ELSEVIER

Random matrix theory and fund of funds portfolio optimisation

T. Conlon*, H.J. Ruskin, M. Crane

Dublin City University, Glasnevin, Dublin 9, Ireland

Abstract

The proprietary nature of Hedge Fund investing means that it is common practise for managers to release minimal information about their returns. The construction of a fund of hedge funds portfolio requires a correlation matrix which often has to be estimated using a relatively small sample of monthly returns data which induces noise. In this paper, random matrix theory (RMT) is applied to a cross-correlation matrix C, constructed using hedge fund returns data. The analysis reveals a number of eigenvalues that deviate from the spectrum suggested by RMT. The components of the deviating eigenvectors are found to correspond to distinct *groups of strategies* that are applied by hedge fund managers. The inverse participation ratio is used to quantify the number of components that participate in each eigenvector. Finally, the correlation matrix is cleaned by separating the noisy part from the non-noisy part of C. This technique is found to greatly reduce the difference between the predicted and realised risk of a portfolio, leading to an improved risk profile for a fund of hedge funds.

1. Introduction

A hedge fund is a lightly regulated private investment vehicle that may utilise a wide range of investment strategies and instruments. These funds may use short positions, derivatives, leverage and charge incentive-based fees. Normally, they are structured as limited partnerships or offshore investment companies. Hedge funds pursue positive returns in all markets and hence are described as "absolute return" strategies.

Hedge funds are utilised by pension funds, high net-worth individuals and institutions, due to their low correlation to traditional long-only investment strategies. The incentive-based performance fees, earned by hedge fund managers, align the interest of the hedge fund manager with that of the investor. The performance of hedge funds has been impressive, with the various hedge fund indices providing higher returns, with lower volatility, than traditional assets over many years. As of the end of the first quarter 2006 the total assets managed by hedge funds world wide is estimated at $1.25 trillion [1]. Hedge funds generally only report their

---

*Corresponding author.

E-mail addresses: tconlon@computing.dcu.ie (T. Conlon), hruskin@computing.dcu.ie (H.J. Ruskin), mcrane@computing.dcu.ie (M. Crane).

returns on a monthly basis and this means that there is a very limited amount of data available to study as databases of hedge fund returns have only been in operation for about 15 years. This is in keeping with the highly secretive, proprietary nature of hedge fund investing. The amount of information reported by a hedge fund about how and where it is producing its returns is often limited to sectoral overviews and strategy allocations. For an introduction to hedge funds see Refs. [2,3].

Significant diversification benefits can be gained by investing in a variety of hedge fund strategies, due to the presence of low and even negative correlations between different hedge fund strategies. Such strategies can be broken up into two general categories: directional and market neutral. Directional strategies, (for example long/short equity, emerging markets, macro and managed futures) have a high risk, high return profile and act as return enhancers to a traditional portfolio. Market neutral strategies, (for example convertible arbitrage, equity market neutral and fixed income arbitrage) deploy a low risk profile and act as a substitute for some proportion of the fixed income holdings in an investors portfolio [2,3].

A fund of hedge funds allows investors to have access to a large diverse portfolio of hedge funds without having to carry out due diligence on each individual manager. The diversification benefits provided by fund of funds are brought about by investing in a number of funds that have a low correlation to each other. These correlations are often calculated by using equally weighted fund returns and can contain a significant amount of noise due to the very small amount of returns data available for hedge funds [3].

In this paper we apply random matrix theory (RMT) to hedge fund returns data with the aim of reducing the levels of noise in these correlation matrices formed from this data and hence constructing a fund of hedge funds with an improved risk profile. Previous studies have used the information found in the RMT defined deviating eigenvalues of a correlation matrix as inputs into a minimum spanning tree [4] to enable characterisation of hedge fund strategies. In this paper, the components of the deviating eigenvectors are shown to correspond to distinct groups of strategies that are applied by hedge fund managers and this is exploited to construct a portfolio with reduced levels of risk.

