

## ECON 626: Empirical Microeconomics

### Post Double Selection LASSO

Department of Economics  
University of Maryland  
Fall 2019

The do file `econ626-2019-L6-A3-post-double-lasso.do` generates standard normal random variables  $A1-A9$ ,  $B1-B9$ , and  $C1-C9$  and then uses them to generate a treatment dummy,  $T$  and an outcome variable  $Y$ . The  $A$  variables predict treatment, the  $B$  variables predict both treatment and  $Y$ , and the  $C$  variables predict  $Y$  but not  $T$ . The program is a loop, but the local `iters` is set to 1, so the loop will only run once.

1. Run the do file to generate all the variables, and then regress  $Y$  on  $T$ . What does the coefficient suggest about the impact of  $T$ ? Now regress  $Y$  on  $T$  controlling for the  $B$  variables. How does the coefficient on  $T$  change? Why?
2. Review the lasso estimation output (from running the code the first time). Which variables does lasso select as predictors of  $Y$ ? Which variables does lasso select as predictors of  $T$ ? Are these controls the ones you would expect?
3. Set `iters` to 50 and run the code. Compare the average estimate of beta from post single lasso estimation and post double lasso estimation to the true  $\beta$ .
4. Now set `scalefactor` to 0.1 and rerun the code. How do your results differ from (3)?
5. Generate a new treatment dummy,  $P$ , that is randomly-assigned (and hence not correlated with any of the other variables, in expectation). How do PSL and PDL compare under the null and when there is a true treatment effect (for example, if  $Y2 = Y + P$ )?
6. Create a data set that is genuinely sparse by modifying the code to include variables  $D1-D99$ , standard normals that have no relationship to  $T$  or  $Y$ . How do the PSL and PDL estimates of  $\beta$  compare when you include the  $D$  variables as potential covariates?
  - (a) If the results are unsatisfactory, set `iters` to one to examine the number of covariates selected as predictors of  $T$ . Re-estimate using the  $\lambda$  chosen by AIC rather than EBIC — how do your results change? You might also try replacing `lasso2` with `rlasso` throughout.