Available online at www.sciencedirect.com
SCIENCE DIRECT®
Physica A 335 (2004) 629–643
PHYSICA A
www.elsevier.com/locate/physa

Random matrix theory for portfolio optimization: a stability approach

S. Sharifi, M. Crane, A. Shamaie, H. Ruskin*

Dublin City University, Glasnevin, Dublin 9, Ireland

Received 4 July 2003

Abstract

We apply random matrix theory (RMT) to an empirically measured financial correlation matrix, C, and show that this matrix contains a large amount of noise. In order to determine the sensitivity of the spectral properties of a random matrix to noise, we simulate a set of data and add different volumes of random noise. Having ascertained that the eigenspectrum is independent of the standard deviation of added noise, we use RMT to determine the noise percentage in a correlation matrix based on real data from S&P500. Eigenvalue and eigenvector analyses are applied and the experimental results for each of them are presented to identify qualitatively and quantitatively different spectral properties of the empirical correlation matrix to a random counterpart. Finally, we attempt to separate the noisy part from the non-noisy part of C. We apply an existing technique to cleaning C and then discuss its associated problems. We propose a technique of filtering C that has many advantages, from the stability point of view, over the existing method of cleaning.

Keywords: Random matrix theory; Portfolio optimization; Correlation matrix; Eigenvalues and eigenvectors

1. Introduction

Random matrix theory (RMT), originally developed for use in nuclear physics, has been described by, among others, Dyson in a series of papers beginning with Dyson [1] and subsequently in collaboration with Mehta (Mehta and Dyson [2], Dyson and Mehta [3], Mehta [4]) as the matrix representation of the average of all possible interactions in a nucleus. It can be used to identify non-random properties which are

* Corresponding author.
E-mail addresses: ssaba@computing.dcu.ie (S. Sharifi), mcrane@computing.dcu.ie (M. Crane), ashamaie@computing.dcu.ie (A. Shamaie), hruskin@computing.dcu.ie (H. Ruskin).

0378-4371/$ - see front matter
doi:10.1016/j.physa.2003.12.016

deviations from the universal predictions of RMT; properties that are specific to the considered system. Close agreement between the distribution of the eigenvalues of a matrix M, with those from a matrix made up of random entries implies that M has entries that contain a considerable degree of randomness as has been shown in the literature [5–8]. This matrix consisting of random elements with unit variance and zero mean is called a random matrix [4]. In the case of a correlation matrix C, the level of agreement between its eigenvalue distribution and those of a random matrix, represents the amount of randomness (or noise) in C and thus, deviations from RMT represent genuine correlation (cf. [5–7]). This is precisely the problem that we wish to address, i.e., the identification of the true information (correlated assets) in a noisy financial correlation matrix. The method tests the null hypothesis that the distribution of eigenvalues of the correlation matrix is random. Since the correlation matrix is symmetric, the random matrix, with which it is compared, should also be symmetric [8].

Before applying the cleaning method to real correlation matrices, we need to determine the role of the *amount* of random noise on the spectral properties of a random matrix. This is done by examining the difference between the eigenspectrum of a correlation matrix made up of simulated data with different amounts of random noise and that of a random matrix. We can then proceed with confidence to examine the stability of real correlation matrices using real data. The empirical data set we use consists of 30-min intraday prices from the S&P500 Index from the beginning of April 1997 to the beginning of April 1999. This provides about 1500 data points for about 450 companies.

In this paper our initial objective, therefore, is to separate the noisy part from non-noisy part in C. Removal of the noise makes the optimization process more reliable, leaving the analyst in a better position to estimate the risk associated with the constructed portfolio. However, the techniques for removing noise from C should be considered carefully. A standard technique is initially applied to clean C but assessment of the results achieved reveal that it is not particularly satisfactory on the grounds of stability. We therefore, go on to discuss a filtering technique that takes account of the stability in a more precise way. Advantages of the new approach are validated by application to a financial data set from the S&P500.

2. Background and theory

2.1. Introduction: applications of RMT

Recently, a growing number of physicists have attempted to analyse and model financial markets. The history of this interest goes back to the work of Majorana [9] on analogies between statistical laws in physics and in the social sciences. In nuclear physics, the problem of understanding properties of matrices with random entries (which attempt to describe the interactions between sub-atomic particles) has a rich history (e.g. Wigner [10–12], Dyson [1], Mehta and Dyson [2], Dyson and Mehta [3] and Mehta [4]); in these, the assumption is made that the interactions between nuclear components are so complex that they can be taken as random. In