This paper is organised as follows: in Section 2 we review RMT and discuss its use in the extraction of information from a correlation matrix of hedge fund returns using RMT techniques. In Section 3, we look at the results obtained applying RMT to hedge funds and, in the final section, we draw our conclusions.

2. Methods

2.1. Random matrix theory

Given returns $G_i(t)$, $i = 1, \ldots, N$, of a collection of hedge funds we define a normalised return in order to standardise the different fund volatilities. We normalise G_i with respect to its standard deviation σ_i as follows:

$$g_i(t) = \frac{G_i(t) - \widehat{G_i(t)}}{\sigma_i}, \tag{1}$$

where σ_i is the standard deviation of G_i for assets $i = 1, \ldots, N$ and $\widehat{G_i}$ is the time average of G_i over the period studied.

Then the equal time cross-correlation matrix is expressed in terms of $g_i(t)$

$$C_{ij} \equiv \langle g_i(t) g_j(t) \rangle. \tag{2}$$

The elements of C_{ij} are limited to the domain $-1 \leqslant C_{ij} \leqslant 1$, where $C_{ij} = 1$ defines perfect correlation between funds, $C_{ij} = -1$ corresponds to perfect anti-correlation and $C_{ij} = 0$ corresponds to uncorrelated funds. In matrix notation, the correlation matrix can be expressed as

$$\mathbf{C} = \frac{1}{T} \mathbf{G} \mathbf{G}^t, \tag{3}$$

where $\mathbf{G}$ is an $N \times T$ matrix with elements g_{it}.

The spectral properties of $\mathbf{C}$ may be compared to those of a "random" Wishart correlation matrix [8,11],

$$\mathbf{R} = \frac{1}{T}\mathbf{A}\mathbf{A}^t, \tag{4}$$

where $\mathbf{A}$ is an $N \times T$ matrix with each element random with zero mean and unit variance. Statistical properties of random matrices have been known for many years in the physics literature [5] and have been applied to financial problems relatively recently [6–16].

In particular, the limiting property for the sample size $N \to \infty$ and sample length $T \to \infty$, providing that $Q = T/N \geqslant 1$ is fixed, has been examined to show analytically that the distribution of eigenvalues λ of the random correlation matrix $\mathbf{R}$ is given by

$$P_{rm}(\lambda) = \frac{Q}{2\pi\sigma^2}\frac{\sqrt{(\lambda_+ - \lambda)(\lambda - \lambda_-)}}{\lambda}, \tag{5}$$

for λ within the region $\lambda_- \leqslant \lambda_i \leqslant \lambda_+$, where λ_- and λ_+ are given by

$$\lambda_{\pm} = \sigma^2\left(1 + \frac{1}{Q} \pm 2\sqrt{\frac{1}{Q}}\right), \tag{6}$$

where σ^2 is the variance of the elements of $\mathbf{G}$ (for $\mathbf{G}$ normalised this is equal to unity).

$\lambda_{\pm}$ are the bounds of the theoretical eigenvalue distribution. Eigenvalues that are outside this region are said to deviate from RMT. Hence by comparing the empirical distribution of the eigenvalues of the funds correlation matrix to the distribution for a random matrix as given in Eq. (5), we can identify the deviating eigenvalues. These deviating eigenvalues are said to contain information about the system under consideration and we use eigenvector analysis to identify specifically the information present.

