One Weird Trick

for Better Inference in Experimental Data

Austin Nichols
austinnichols@gmail.com
@austnnchols

How to analyze experiments

- The only way to be sure we are estimating unbiased causal impacts of a “treatment” (intervention, policy, program) is to compare means via an experiment (Freedman 2018a,b, Lin 2013)

- But we can always do better by conditioning on observable (pre-treatment) characteristics: these “covariates” can reduce MSE
  - Stratification/blocking preferred to post hoc statistical adjustment but has its own limitations (Kallus 2018)
  - How should one adjust for covariates *if using a regression* to analyze the experimental data? What variables should be included?

❖ Use the LASSO! Specifically, poregress, dsregress, xporegress, etc.

- New to Stata as of Stata 16, explained in the new [LASSO] manual and in Drukker (2019)

Partialing out

- A series of seminal papers by Belloni, Chernozhukov, and many others (see references) derived partialing-out estimators that provide reliable inference for δ after one uses covariate selection to determine which of many covariates "belong" in the model for outcome Y

$$Y = A\,\delta + X\,\gamma + e$$

where A is a treatment variable of interest and X measures the (possibly very large) set of potential covariates, but many elements of γ are zero

- Essentially, run separate LASSO regressions of Y and A on X and regress residualized $\ddot{Y}$ on residualized $\ddot{A}$ (where $\ddot{A} = A - \hat{A}$)
- The cost of using these poregress, dsregress, xporegress methods is that they do not produce estimates for the covariate coefficients γ

Stata 16 LASSO manual page 12

Solutions that focus on the true model

If your interest is inference about z1 and z2 in the true model that generated the data, the solution is to type

```
. dsregress y z1 z2, controls(x1-x500)
```

or

```
. poregress y z1 z2, controls(x1-x500)
```

or

```
. xporegress y z1 z2, controls(x1-x500)
```

These commands produce the double selection, partialing-out, and cross-fit partialing-out solutions, respectively, for the linear model, but commands also exist for logistic, Poisson, and instrumental-variables regression. These solutions all use multiple lassos and moment conditions that are robust to the model-selection mistakes that lasso makes; namely, that it does not select the covariates of the true model with probability 1. Of the three, the cross-fit partialing-out solution is best, but it can take a long time to run. The other two solutions are most certainly respectable. The cross-fit solution allows the true model to have more coefficients, and it allows the number of potential covariates, x1-x500 in our examples, to be much larger. Technically, cross-fit has a less restrictive sparsity requirement.

Add'l Stata implementations

- ssc desc lassopack, ssc desc pdslasso (Ahrens, Hansen, and Schaffer 2018) released prior to Stata 16 implementations
  - They implement the LASSO (Tibshirani 1996) and the square-root-lasso (Belloni et al. 2011, 2014).
  - These estimators can be used to select controls (pdslasso) or instruments (ivlasso) from a large set of variables (possibly numbering more than the number of observations), in a setting where the researcher is interested in estimating the causal impact of one or more (possibly endogenous) causal variables of interest.
  - Two approaches are implemented in pdslasso and ivlasso: (1) The "post-double-selection" (PDS) methodology of Belloni et al. (2012, 2013, 2014, 2015, 2016). (2) The "post-regularization" (CHS) methodology of Chernozhukov, Hansen and Spindler (2015). For instrumental variable estimation, ivlasso implements weak-identification-robust hypothesis tests and confidence sets using the Chernozhukov et al. (2013) sup-score test.

Regression for experiments

- Note that in the model for outcome Y

$$Y = A\,\delta + X\,\gamma + e$$

- We really should never care about the "effect" of any element of X conditional on A and other elements of X, i.e. we should not care one whit about estimates of γ

- In expectation, A and X are uncorrelated; we just want a data-driven way to eliminate chance correlation between X and A for any X that also has effects on Y in order to reduce the variance of our estimates of δ

- These and other points arose in email correspondence in 2016-2017 with David Judkins who has used LASSO in subsequent studies (Judkins 2019)

Okay, LASSO, but what kind?

