Workshop on Data Science 2015
Aachen

October 2015

Speaker: Ansgar Steland
Institute of Statistics,
RWTH Aachen University

Title: Large Sample Approximations for Variance–Covariance Matrices of High–Dimensional Time Series and Applications

Abstract:

High–dimensional time series consisting of n observations of a large number, d_n, of variables (features), thus forming a high–dimensional vector time series, are ubiquitous in data science. A major issue is the analysis of the dependencies of the observed variables, especially for the case that d_n is large compared to n which is the typical situation when analysing massive data.

In this talk I present new results providing large sample distributional approximations for bilinear forms of the sample variance-covariance matrix and a couple of key applications. Those approximations aim at providing a sound basis for future construction of inferential procedures to analyze dependencies in high–dimensional time series by means of statistically valid inferential procedures. Our approach particularly allows us to handle the case of projections, which is a commonly used approach to deal with high–dimensional data.

Within a high-dimensional time series model that allows for full covariance matrices, we propose novel large sample approximations for bilinear forms of the sample variance-covariance matrix. The results cover weakly as well as many long–range dependent linear processes and are valid for a large class of projection vectors that arise, naturally or by construction, in many problems extensively studied for high–dimensional vector time series.

The results can be directly applied to a large number of key applications such as financial portfolio optimization, sparse principal component analysis and shrinkage estimation. For sparse financial portfolio optimization, [2] proposed to construct regularized portfolio vectors. In order to conduct sparse principal component analysis, there are several recent proposals to construct ℓ_1-bounded components such as the SCotLASS (simplified component technique-lasso) approach of [4], the penalized matrix decomposition problem (PMD) studied by [8] or the lassoed principal components (LPC) method of [**?**]. Our results are also directly applicable to the problem of shrinkage estimation ([5], [6], [3]), which is widely used alternative to approaches such as banding or tapering, see [1].

Acknowledgement:

This is joint work with Rainer v. Sachs, Universite catholique Louvain, Belgium.

References

[1] Bickel, P.J. and Levina, E. (2008). Covariance regularization by thresholding. *Annals of Statistics*, 36(6): 2577–2604.

[2] Brodie, J., Daubechies, C., De Mol, C., Giannne, D. and Loris, I. (2009). Sparse and stable Markowitz portfolios. *Proceedings of the National Academy of Sciences of the U.S.A.* (PNAS), 106(30): 12267–12272.

[3] Boehm, H. and von Sachs, R. (2009). Shrinkage estimation in the frequency domain of multivariate time series. *Journal of Multivariate Analysis*, 100(5): 913–935.

[4] Jolliffe, L., Trendafilov, N. and Uddin, M. (2003). A modified principal component technique based on the lasso. *Journal of Computational and Graphical Statistics*, 12: 531–547.

[5] Ledoit, O. and Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, 88(2): 365–411.

[6] Sancetta, A. (2008). Sample covariance shrinkage for high-dimensional dependent data. *Journal of Multivariate Analysis*, 99(5): 949–967.

[7] Steland, A. and von Sachs, R. (2015). Large sample approximations for variance-covariance matrices of high-dimensional time series, *revised version submitted.*

[8] Witten, D.M., Tibshirani, R. and Hastie, T. (2009). A penalized decomposition, with applications to sparse principal components and canonical correlation analysis. *Biostatistics* 10: 515–534.