Copula Shrinkage and Portfolio Allocation in Ultra-High Dimensions

Stanislav Anatolyev *[1] and Vladimir Pyrlik †[1]

[1]CERGE-EI.

Abstract

Many problems in finance, in particular the problem of (many) asset allocation, require use of (high-dimensional) multivariate distributions. Copulas are a convenient framework to synthesize joint distributions, especially in high dimensions. Currently, copula-based high dimensional settings are used for as many as a few hundred variables and require large data samples for estimation to be precise. Until recently, the dimensionality of data in empirical applications rarely had exceeded a few dozen variables, with only several settings with more than a hundred variables. Most studies focus on sparse structures that are identified heuristically from the data, which is an excessively strong assumption.

In this paper, we focus on elliptical copulas in high dimensions, specifically, two most commonly used in modeling and practical applications: Gaussian and t copulas, which are used in a variety of finance applications. Most often, they are applied to model joint distributions of financial assets or indices returns for the task of portfolio allocation, but also in studies of tail dependence and asset pricing. While most settings based on the Gaussian and t copulas are low-dimensional, where the number of dimensions varies from two to a few dozen, and the ratio to corresponding sample sizes is considerably less than unity, some settings are high-dimensional with the ratio reaching five. In the case of Gaussian and t copulas, the dimensionality of the parameter space is directly connected to the data dimensionality, with the matrix parameter naturally interpretable in the description of the degree of pairwise dependence among the variables. In low dimensions, copulas are effectively estimated via computationally very practical method-of-moments-like techniques based on rank correlations and sample correlation matrices. However, in high dimensions the settings and their estimates inherit the same problems as the traditional covariance matrix estimators.

Recently, a substantial amount of research has focused on developing covariance matrix estimators that are robust to and well-conditioned under the data dimensionality growing along with the sample size. One approach to solving the problem is to adjust the traditional sample covariance matrix by directly restricting its structure, eigenvalues or the inverse to achieve better properties under moderate or high data dimensionality. This machinery has been developed in a series of papers of Olivier Ledoit and Michael Wolf, whose estimator relies on the random

*E-mail address: stanislav.anatolyev@cerge-ei.cz

†E-mail address: vladimir.pyrlik@cerge-ei.cz

matrix theory and leads to fast and relatively easy estimation of large covariance matrices of dimensionality higher than had been feasible ever before.

These advances in large covariance matrix estimation rather conveniently match with the structure of Gaussian and t copulas. An important property of these copulas is that their matrix parameter is very close the correlation matrix of pseudo-observations, which allows one to use the shrinkage estimators to estimate the matrix parameters of Gaussian and t copulas in high dimensional datasets. In particular, we consider datasets with up to thousands of variables that use up to 30 times lower sample sizes. Thus, we take the data dimensionality well beyond what is studied in the copula literature; hence the prefix "ultra-" in "high dimensions" in the title.

In a simulation study, we compare the quality of performance of different estimators for various ratios of data dimensionality to sample size. We show that the shrinkage estimators significantly outperform the traditional copula matrix parameter estimators, as measured in terms of both the closeness of estimated parameter values to their actual values and the closeness of the entire estimated copula function to its true counterpart.

We apply shrinkage-based estimators of copula correlation matrices in high dimensions to a large portfolio allocation problem and compare emerging portfolios to those from a multivariate normal model and copula models based on traditional estimators. Using daily data on prices of over 3600 U.S. stocks, we construct portfolios of up to 3600 assets and simulate buy-and-hold portfolio strategies. The joint distributional models of asset returns are estimated over the period of six months (120 observations), hence the problem is ultra-high dimensional, with the dimensionality ratio of 30. The comparison of the portfolios based on different models to equally weighted portfolios shows that the shrinkage-based estimators applied to t copula based models of return distribution deliver better portfolios in terms of both cumulative return and maximum downfall over the portfolio lifetime than the corresponding portfolios derived from the multivariate normal or copula-based models estimated via traditional estimators.

Keywords: Portfolio allocation, elliptical copulas, high dimensionality, shrinkage.

Acknowledgements: This research was supported by the grant 20-28055S from the Czech Science Foundation.