- Chetverikov, Liao, and Chernozhukov (2019) show “the cross-validated LASSO estimator achieves the fastest possible rate of convergence in the prediction norm up to a small logarithmic factor”
- Drukker (2019) suggests the plug-in estimator has better small-sample performance in simulations (not reported)
- A bootstrap could give out-of-sample performance measures akin to RandomForest regressions

Simulations

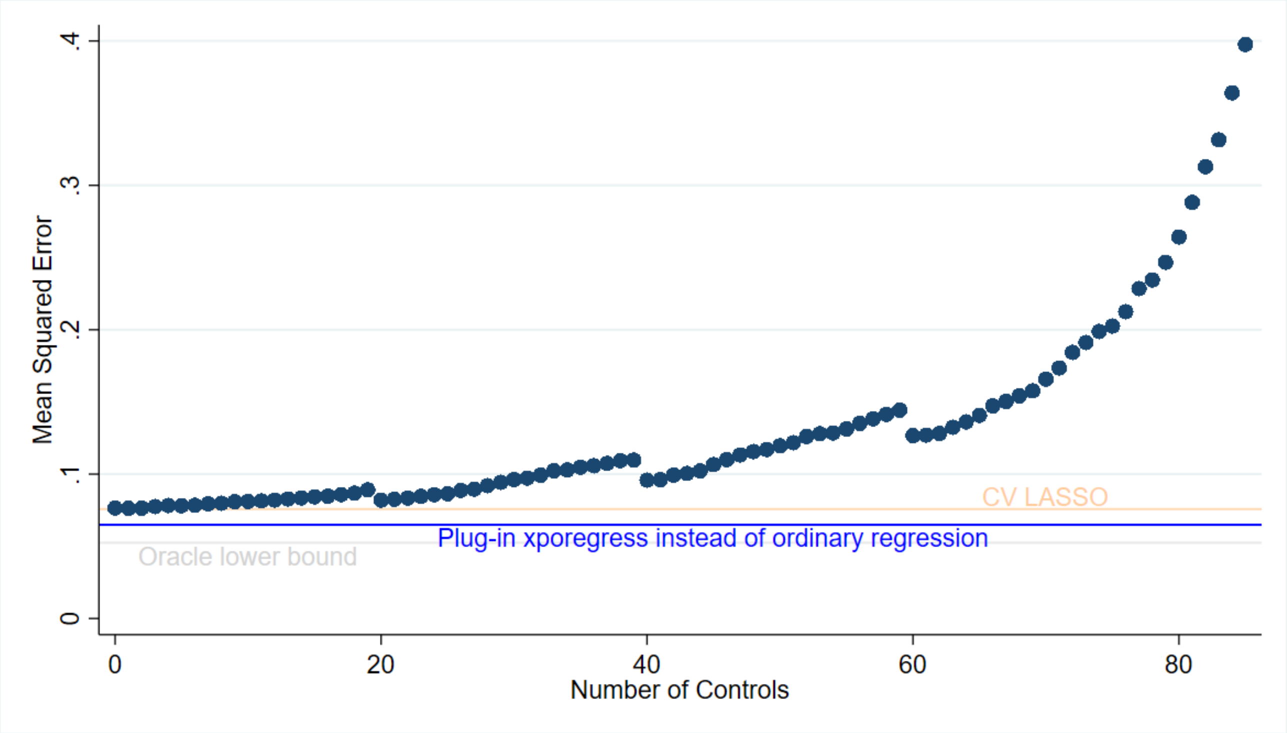
- Suppose we have hundreds of candidate regressors, all distributed lognormal, all uncorrelated with each other
- A few are correlated with Y (every 20^{th})
- How big an improvement might we expect with the xporegress cross-fit partialing-out lasso linear regression with plug-in optimal lambda?

Typical Simulation Results

10,000 iterations with N=100

Regressions use all available controls, zero to 80+

Horizontal lines show performance of xporegress with CV or plug-in selection options



Conclusions

- As we add useless regressors, MSE increases and the occasional useful regressor does not (necessarily) make up for that, but xporegress does better in every realistic case examined
- Alternatives in e.g. Judkins (2019) can introduce bias or introduce size errors (rejection rates deviating from nominal size) but xporegress is safe on both fronts

Credit (blame) for the title to Tim

Austin Nichols
@AustnNchols

Replying to @tmorris_mrc

Wait for my paper "one weird trick"

6:37 AM · 3/10/21 · Twitter for iPhone

View Tweet activity

1 Like

Tim Morris @tmorris_mrc · 3/10/21
Replying to @AustnNchols

I can't wait!
AJE is a good bet

Thirteen Questions About Using Machine Learning...
academic.oup.com

References

Ahrens, A., C. Hansen, and M.E. Schaffer. 2018. pdslasso and ivlasso: Progams for post-selection and post-regularization OLS or IV estimation and inference. http://ideas.repec.org/c/boc/bocode/s458459.html

Belloni, A., D. Chen, V. Chernozhukov, and C. Hansen. 2012. Sparse models and methods for optimal instruments with an application to eminent domain. *Econometrica* 80: 2369–2429.