2.2. Eigenvector analysis

The deviations of $P(\lambda)$ from the RMT result $P_{rm}(\lambda)$ implies that these deviations should also be displayed in the statistics of the corresponding eigenvector components [8]. In order to interpret the meaning of the deviating eigenvectors, we note that the largest eigenvalue is of an order of magnitude larger than the others, which constrains the remaining $N - 1$ eigenvalues since $\text{Tr}[C] = N$. Thus, in order to analyse the contents of the remaining eigenvectors, we first remove the effect of the largest eigenvalue. To do this we use the linear regression (as discussed in Ref. [11])

$$G_i(t) = \alpha_i + \beta_i G^{large}(t) + \varepsilon_i(t), \tag{7}$$

where $G^{large} = \sum_1^N u_i^{large} G_i(t)$ and $N = 49$ is the number of funds in our sample. Here u_i^{large} corresponds to the components of the largest eigenvector. We then recalculate the correlation matrix C using the residuals $\varepsilon_i(t)$. If we quantify the variance of the part not explained by the largest eigenvalue as $\sigma^2 = 1 - \lambda_{large}/n$, [8], we can use this value to recalculate our values of $\lambda_{\pm}$.

Using techniques for sector identification, [10], we try to analyse the information contained in the eigenvectors. We partition the funds into groups $l = 1, \ldots, 10$ as defined by the managers strategy. (e.g., equity long/short, managed futures. See Appendix A for a complete strategy breakdown for the sample.) We define a projection matrix $P_{li} = 1/n_l$, if fund i belongs to group l and $P_{li} = 0$ otherwise. For each deviating eigenvector u^k we compute the contribution $X_l^k \equiv \sum_{i=1}^N P_{li}^k[u_i^k]^2$ of each strategy group, where this represents the product of the projection matrix and the square of the eigenvector components. This allows us to measure the contribution of the different hedge fund strategies to each of the eigenvectors.

As suggested in Ref. [11], we also aim to assess how the effects of randomness lessen as we move further from the RMT upper boundary, λ_+. To do this we use the inverse participation ratio (IPR). The IPR allows quantification of the number of components that participate significantly in each eigenvector and tells us more

about the level and nature of deviation from RMT. The IPR of the eigenvector u^k is given by $I^k \equiv \sum_{l=1}^{N} (u_l^k)^4$ and allows us to compute the inverse of the number of eigenvector components that contribute significantly to each eigenvector.

2.3. Application to portfolio optimisation

The diversification of an investment into independently fluctuating assets reduces its risk. However, since cross-correlations between asset prices exist, the accurate calculation of the cross-correlation matrix is vitally important. The return on a portfolio with N assets is given by $\Phi = \sum_{i=1}^{N} w_i G_i$ where $G_i(t)$ is the return on asset i, w_i is the fraction of wealth invested in asset i and $\sum_{i=1}^{N} w_i = 1$. The risk of holding this portfolio is then given by

$$\Omega^2 = \sum_{i=1}^{N} \sum_{j=1}^{N} w_i w_j C_{ij} \sigma_i \sigma_j, \tag{8}$$

where σ_i is the volatility of G_i and C_{ij} are the elements of the cross-correlation matrix. In order to find an optimal portfolio, using the Markowitz theory of portfolio optimisation [6,17,18], we minimise Ω^2 under the constraint that the return on the portfolio, Φ, is some fixed value. This minimisation can be implemented by using two Lagrange multipliers, which leads to a set of N linear equations which can be solved for w_i. If we minimise Ω for a number of different values of Φ, then we can obtain a region bounded by an upward-sloping curve, called the *efficient frontier*.

In the case of optimisation with a portfolio containing only hedge funds, the additional constraint of no short-selling is natural due to the difficulties in short-selling of funds (note that short-selling may be achievable by the use of swaps but is uncommon) [2,3].

In order to demonstrate the effects of randomness on the cross-correlation matrix of hedge funds we first divide the time series studied into two equal parts. We assume that we have *perfect knowledge* on the future average returns m_i by taking the observed returns on the second sub-period. We calculate:

(1) the predicted efficient frontier using the correlation matrix for the first sub-period and the expected returns m_i;
(2) the realised efficient frontier using the correlation matrix for the second sub-period and the expected returns m_i.

The portfolio risk due to the noise in the correlation matrix can then be calculated using

$$R_p = \frac{\Omega_r^2 - \Omega_p^2}{\Omega_p^2}, \tag{9}$$

where Ω_r^2 is the risk of the realised portfolio and Ω_p^2 is the risk of the predicted portfolio.

