

Project Summary

Prof. Alexei Onatski and Dr. Chen Wang - High-dimensional Cointegration Analysis (JHOY).

Summary of work and results

Most economic data are nonstationary. The major share of their variation often comes from stochastic trends, slowly varying persistent movements that reflect deep economic changes, such as changes in preferences and technology. Analysis of such trends is indispensable for long-run economic prediction and modelling. Its importance has been recently reemphasized in Müller and Watson (2018).

The availability of a large amount of economic and financial data brings new opportunities for the analysis of stochastic trends. For example, Engel et al. (2015) propose to use factors extracted from dozens of exchange rates to improve forecasting of individual exchange rates. Banerjee et al. (2017) use information from a large nonstationary macroeconomic dataset for the identification of structural shocks and their propagation mechanisms.

However, the large nonstationary data bring new theoretical and methodological challenges. One of them is that the standard cointegration analysis, based on the assumption of the fixed small number of studied series, breaks down. Furthermore, the factor analysis of large nonstationary panels may be spurious.

In this project we focus on two issues. First, we develop a new cointegration test that is robust to high dimensionality. Second, we analyze in detail the phenomenon of the spurious factor analysis.

Specifically, we first study the likelihood ratio (LR) statistic for testing no cointegration in high-dimensional vector autoregressions. It has the form of a linear spectral statistic of a matrix $C'ACB$, where A is a sample covariance matrix of high-dimensional random walk, B is a sample covariance matrix of the random walk's innovations, and C is the sample cross-covariance between the random walk and its own innovations. We show that linear spectral statistics for $C'ACB$ are asymptotically normal, and derive formulae for the corresponding asymptotic mean and variance. The formulae can be used to quickly obtain critical values of the LR test of no cointegration in high dimensions from the standard normal tables. This test substantially improves over the standard Bartlett-corrected LR tests based on complicated low-dimensional asymptotics.

Next, we draw parallels between the Principal Components Analysis of factorless high-dimensional nonstationary data and the classical spurious regression. We show that a few of the principal components of such data absorb nearly all the data variation. The corresponding scree plot suggests that the data contain a few factors, which is collaborated by the

standard panel information criteria. Furthermore, the Dickey-Fuller tests of the unit root hypothesis applied to the estimated “idiosyncratic terms” often reject, creating an impression that a few factors are responsible for most of the non-stationarity in the data. We warn empirical researchers of these peculiar effects and suggest to always compare the analysis in levels with that in differences.

Impact and Outputs

The work on the outputs and dissemination is ongoing. The project promoted our collaboration with Iain Johnstone from Stanford University and Yegor Klochkov from Humboldt University on statistics of high-dimensional data.

Dissemination

Eleven presentations of preliminary work:

1. Econometrics seminar, Singapore Management University (April 2019)
2. Econometrics seminar, National University of Singapore (April 2019)
3. Econometrics seminar, Hong Kong University of Science and Technology (April 2019)
4. Big Data Methods in Econometrics and Finance, INET conference, Cambridge (May 2019)
5. 6th RCEA Time Series Econometrics Workshop, invited keynote talk, Cyprus (June 2019)
6. 32nd European Meeting of Statisticians, invited talk, Palermo, Italy (July 2019)
7. Joint Statistical Meetings, 2019. IMS-sponsored invited session – Random matrices and high dimensional statistics. Denver, Colorado, USA (July-August 2019)
8. Econometrics seminar, Harvard-MIT (September 2019)
9. Statistics seminar, Weierstrass Institute, Berlin (October 2019)
10. Econometrics seminar, University of Pennsylvania (November 2019)
11. Econometrics seminar, Princeton University (November 2019)

Planned academic outputs

1. Onatski, A. and Wang, C. “Spurious Factor Analysis,” Revise and Resubmit in *Econometrica*.
2. Onatski, A. and Wang, C. “Testing high-dimensional cointegration”
We are in the process of finishing the first draft of this paper. We plan to submit it to a top econometrics or statistics journal.

Any possible future plans

One project that would be a natural continuation of our “Spurious Factor Analysis” paper is to develop a test for the number of factors in large dimensional stationary data based on the comparison of the factors extracted from filtered data and filtered factors extracted from the original data, where the same filter is used in both cases. We hope that such a test would provide a very powerful technique for deciding on how many factors to extract from various macroeconomic and financial datasets.

References

1. Banerjee, A., Marcellino, M., and Masten, I. (2017) “Structural FECM: Cointegration in large-scale structural FAVAR models,” *Journal of Applied Econometrics* 32, 1069-1086.
2. Engel, C., N.C. Mark, and K.D. West (2015) “Factor Model Forecasts of Exchange Rates,” *Econometric Reviews* 34, 32-55.
3. Müller, U. K. and Watson, M. W. (2018) “Long-Run Covariability,” *Econometrica* 86, 775-804.