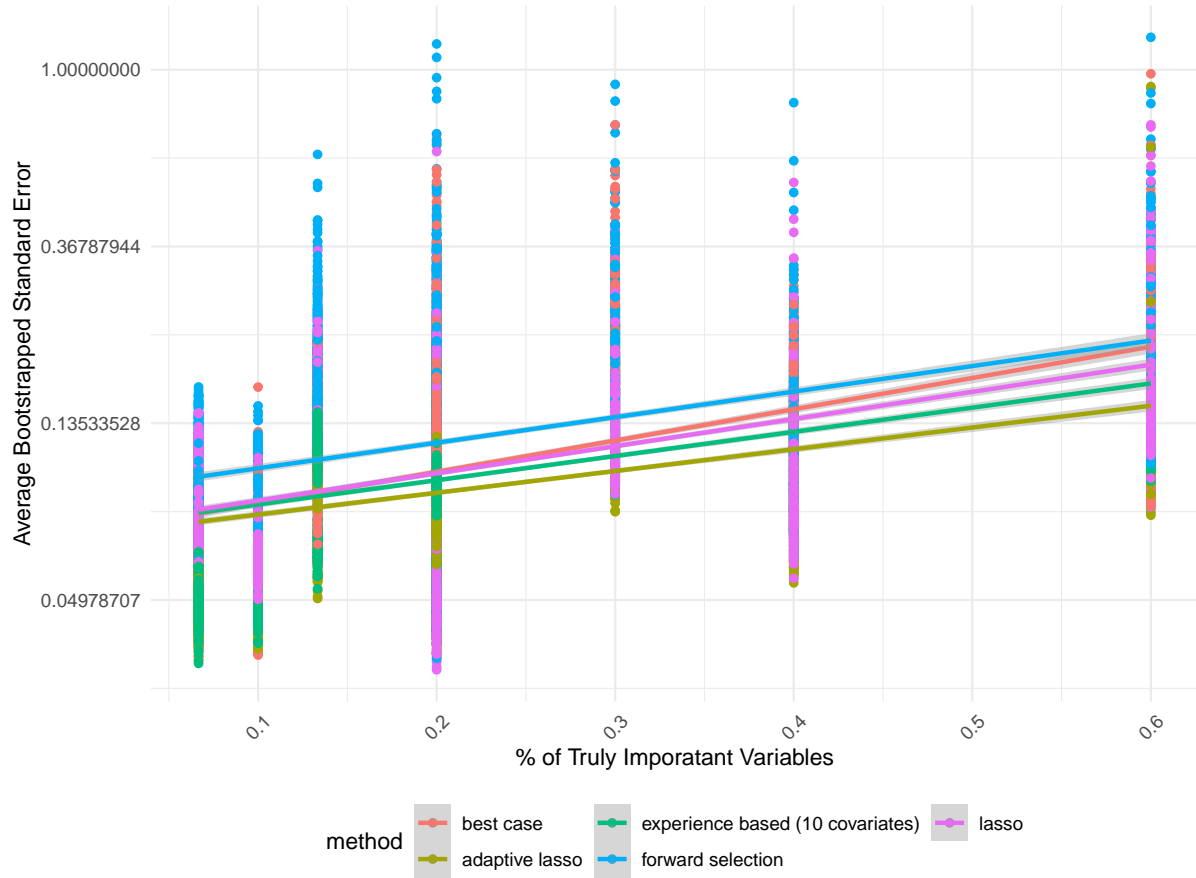
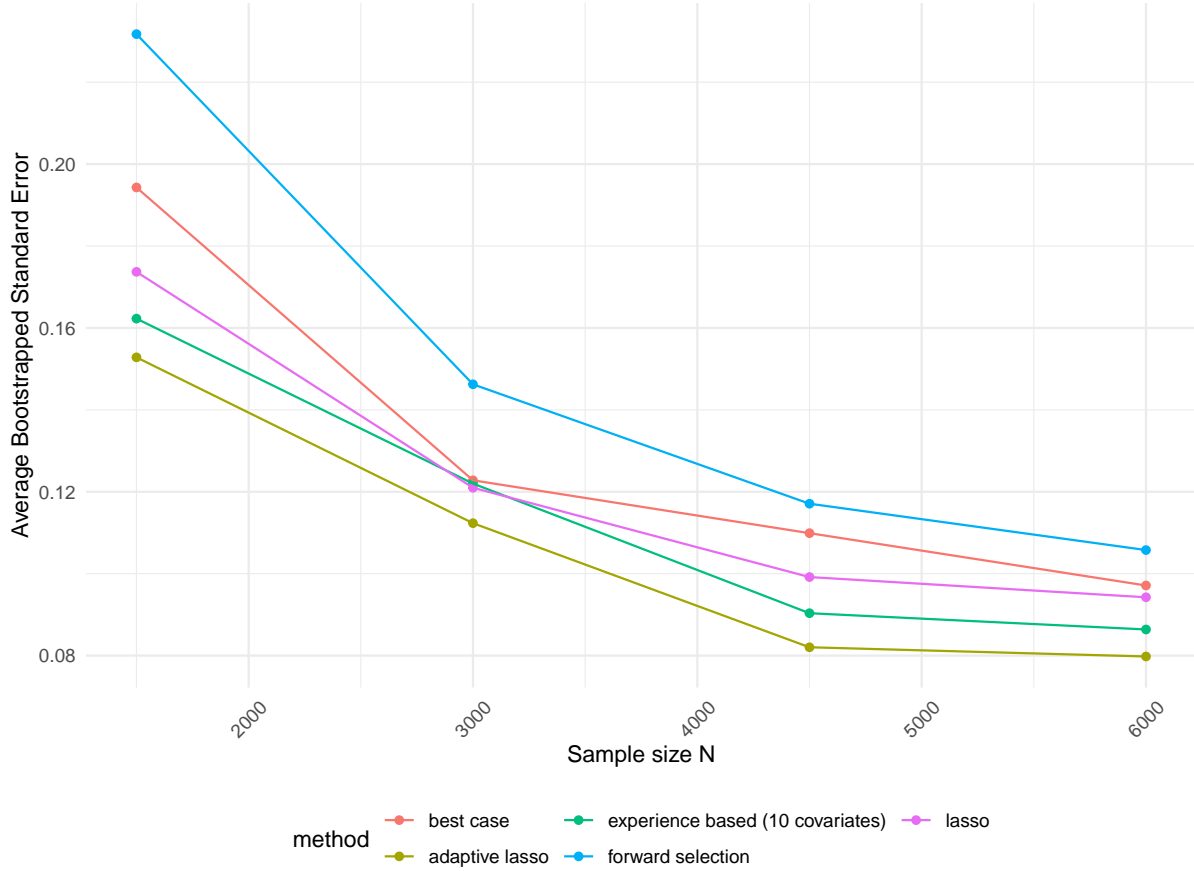


