

Studies in Advanced Econometrics (212.664 - 002)
Spring 2023

Course:

Lectures: Tuesday, 09:30-12:15 (Classroom: M306)
TA Sessions: TBA

Instructor: Prof. Myung Hwan Seo

Office: 16.644
Email: myunghseo@snu.ac.kr
Office Hours: Tuesday 13:30–14:30 or by appointment

Course Description: This course intends to expose the students to the current econometric and statistical approaches to the machine learning methods and theory that deal with high-dimensional models and big data environment. It introduces various high-dimensional estimation and model selection methods for general regression models such as the linear regression, instrumental variable regression, logistic regression, quantile regression with a particular emphasis on the least absolute shrinkage and selection operator (aka lasso). It develops various distribution theory to cover related issues on prediction/forecasting and statistical inference. And it also discusses numerical algorithms for the methods and demonstrates the execution of the algorithms. The lectures are based mainly on the first three books in the reference list below.

Evaluation: In-class Written Examination, , Assignment, Presentations.

- Presentation can be given on the following:
 1. Application on real data.
 2. Monte Carlo simulation (for both a new method and an existing one).
 3. Give a short lecture on a related method in machine learning literature that is not covered in the class.
 4. Theoretical extension of a recent paper.

Outline

1. Review of the (linear) regression
 - (a) Variable selection and post-selection Inference (Hansen Ch.28)

- (b) Series Estimation (Hansen Ch.20)
 - (c) Logit/Probit (Hansen Ch.25)
 - (d) Linear Regression with Change-Points (Structural Break/Threshold/Tipping)
2. Introduction to the Lasso ([BG])
 - (a) Numerical property of a lasso solution
 - (b) Theoretical property
 - (c) Optimization Algorithm
 3. Variants of Lasso
 - (a) Adaptive lasso
 - (b) Weighted lasso
 - (c) Fused lasso
 - (d) SCAD
 4. Inference
 - (a) Double selection
 - (b) Residual Regression
 5. Other Prediction Methods (Machine Learning)
 - Some generic big data environment and applications ([5, 6, 23])
 - Supervised vs Un-supervised Learning
 - Factor Analysis
 - Regression Trees
 - Bagging:
 - Review of Bootstrap
 - Random forest
 - Neural Networks
 - Boosting
 - Support Vector Machines
 6. Stochastic Gradient Descent
 7. Inference on potentially many parameters

References

- [BG] Bühlmann, P. and S. van de Geer (2011). *Statistics for High-Dimensional Data*. DE, Springer.
- [1] Hastie, T., Tibshirani, R., Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Second Edition, Springer.
- [2] B. Hansen (2021). *Econometrics*. Available at <http://www.ssc.wisc.edu/~bhansen/econometrics/>
- [3] JAMES, G., D. WITTEN, T. HASTIE, and R. TIBSHIRANI (2014): *An Introduction to Statistical Learning: With Applications in R*. Springer New York.
- [4] van der Vaart, A. and J. Wellner (1996). *Weak Convergence and Empirical Processes*. Springer.
- [5] Athey, S. (2017). Beyond prediction: Using big data for policy problems. *Science*, 355(6324), 483-485.
- [6] Athey, S. (2019). 21. The Impact of Machine Learning on Economics. *The economics of artificial intelligence*, 507-552.
- [7] ANDREWS, D. W. K. (1994): Empirical Process Methods in Econometrics, in *Handbook of Econometrics*, Vol. \ Iv. Amsterdam: North-Holland, 2247–2294.
- [8] ATHEY, S., and G. IMBENS (2015): A Measure of Robustness to Misspecification, *American Economic Review*, 105, 476-480.
- [9] Bai, J. and P. Perron (1998). "Estimating and Testing Linear Models with Multiple Structural Changes." *Econometrica* 66(1): 47-78.
- [10] Bai, J. and P. Perron (2003). Computation and analysis of multiple structural change models. *Journal of Applied Econometrics* 18(1): 1-22.
- [11] P. Perron (2006) Dealing with Structural Breaks. in *Palgrave Handbook of Econometrics*, Vol. 1, Palgrave Macmillan, 278-352.
- [12] BAJARI, P., D. NEKIPELOV, S. P. RYAN, and M. Y. YANG (2015): Machine Learning Methods for Demand Estimation, *American Economic Review*, 105, 481-485.
- [13] 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.

- [14] 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.
- [15] — (2014): High-Dimensional Methods and Inference on Structural and Treatment Effects, *Journal of Economic Perspectives*, 28, 29-50.
- [16] BOUCHERON, S., G. LUGOSI, and O. BOUSQUET (2004): Concentration Inequalities, *Advanced Lectures on Machine Learning*, 3176, 208-240.
- [17] Chen, Xiaohong (2007): “Large Sample Sieve Estimation of Semi-Nonparametric Models,” In: James J. Heckman and Edward E. Leamer, Editor(s), *Handbook of Econometrics*, Elsevier, Volume 6, Part B, 5549-5632,
- [18] EINAV, L., and J. LEVIN (2014): Economics in the Age of Big Data, *Science*, 346.
- [19] Giacomini, R., Politis, D.N., and H. White (2013). A Warp-Speed Method for Conducting Monte Carlo Experiments Involving Bootstrap Estimators, *Econometric Theory* 29(3): 567-589.
- [20] KLEINBERG, J., J. LUDWIG, S. MULLAINATHAN, and Z. OBERMEYER (2015): Prediction Policy Problems, *American Economic Review*, 105, 491-495.
- [21] KNIGHT, K., and W. FU (2000): Asymptotics for Lasso-Type Estimators, *Ann. Statist.*, 28, 1356–1378.
- [22] LEEB, H., and B. M. POTSCHER (2005): Model Selection and Inference: Facts and Fiction, *Econometric Theory*, 21, 21-59.
- [23] Mullainathan, S., & Spiess, J. (2017). Machine learning: an applied econometric approach. *Journal of Economic Perspectives*, 31(2), 87-106.
- [24] RUPPERT, D. (2002): Selecting the Number of Knots for Penalized Splines, *Journal of Computational and Graphical Statistics*, 11, 735-757.
- [25] TIBSHIRANI, R. (1996): Regression Shrinkage and Selection Via the Lasso, *J. Roy. Statist. Soc. Ser. B*, 58, 267–288.
- [26] VARIAN, H. R. (2014): Big Data: New Tricks for Econometrics, *Journal of Economic Perspectives*, 28, 3-28.