It was shown in [13,14] that correlations may also be measured in the random part of the eigenvalue spectrum. However, since our aim here is to demonstrate how RMT can be used to improve the risk/return profile for a portfolio of hedge funds, we assume that the eigenvalues corresponding to the noise band in RMT, $\lambda_- \leqslant \lambda \leqslant \lambda_+$, are not expected to correspond to real information following [6–12,15,16]. We then use this assumption to remove some of the noise from the correlation matrix. Although the technique used in [6,8] has been shown to lead to problems with the *stability* of the correlation matrix [16], we apply it here as a simple test case to demonstrate how noise can be removed from the cross-correlation matrix formed from hedge fund returns. This technique involves separating the noisy and non-noisy eigenvalues and keeping the non-noisy eigenvalues the same. The noisy eigenvalues are then replaced by their average and the correlation matrix is reconstructed from the cleaned eigenvalues. We can then compare the risk of both the predicted and realised portfolios for the original and cleaned correlation matrices.

3. Results

3.1. Equally weighted correlation matrix

The data set studied here is a collection of 49 hedge funds with varying strategies over a synchronous period from January 1997 to September 2005 ($T = 105$). The original data set was much larger (approximately 1500 funds) but since the length of data available was much less than the number of funds we were forced to choose a subset. The subset chosen were the 49 funds with the longest track records giving us a fund to data ratio $Q = 2.143$. Reducing the data set in this way is not unrealistic as a typical fund of hedge funds would monitor a subset of funds and choose a portfolio from these [2,3]. Often one of the criteria used in choosing this subset of investable funds would be the completion of a minimum track record. Other data sets, such as a portfolio made up of hedge fund strategy indices, were also studied by the authors with notably similar results.

First we calculated the empirical correlation matrix between the funds using equally weighted returns and from this found the spectrum of eigenvalues. This is then compared with the theoretical spectrum for random Wishart matrices (as per Eq. (5)), using $\lambda_- = 0.1$ and $\lambda_+ = 2.83$. As can be seen from Fig. 1, the bulk of the eigenvalues conform to those of the random matrix. There are three deviating eigenvalues, at 10.9886, 8.2898 and 2.944. This means that 6.1% of eigenvalues deviate from the RMT prediction which is consistent with the findings of Ref. [8], where the authors argue that at most 6% of eigenvalues are non-noisy.

3.2. Bootstrapping

In order to show that there was no dependence on the choice of time period or the length of the series we broke the time series up into two segments and again compared the eigenvalue spectrum of the correlation matrix with that of a random Wishart matrix. For both time periods studied we found just two eigenvalues that deviated from the RMT prediction. As can be seen in Fig. 2, the anomalous eigenvalue contributions are very similar for both periods chosen, which implies independence from the choice of time period and stationarity of the data. The values of the deviating eigenvalues are shown in Table 1.
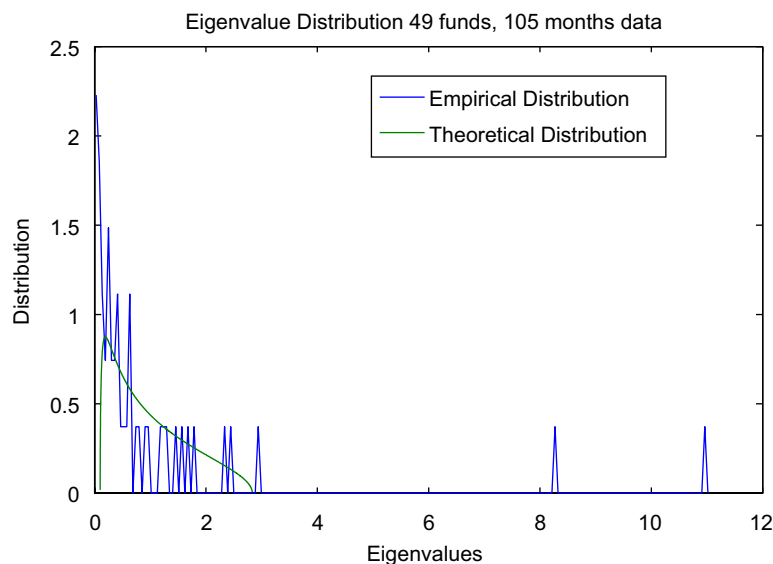

