

Econometrics of High-Dimensional Models

Lecture: Monday and Wednesday, 9:30–10:45 am

Professor: Denis Chetverikov (chetverikov@econ.ucla.edu)

Office Hours: by appointment

Course Description: This class provides an introduction to econometrics of high-dimensional models. The class will cover the following topics:

1. relevant results in probability theory;
2. estimation of linear high-dimensional models using Lasso and Dantzig selector methods;
3. estimation of generalized linear high-dimensional models using l1-penalized M-estimators;
4. basics of machine learning;
5. inference in high-dimensional models via double machine learning;
6. related topics in econometrics such as grouped fixed effect estimators in panel data and many moment inequalities.

Evaluation: Class participation and home assignments

Textbooks: The class will be primarily based on research papers, but as a general reference, highly recommended textbooks are

- Wainwright (2019), *High-dimensional statistics: a non-asymptotic view point*. Cambridge University Press.
- Vershynin (2018), *High-dimensional probability: an introduction with applications in data science*. Cambridge University Press.

Other relevant textbooks are

- Buhlmann and van de Geer (2011), *Statistics for high-dimensional data: methods, theory, and applications*. Springer Series in Statistics.
- Hastie, Tibshirani, and Friedman (2009), *The elements of statistical learning: data mining, inference, and prediction*. Springer Series in Statistics.
- Giraud (2015), *Introduction to high-dimensional statistics*. CRC Press.
- van de Geer (2016), *Estimation and testing under sparsity*. Springer: Lecture Notes in Mathematics.

References:

1. Adamczak (2008), A tail inequality for suprema of unbounded empirical processes with applications to Markov chains, *Electronic Journal of Probability*.
2. Athey, Tibshirani, and Wager (2019), Generalized random forests, *Annals of Statistics*.

3. Bonhomme and Manresa (2015), Grouped patterns of heterogeneity in panel data, *Econometrica*.
4. Belloni, Chen, Chernozhukov, and Hansen (2012), Sparse models and methods for optimal instruments with an application to eminent domain, *Econometrica*.
5. Belloni and Chernozhukov (2011), l1-penalized quantile regression in high-dimensional sparse models, *Annals of Statistics*.
6. Belloni and Chernozhukov (2011), High-dimensional sparse econometric models: an introduction, in *Inverse Problems and High-Dimensional Estimation*.
7. Belloni, Chernozhukov, Fernandez-Val, and Hansen (2017), Program evaluation and causal inference with high-dimensional data, *Econometrica*.
8. Belloni, Chernozhukov, and Hansen (2014), Inference on treatment effects after selection among high-dimensional controls, *Review of Economic Studies*.
9. Belloni, Chernozhukov, Chetverikov, Hansen, and Kato (2018), High-dimensional econometrics and regularized GMM, *prepared for Handbook of Econometrics*.
10. Belloni, Chernozhukov, Chetverikov, Demirer, Duflo, Hansen, Newey, and Robins (2018), Double/debiased machine learning for treatment and structural parameters, *Econometrics Journal*.
11. Bickel, Ritov, and Tsybakov (2009), Simultaneous analysis of lasso and dantzig selector, *Annals of Statistics*.
12. Chernozhukov, Chetverikov, and Kato (2013), Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors, *Annals of Statistics*.
13. Chernozhukov, Chetverikov, and Kato (2017), Central limit theorems and bootstrap in high dimensions, *Annals of Probability*.
14. Chernozhukov, Chetverikov, Kato (2019), Inference on causal and structural parameters using many moment inequalities, *Review of Economic Studies*.
15. Chetverikov and Sorensen (2019), Bootstrap after (cross-)validation, *in preparation*.
16. Chudik, Kapetanios, and Pesaran (2018), A one-covariate at a time, multiple testing approach to variable selection in high-dimensional linear regression models, *Econometrica*.
17. Fan, Liao, and Yao (2015), Power enhancement in high-dimensional cross-sectional tests, *Econometrica*.
18. Farrell, Liang, and Misra (2018), Deep neural networks for estimation and inference: application to causal effects and other semiparametric estimands, *working paper*.
19. Gautier, Rose, and Tsybakov (2018), High-dimensional instrumental variable regression and confidence sets, *working paper*.

20. Javanmard and Montanari (2014), Confidence intervals and hypothesis testing for high-dimensional regression, *Journal of Machine Learning Research*.
21. Meinshausen and Bühlmann (2006), High-dimensional graphs and variable selection with the lasso, *Annals of Statistics*.
22. Van de Geer (2008), High-dimensional generalized linear models and the lasso, *Annals of Statistics*.
23. Van de Geer, Bühlmann, Ritov, and Dezeure (2014), On asymptotically optimal confidence regions and tests for high-dimensional models, *Annals of Statistics*.
24. Wainwright (2009), Sharp thresholds for high-dimensional and noisy sparsity recovery using ℓ_1 -constrained quadratic programming (lasso), *IEEE transactions on information theory*.
25. Zhao and Yu (2006), On model selection consistency of lasso, *Journal of Machine Learning Research*.
26. Zhang and Zhang (2014), Confidence intervals for low-dimensional parameters in high-dimensional linear models, *Journal of the Royal Statistical Society, Series B*.
27. Zou (2006), The adaptive lasso and its oracle properties, *Journal of the American Statistical Association*.