Machine Learning and Big Data with Macroeconomics Applications

Center for Monetary and Financial Economics (CMFE) Workshop

I believe that machine learning will have a dramatic impact on the field of economics within a short time frame.
Professor Susan Athey in Impact of Machine Learning in Economics (2018).

The intersection of Machine Learning (ML) with econometrics and applied statistics is rapidly shaping up the research landscape in economics (Athey (2018), Mullainathan (2017)). ML has gained prominence due to the availability of large datasets (*Big Data*) that can be studied to improve our understanding of consumer and firm behaviour, financial markets, labour markets, trade, among many other areas. These developments have opened up new research avenues, and led to interest from both public and private sectors.

However, little progress has been made to understand the properties of ML methods when applied to macroeconomic forecasting. That is, the black box remains closed. The objective of this course is to bring an understanding of machine learning models that goes beyond the coronation of a single winner for a specific forecasting target. This will be useful for subsequent predictive model building.

Participants will first learn how to define the forecasting problem in presence of large sets of predictors. Then, how to use machine learning methods to reduce dimensionality and to extract relevant information for the prediction of important macroeconomic outcomes. Finally, they will learn how to evaluate the competing models and to construct and understand their own forecasts.

This course has two parts. The first day, we cover econometric forecasting models combined with *light* machine learning techniques that have been used to predict macroeconomic aggregates. During the second day we cover some *heavier* non-linear and non-parametric ML methods. The first day course level is intermediate, while the second day is of advanced nature. Empirical examples and simulations will be used to illustrate the methods. These are coded in Matlab, but similar techniques are available for R and Python as well.

A typical day consists of the lecture in the morning part followed by applications and the practical exercises in the second part. Applications in economics and finance will be considered to illustrate the different methodologies. Exchanges and discussions are encouraged.

Outline

Day 1: Light Machine Learning for Macroeconomic Forecasting

Theory of forecasting

Model (variables) selection:

- In-sample criteria
- Regularized regressions
- Cross-validation for time series

Forecasting methods

- Factor-augmented regressions
    - Diffusion indices and variations
    - Three-pass regression filter
- Factor models
    - Factor-augmented VAR
    - Dynamic Factor Model
- Complete subset regressions

Applications in class:

- High-dimensional macroeconomic panels
- Forecasting industrial production growth (real activity) and CPI inflation

Main reference: Leroux, M., Kotchoni, R. et D. Stevanovic (2017), "Macroeconomic forecasts accuracy in data-rich environment", UQAM

Day 2: Light Machine Learning for Macroeconomic Forecasting

Forecasting methods

- Factor-augmented regressions: extensions
    - Ridge regression
    - Kernel ridge regression
    - Random forests
- Regularized Data-Rich Model Averaging
- Forecasts combinations

Applications in class:

- High-dimensional macroeconomic panels
- Forecasting industrial production growth (real activity) and CPI inflation

Main reference: Goulet-Coulombe P., Leroux, M., Stevanovic D. and S. Surprenant (2018), "How is Machine Learning Useful for Macroeconomic Forecasting", UQAM

Other references

- James, Witten, Hastie, Tibshirani (JWHT), *An Introduction to Statistical Learning*, Springer texts, http://www-bcf.usc.edu/~gareth/ISL/ISLR%20First%20Printing.pdf

- Hastie, Tibshirani, Friedman, (HTF), *The Elements of Statistical Learning*, Springer Series in Statistics, https://web.stanford.edu/~hastie/Papers/ESLII.pdf

- Additional references for Support Vector Machine Regression:
  - http://kernelsvm.tripod.com
  - Smola, A. et B. Scholkopf (2004), « A tutorial on support vector regression », *Statistics and Computing* 14, 199-222.

- Additional references for Dynamic factor model:
  - Stock, J.H. et M.H. Watson (2016), "Factor Models and Structural Vector Autoregressions in Macroeconomics" in *Handbook of Macroeconomics,* Vol2A, John B. Taylor and Harald Uhlig (eds), 2016, Chapter 8, pp 415-526
  - Bai, J. et Ng, S. (2008), "Forecasting economic time series using targeted predictors", *Journal of Econometrics* 146, 304–317
  - Bai, J. et Ng, S. (2008), "Boosting Diffusion Indices", *Journal of Applied Econometrics*, 24:4, 607-629
- Additional references for forecasting:
  - Giannone, D., Lenza, M. and G. Primiceri (2018), "Economic Predictions with Big Data: The Illusion of Sparsity", Staff Report No. 847. NYFed
  - Diebold, F.X., *Forecasting*, http://www.ssc.upenn.edu/~fdiebold/Textbooks.html
  - Stock, J.H. et M.H. Watson (2017) "Twenty Years of Time Series Econometrics in Ten Pictures" *Journal of Economic Perspectives*, Vol. 31 No. 2.