Available online at www.sciencedirect.com
SCIENCE DIRECT®
Physica A 343 (2004) 623–634
PHYSICA A
www.elsevier.com/locate/physa

Estimated correlation matrices and portfolio optimization

Szilárd Pafka[a,b,*], Imre Kondor[a,c]

[a] *Department of Physics of Complex Systems, Eötvös University, Pázmány P. sétány 1/a, H-1117 Budapest, Hungary*
[b] *Risk Management Department, CIB Bank, Medve u. 4-14., H-1027 Budapest, Hungary*
[c] *Collegium Budapest - Institute for Advanced Study, Szentháromság u. 2., H-1014 Budapest, Hungary*

Received 16 January 2004

Available online 21 June 2004

Abstract

Correlations of returns on various assets play a central role in financial theory and also in many practical applications. From a theoretical point of view, the main interest lies in the proper description of the structure and dynamics of correlations, whereas for the practitioner the emphasis is on the ability of the models to provide adequate inputs for the numerous portfolio and risk management procedures used in the financial industry. The theory of portfolios, initiated by Markowitz, has suffered from the "curse of dimensions" from the very outset. Over the past decades a large number of different techniques have been developed to tackle this problem and reduce the effective dimension of large bank portfolios, but the efficiency and reliability of these procedures are extremely hard to assess or compare. In this paper, we propose a model (simulation)-based approach which can be used for the systematical testing of all these dimensional reduction techniques. To illustrate the usefulness of our framework, we develop several toy models that display some of the main characteristic features of empirical correlations and generate artificial time series from them. Then, we regard these time series as empirical data and reconstruct the corresponding correlation matrices which will inevitably contain a certain amount of noise, due to the finiteness of the time series. Next, we apply several correlation matrix estimators and dimension reduction techniques introduced in the literature and/or applied in practice. As in our artificial world the only source of error is the finite length of the time series and, in addition, the "true" model, hence also the "true" correlation matrix, are precisely known, therefore in sharp contrast with empirical studies, we can precisely compare the performance of the various noise reduction techniques. One of our recurrent observations is that the recently

* Corresponding author. Department of Physics of Complex Systems, Eötvös University, Pázmány P. sétány 1/a, H-1117 Budapest, Hungary.
E-mail addresses: syl@complex.elte.hu (S. Pafka), kondor@colbud.hu (I. Kondor).

0378-4371/$ - see front matter
doi:10.1016/j.physa.2004.05.079

introduced filtering technique based on random matrix theory performs consistently well in all the investigated cases. Based on this experience, we believe that our simulation-based approach can also be useful for the systematic investigation of several related problems of current interest in finance.

PACS: 87.23.Ge; 05.45.Tp; 05.40.−a

Keywords: Estimated covariance matrices; Estimation noise; Noise filtering; Random matrix theory; Portfolio optimization; Risk management

1. Introduction

Correlation matrices of financial returns play a crucial role in several branches of modern finance such as investment theory, capital allocation and risk management. For example, financial correlation matrices are the key input parameters to Markowitz's classical portfolio optimization problem [1], which aims at providing a recipe for the selection of a portfolio of assets so that risk (quantified by the standard deviation of the portfolio's return) is minimized for a given level of expected return. For any practical use of the theory it would therefore be necessary to have reliable estimates for the correlations of returns (of the assets making up the portfolio), which are usually obtained from historical return series data. However, if one estimates a $n \times n$ correlation matrix from n time series of length T each, with T bounded for evident practical reasons, one inevitably introduces estimation error, which for large n can become so overwhelming that the whole applicability of the theory becomes questionable.

This difficulty has been well known by economists for a long time (see e.g. [2] and the numerous references therein). Several aspects of the effect of noise (in the correlation matrices determined from empirical data) on the classical portfolio selection problem have been investigated e.g. in Ref. [3]. One way to cope with the problem of noise is to impose some structure on the correlation matrix, which may certainly introduce some bias in the estimation, but by effectively reducing the dimensionality of the problem, could, in fact, be expected to improve the overall performance. Such a best-known structure is that imposed by the single-index (or market) model, which has stimulated strong interest in the academic literature (see e.g. Ref. [2] for an overview and references) and has also become widely used in the financial industry (the coefficient "beta", relating the returns of an asset to the returns of the corresponding wide market index, has long been a widespread tool in the financial community). On economic or statistical grounds, several other correlation structures have been experimented with in the academic literature and financial industry, for example multi-index models, grouping by industry sectors, macroeconomic factor models, models based on principal component analysis, etc. Several studies (see e.g. Ref. [4]) attempt to compare the performance of these correlation estimation procedures as input providers for the portfolio selection problem, although all these studies have been restricted to the use of given specific empirical samples. More recently, additional procedures to impose