Belloni, A., and V. Chernozhukov. 2013. Least squares after model selection in high-dimensional sparse models. *Bernoulli* 19: 521–547.

Belloni, A., V. Chernozhukov, and C. Hansen. 2013. Inference for high-dimensional sparse econometric models. In *Advances in Economics and Econometrics: 10th World Congress, Vol. 3: Econometrics,* Cambridge University Press: Cambridge, 245-295.

Belloni, A., V. Chernozhukov, and C. Hansen. 2014a. High-dimensional methods and inference on structural and treatment effects. *Journal of Economic Perspectives,* 28(2): 29–50.

Belloni, A., V. Chernozhukov, and C. Hansen. 2014b. Inference on treatment effects after selection among high-dimensional controls. *The Review of Economic Studies,* 81(2):608–650.

Belloni, A., Chernozhukov, V., Hansen, C. and Kozbur, D. 2016. Inference in High Dimensional Panel Models with an Application to Gun Control. Journal of Business and Economic Statistics 34(4):590-605.

Belloni, A., Chernozhukov, V. and Wang, L. 2011. Square-root lasso: Pivotal recovery of sparse signals via conic programming. Biometrika 98:791-806.

Belloni, A., V. Chernozhukov, and L. Wang. 2014. Pivotal estimation via square-root-lasso in nonparametric regression. Annals of Statistics 42(2):757-788.

Belloni, A., V. Chernozhukov, and Y. Wei. 2016. Post-selection inference for generalized linear models with many controls. *Journal of Business & Economic Statistics* 34: 606–619.

Bühlmann, P., and S. Van de Geer. 2011. *Statistics for High-Dimensional Data: Methods, Theory and Applications*. Berlin: Springer.

Chernozhukov, V., D. Chetverikov, M. Demirer, E. Duflo, C. Hansen, W. Newey, and J. Robins. 2018. Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal,* 21(1): C1–C68.36 / 36

Chernozhukov, V., D. Chetverikov, K. and Kato. 2013. Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. Annals of Statistics 41(6):2786-2819.

Chernozhukov, V. Hansen, C., and Spindler, M. 2015. Post-selection and post-regularization inference in linear models with many controls and instruments. *American Economic Review: Papers & Proceedings,* 105(5):486-490.

Chetverikov, D., Z. Liao, and V. Chernozhukov. 2019. On cross-validated Lasso. arXiv Working Paper No. arXiv:1605.02214. http://arxiv.org/abs/1605.02214.

Drukker, D. 2019. Using the lasso in Stata for inference in high-dimensional models. Presentation at London Stata Conference 5-6 September 2019.

Freedman D. A. (2008a). On regression adjustments to experimental data. Adv. in Appl. Math. 40: 180–193. MR2388610

Freedman, D. A. (2008b). On regression adjustments in experiments with several treatments. Ann. Appl. Stat. 2: 176–196. MR2415599

Hastie, T., R. Tibshirani, and J. Friedman. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2nd ed. New York: Springer.

Hastie, T., R. Tibshirani, and M. Wainwright. 2015. Statistical Learning with Sparsity: The Lasso and Generalizations. Boca Rotaon, FL: CRC Press.

Judkins, D. 2019. "Covariate Selection in Small Randomized Studies." https://ww2.amstat.org/meetings/jsm/2019/onlineprogram/AbstractDetails.cfm?abstractid=307372

Kallus, N. 2018. Optimal *a priori* balance in the design of controlled experiments. *Journal of the Royal Statistical Society: Series B (Statistical Methodology),* 80(1), 85–112.

Lin, Winston. 2013. "Agnostic notes on regression adjustments to experimental data: Reexamining Freedman's critique." *Ann. Appl. Stat.* 7(1): 295 - 318.

Spindler, M., V. Chernozhukov, and Hansen, C. 2016. High-dimensional metrics. https://cran.r-project.org/package=hdm.

Tibshirani, R. 1996. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Series B* 58: 267–288.

Yamada, H. 2017. The Frisch-Waugh-Lovell Theorem for the lasso and the ridge regression. Communications in Statistics - Theory and Methods 46(21):10897-10902.

Zou, H. 2006. The adaptive Lasso and its oracle properties. *Journal of the American Statistical Association* 101: 1418–1429.

Zou, H., and T. Hastie. 2005. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society, Series B* 67: 301–320.

Austin Nichols
austinnichols@gmail.com
@austnnchols

ASSOCIATES

abtassociates.com