PhD Talks at Queen's University, Canada

Hashem Pesaran

Trinity College, Cambridge University, and University of Southern California

September 16, 2025

Title: Big Data Analytics: Alternative Approached and New Perspectives

Abstract:

Large data sets, "Big Data", are available in many forms (numeric, textual, bar charts, video), and cover many different dimensions (space, time, firms, households, sectors). In my talk I consider the different challenges we face in analyzing such data sets. I will argue that we need to consider different approaches depending on the type of data available and the objective(s) of the study. High dimensional spatiotemporal panels with large n (cross section units) and T (time) require a different treatment as compared to large online data sets that are mainly cross-sectional. Models with many data points are also to be distinguished from models with many parameters. With this in mind I plan to cover:

- Machine learning techniques (penalized regressions with focus on Lasso, partial least squares, boosting, clustering, and random forests).
- Econometric techniques (PCA and factor models, OCMT, one covariate multiple testing, Bayesian shrinkage techniques, large high dimensional VARs).
- High-dimensional spatiotemporal models, with focus on estimation of heterogeneous spatiotemporal models with applications to the analysis of ripple effects.
- Global VAR modelling with application to the analysis of common shocks and their transmission and spill-over effects across countries, regions and counties.

Selected References

*Aquaro M. , N. Bailey, and M. H. Pesaran (2021), Estimation and Inference for Spatial Models with Heterogeneous Coefficients: An Application to U.S. House Prices, Journal of Applied Econometrics. https://doi.org/10.1002/jae.2792

Bailey, N. G. Kapetanios, and M.H. Pesaran (2021), Measurement of Factor Strength: Theory and Practice, Journal of Applied Econometrics. https://doi.org/10.1002/jae.2830

Buhlmann, P. and S. van de Geer (2011), Statistics for High-Dimensional Data: Methods, Theory and Applications, Springer Series in Statistics. https://doi.org/10.1007/978-3-642-20192-9

*Chudik, A., K. Mohaddes, M. H. Pesaran, M. Raissi, and A. Rebucci (2021) A counterfactual economic analysis of Covid-19 using a threshold augmented multi-country model, Journal of International Money and Finance. https://doi.org/10.1016/j.jimonfin.2021.102477.

Chudik, A., G. Kapetanios, and M. H. Pesaran (2018), A One Covariate at a Time, Multiple Testing Approach to Variable Selection in High-Dimensional Linear Regression Models, Econometrica, 86, 1479-1512. https://doi.org/10.3982/ECTA14176

Chudik, A., M.H. Pesaran and M. Sharifvaghefi (2024), Variable Selection in High Dimensional Linear Regressions with Parameter Instability, Journal of Econometrics, 246, published online, also https://arxiv.org/abs/2312.15494

Hastie, T., R Tibshirani and J Friedman (2009) The Elements of Statistical Learning - Data Mining, Inference & Prediction, Second Edition, Springer Series in Statistics.

*Koch I.(2014) *Analysis of Multivariate and High-Dimensional Data*. Cambridge University Press, Cambridge. https://doi.org/10.1017/CBO9781139025805 Chapters 1,2, 6 and 7.

*Pesaran, M. H. and R. Smith (2024) High-dimensional forecasting with known knowns and known unknowns. National Institute Economic Review, 2024;267:1-25. doi:10.1017/nie.2024.1, also at https://arxiv.org/abs/2401.14582

Varian, H. (2014) Big Data: new tricks for econometrics, Journal of Economic Perspectives, 28, 3-28. https://www.aeaweb.org/articles?id=10.1257/jep.28.2.3

*Wainwright M.J. (2019) *High-Dimensional Statistics: A Non-Asymptotic Viewpoint*. Cambridge University Press, Cambridge. https://doi.org/10.1017/9781108627771 Chapters 7 and 8.