Figure 1: Relationship between % of true covariates and the IPW2 bootstrap standard error. Y-axis on the log scale for more clear visualization



Gaussian Linear Model for Factor Effects

We fit a general regression model to interpret marginal effects of variable factors on the variance estimates using bootstrap. For the purpose of regression modeling, we transform the response, a bootstrap standard error of an estimator from each replication to the logarithmic scale. Due to our data generating mechanism, interpretation of point change in variance would not provide interpretable enough main effects. In contrast, log-transformation allows us to get effect of variable parameters in terms of percentage change, which allows us to compare variances of estimators that have varying magnitude.

Table 1 shows main effects of considered varying parameters on the log-bootstrap error. We consider interaction between variable selection method and the number of true covariates to evaluate how well each method does when the number of confounders that we need to capture varies.

Best case scenario was used as a reference level for the variable selection methods. The variance of best case scenario method is then defined by the variance of residuals specified

Table 1: Gaussian GLM for main effects on Log-Bootstrap Standard Error. All coefficients and interval bounds are exponentiated. Predictors explain 81% of variation in log-transformed bootstrap standard errors of estimators.

Model Term	Exp. Estiamte	Lower CI Bound	Higher CI Bound
(Intercept)	0.0667	0.0652	0.0682
Adaptive Lasso	1.1678	1.1368	1.1997
Experience-Based (10 covariates)	1.1933	1.1616	1.2258
Forward selection	1.2403	1.2066	1.2749
Lasso	1.1729	1.1410	1.2056
m	1.0000	0.9999	1.0001
n	0.9998	0.9998	0.9998
s	1.0557	1.0548	1.0567
Adaptive Lasso:s	0.9852	0.9840	0.9865
Experience-Based (10 covariates):s	0.9880	0.9868	0.9892
Forward selection:s	0.9959	0.9947	0.9972
Lasso:s	0.9903	0.9891	0.9916

in the data generating mechanism. Each other method that is tasked with selecting best predictors under sampling uncertainty is compared to the best case scenario. Main effects for variable selection methods show that Lasso and Adaptive Lasso {causal variable selction method} perform about the same, with average bootsrap stanrd error being about 16% higher than the best case scenario.

Purely data driven method of forward variable selection performs worse than all other methods. The nature of “Experience Based” method, which always selects 5 true and 5 random predictors, makes me think that a step wise forward variable selection method must have a ratio of false positive to true positive predictors that is less than 50%, which then leads to the higher degree of model misspecification, and therefore higher variance of the estiamtor.

We consider parameter s , the number of true predictors for treatment and outcome, as well as its interaction with variable selection methods. As the number of true predictor, or true *confounders*, increases, we expect that the task of selecting true confounders through considered methods becomes harder. Therefore, we want to know how each method behaves as the number of counders increases.

- Main effect of s shows that for the best case scenario, an extra additional predictor resulted in about 5% increase of bootstrap standard error, after adjusting for other vary-ing parameters. It also appears that this effect is stronger than other variable selection methods, as shown on Figure 1. This means that even we are able to correctly specify the model and capture all true confoudners, more complicated data generating schemas with a larger number of truly important predictors still inheret a higher degree of variability.

- For each other variable selection, the effect of an additional significant predictor increased the standard error of an estimator by about 3-3.5%.
- It is interesting to note that the number of total covariates present in the data set has no effect on the bootstrap standard error.