International Econometric Review (IER)

A Starting Note: A Historical Perspective in Lasso

Mehmet Caner[®]

North Carolina State University
Submission Time: 12 February 2021 | **Acceptance Time:** 13 July 2021

ABSTRACT

In this note, I will look at history of Lasso estimation, which is the benchmark high dimensional estimation technique. I want to also give a perspective where Lasso may be evolving.

JEL Classifications: C1, C5, C7

LASSO (least absolute subset selection operator) was founded by Tibshirani (1996), and this is tool to estimate and select the model simultaneously. This one step approach unlike what we did in the past, made Lasso very popular. A two-step approach, first selecting the model, and then estimate it, resulted in bad finite sample properties, since the first step mistakes carried over to the second step. Lasso also proved to be a good tool in large scale prediction problems since it can prevent an overfit. Take a simple case of p regressors, and larger than n observations, then simple OLS will not work, due to singularity of Empirical Gram matrix. Then a penalized approach will delete useless coefficients, and keep the important ones, resulting in a sparse regression which can have good prediction properties. The penalty will be on all coefficients, and a tuning parameter will penalize the parameters. Lasso puts l1 type of penalty, since its convex, it has good computation properties.

However, even in the fixed dimensions, the theoretical property of Lasso or Lasso type (with slightly different penalties) were not known. In Knight and Fu (2000) they discovered the technical foundations of Lasso, and established consistency. They show a non-convex penalty Bridge estimation gave a better model selection perfection in theory. The main issue with Lasso or lasso-based estimators are their limits turn out to be non-normal, and came with no standard errors, no confidence intervals. So that meant if we find some coefficient as 0 that means we delete it, but we do not know in classical econometric sense, whether it is significant or not. In my mind, this is not a major impediment for Lasso. The other two disadvantages of Lasso/Lasso type estimators are, they cut small coefficients to zero by mistake (creating underfit), and this is related to pointwise consistency of Lasso estimators. Lasso type estimators are not uniformly consistent over a small set of coefficients that may reflect local to zero behavior.

Uniform inconsistency of lasso and all lasso type estimators are discovered by several articles by Leeb and Potscher (2003, 2005). Another problem is lasso cannot find the perfect model if there exists one. However, adaptive lasso can find it, and this is a data dependent penalty applied unlike lasso. Adaptive lasso is developed by Zou (2006).

After these developments in statistics, in econometrics a theoretical model involving least squares with lasso-type estimators are developed by Knight (2008). Subsequent to that we see an application to GMM via lasso-type estimation by Caner (2009), and shows that classic J

[®] Mehmet Caner, North Carolina State University, (email: mcaner@ncsu.edu).

tests. AIC, BIC model selection is inferior to lasso-type GMM. Afterwards, Caner et al. (2018) propose adaptive elastic net estimator that can handle many invalid moments in linear GMM.

This new estimator can handle correlated moments, as well as invalid moments (instruments) as long as the system is identifiable by a relevant set of instruments. In adaptive elastic net, a data dependent penalty and a ridge type penalty is combined. The main value of their paper is that in case of many instruments, we will not be concerned much about invalid instruments, since the estimator can ignore the information coming from them and use only information from the valid instruments. All these estimators in the past use large number of moments, or large number of regressors but less than the sample size. All these quantities grow.

In the last couple of years, simple usage of lasso-type estimators start changing, a major article that solved uniform inconsistency of lasso, and provide confidence intervals- standard errors and testing in least squares framework is proposed by van de Geer et al. (2014). They start with lasso and debias it, by removing shrinkage bias of lass, and this results in a classical result, with limits of the simple tests can be standard normal. This is a large leap in history of lasso in my mind. After that Caner and Kock (2018) proposed a conservative lasso for the first step, and then debiasing it, which can have model selection consistency properties as well, and with heteroskedastic data with variance covariance estimation in least squares.

An alternative to debiasing and coming up with tests are proposed by Belloni et al. (2014). They partial out regressors and solve the problem in that way. Also in high dimensional models, Belloni et al. (2012) proposed sparse models for optimal instruments and finite sample prediction bounds. We can say that Victor Chernozhukov of MIT has been the leading figure in economics in high dimensions, he has several other papers, that are not cited here.

Recently, an interesting idea is developed by Esra Ulasan (then Ege University, PhD, NCSU Post-Doc), she proposed merging a technique called nodewise regression with finance. Nodewis regression is a neighborhood regression where we discover our neighbors through lasso. This turned out to be great idea, and Callot et al. (2021) get very good results compared to other portfolio variance calculations in high dimensions.

REFERENCES

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

Callot, L., M. Caner, O. Onder and E. Ulasan (2021). A nodewise regression approach to estimating large portfolios. *Journal of Business and Economic Statistics*, Forthcoming.

Caner, M. (2009). Lasso-type GMM estimation. *Econometric Theory*, 25, 270-290.

Caner, M. and A.B. Kock (2018). Asymptotically honest confidence regions for high dimensional parameters by the desparsified conservative lasso. *Journal of Econometrics*, 203, 143-168.

Caner, M., X. Han and Y. Lee (2018). Adaptive elastic net GMM estimation with many invalid moment conditions: Simultaneous model and moment selection. *Journal of Business and Economic Statistics*, 36, 24-46.

Knight, K. and W. Fu (2000). Asymptotics for lasso-type estimators. *Annals of Statistics*, 28, 1356-1378.

Knight, K. (2008). Shrinkage estimation for nearly singular designs. *Econometric Theory*, 24, 323-337.

Leeb, H. and B. Potscher (2003). The finite sample distribution of post-model selection estimators and uniform versus nonuniform approximations. *Econometric Theory* 19, 100-142.

Leeb, H. and B. Potscher (2005). Model selection and inference: facts and fiction. *Econometric Theory*, 21, 21-59.

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

Van de Geer, S., P. Buhlmann, Y. Ritov and R. Dezeure (2014). On asymptotically optimal confidence regions and tests for high dimensional models. *Annals of Statistics*, 42, 1166-1202.

Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association*, 101, 1418-1429.