

# Essay in High-Dimensional Econometrics and Finance

Vladimir Pyrlík

## Abstract

High dimensionality is a popular contemporary setting in applied statistical analysis of various types of data. Growing dimensionality of the data challenges precise and well-conditioned estimates of statistical models. We address selected high-dimensional methods and their applications to financial market analysis. First, we consider modeling flexible joint distributions using copulas. In Chapter 1, we address estimation of Gaussian and  $t$  copulas in ultra-high dimensions, up to thousands of variables that use up to 30 times shorter sample lengths. We employ large covariance matrix shrinkage tools to obtain precise and well-conditioned estimates of the matrix parameters of the copulas. In Chapter 2, we present a new method for estimating the skew- $t$  copula, known for its advantageous properties in characterizing joint distributions, including asymmetry, heavy tails, and asymmetric tail dependence. Our approach involves a two-step procedure based on the simulated method of moments and analytical non-linear shrinkage estimator for large covariance matrices. In both chapters, we also illustrate the benefits of the copula approach in a large stock portfolio allocation. Our analysis shows that copula-based models deliver better portfolios in terms of cumulative returns and maximum downfalls over the portfolio lifetime than the benchmark alternatives. In Chapter 3, we step away from unconditional distribution modeling, and assess the performance of selected machine learning algorithms in forecasting daily realized volatility. We utilize high-dimensional settings and mixed-frequency data set-ups to improve forecasts of selected stocks returns volatility in the Russian stock market in 2018-2020.

## Keywords

high dimensionality, large covariance matrices, shrinkage, Gaussian copula,  $t$  copula, multivariate skew- $t$  distribution, skew- $t$  copula, simulated method of moments, portfolio allocation, stock market, EUROSTOXX50, heterogeneous autoregressive model, machine learning, lasso, gradient boosting, random forest, long short-term memory, realized volatility, Russian stock market, mixed-frequency data