Available online at www.sciencedirect.com
ScienceDirect
ELSEVIER Journal of Economic Dynamics & Control 32 (2008) 235–258
JOURNAL OF Economic Dynamics & Control
www.elsevier.com/locate/jedc

Cluster analysis for portfolio optimization

Vincenzo Tola[a,b,1], Fabrizio Lillo[c,d,e], Mauro Gallegati[a], Rosario N. Mantegna[c,d,*]

[a]*Dipartimento di Economia, Università Politecnica delle Marche, Piazza Martelli 8, I-60121 Ancona, Italy*
[b]*Banca d'Italia, Servizio Vigilanza sugli Enti Creditizi, Via Piacenza 6, 00184 Rome, Italy*
[c]*Dipartimento di Fisica e Tecnologie Relative, Università di Palermo, viale delle Scienze, I-90128 Palermo, Italy*
[d]*INFN, Sezione di Catania, Catania, Italy*
[e]*Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501, USA*

Received 1 May 2005; received in revised form 1 December 2006; accepted 26 January 2007
Available online 2 April 2007

Abstract

We consider the problem of the statistical uncertainty of the correlation matrix in the optimization of a financial portfolio. By assuming idealized conditions of perfect forecast ability for the future return and volatility of stocks and short selling, we show that the use of clustering algorithms can improve the reliability of the portfolio in terms of the ratio between predicted and realized risk. Bootstrap analysis indicates that this improvement is obtained in a wide range of the parameters N (number of assets) and T (investment horizon). The predicted and realized risk level and the relative portfolio composition of the selected portfolio for a given value of the portfolio return are also investigated for each considered filtering method. We also show that several of the

*Corresponding author. Dipartimento di Fisica e Tecnologie Relative, Universita degli studi di Palermo, Viale delle Scienze, 90128 Palermo, Italy. Tel.: +39 91 661 5074; fax: +39 91 661 5063.
E-mail address: mantegna@unipa.it (R.N. Mantegna).

[1]Opinions expressed in this paper are exclusively by the authors and do not necessarily reflect those of Banca d'Italia.

doi:10.1016/j.jedc.2007.01.034

results obtained by assuming idealized conditions are still observed under the more realistic assumptions of no short selling and mean return and volatility forecasting based on historical data.

JEL classification: G11; C30; C15

Keywords: Portfolio optimization; Clustering methods; Correlation matrices; Random matrix theory

1. Introduction

The problem of portfolio optimization is one of the most important issues in asset management (Elton and Gruber, 1995). Since the seminal work of Markowitz (1959), which solved the problem under a certain number of simplifying assumptions (see also Section 2), many other studies have been devoted to consider several aspects of portfolio optimization both from a theoretical and from an applied point of view. A huge number of studies considering key aspects of portfolio optimization theory are present in the finance literature. Here we refer to a few studies considering, for example, the real performance of portfolios constructed using sample moments (Jorion, 1985), the realized Sharpe ratio of the global minimum variance portfolio (Jagannathan and Ma, 2003), the role of constraints in portfolio optimization (Eichhorn et al., 1998; Jagannathan and Ma, 2003), the shrinkage estimator of large dimensional covariance matrices (Ledoit and Wolf, 2004a,b). The aim of the present study is to focus on the role of the correlation coefficient matrix in portfolio optimization. The estimation of the correlation matrix is unavoidably associated with a statistical uncertainty, which is due to the finite length of the asset return time series. Recently, there have been several contributions in the econophysics literature devoted to quantify the degree of statistical uncertainty present in a correlation matrix. The results of these investigations have been obtained by using concepts and tools of random matrix theory (RMT) (Metha, 1990). The RMT quantification of the statistical uncertainty associated with the estimation of the correlation coefficient matrix of a finite multivariate time series has been recently used to devise a procedure to filter the information present in the correlation coefficient matrix which is robust with respect to the unavoidable statistical uncertainty (in the econophysics literature the term of noise dressing has been used) (Galluccio et al., 1998; Laloux et al., 1999, 2000; Plerou et al., 1999, 2002; Gopikrishnan et al., 2001; Drozdz et al., 2001; Rosenow et al., 2002; Pafka and Kondor, 2003, 2004; Rosenow et al., 2003; Guhr and Kalber, 2003; Malevergne and Sornette, 2004; Sharifi et al., 2004; Burda and Jurkiewicz, 2004). The correlation matrices obtained by this filtering procedure has been used in portfolio optimization in some studies (Laloux et al., 2000; Rosenow et al., 2002), which have shown that under the assumption of perfect forecasting of future returns and volatilities the distance