Topics in Econometrics and Machine Learning (ECON 6121G)

Economics, HKUST

Spring **2026**

Course Description: This PhD course (3 credits) surveys modern econometrics and machine learning methods for high-dimensional prediction and causal analysis. We build from regression with many regressors (Lasso and variants; series/factor methods) to double machine learning, time-series modeling and forecasting, including stationarity, autoregression/ADL, multi-step prediction, HAC inference, and nonstationary behavior (trends, breaks, cointegration, volatility). We also cover dynamic causal effects—distributed lags, local projections, and structural VARs—emphasizing identification and inference under realistic data limitations. We then overview general supervised learning tools (trees, bagging, random forests, boosting, SVMs, neural networks) and optimization algorithms including stochastic optimization for econometrics: stochastic gradient methods for moment conditions, scalable inference (random scaling and plug-in), and inference when the number of target parameters is large.

By the end of this course, students will understand main econometric tools that bridge high-dimensional machine learning with rigorous causal and structural econometrics. They will be able to navigate the "curse of dimensionality" using penalized estimators and stochastic optimization while mastering dynamic modeling through Local Projections and Structural VARs. Ultimately, students will develop the proficiency to execute flexible predictive modeling that maintains valid statistical inference for identifying causal effects and shocks in complex, large-scale datasets.

Assessment will be based on exams (60%), presentations, and/or a term paper (40%).

Prerequisites: graduate level econometrics and statistics (or equivalent)

Outline

1. Regression with Many Regressors.
2. LASSO
3. Factor Analysis

4. Time Series and Forecasting.
5. Time Series Theory (TS01)
   (a) LLN/CLT for time series data.
   (b) Heteroskedasticity Autocorrelation Robust (HAR) inference.
6. Nonstationary Time Series
   (a) Trend, Break, Non-linearity.
   (b) FCLT
7. Dynamic Causal Effects.
   (a) Distributed Lag Model (SW.16)
   (b) Local Projection and Structural VAR
8. SGD and Random Scaling inference

References

[BG] Bühlmann, P. and S. van de Geer (2011). Statistics for High-Dimensional Data. DE, Springer.

[HTF] Hastie, T., Tibshrani, R., Friedman, J. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition, Springer.

[Hansen] Hansen, B. (2022). Econometrics. Princeton University Press.

[JWHT] James, G., D. Witten, T. Hastie, and R. Tibshirani (2014): An Introduction to Statistical Learning: With Applications in R. Springer New York.

[SW] Stock, J. H., & Watson, M. W. (2020). Introduction to econometrics. Pearson.

Others will be given during lectures.