Fig. 1. Spectral distribution for equally weighted hedge fund correlation matrix.
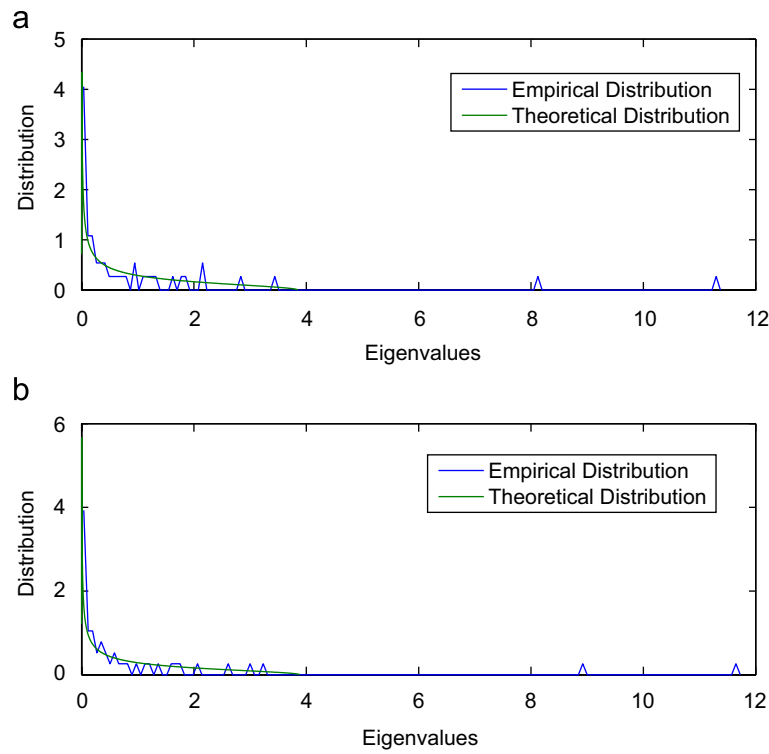
Fig. 2. Bootstrapped spectral distribution for consecutive periods.

Table 1
Deviating eigenvalues

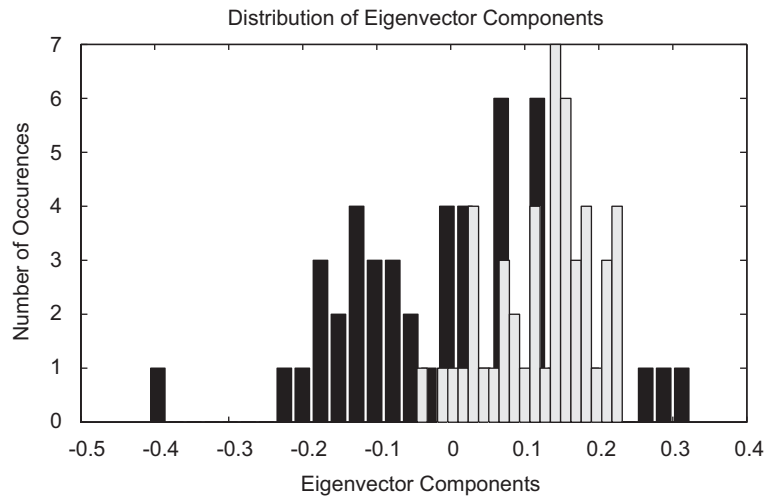| Eigenvalue rank	105 months returns	First 53 months	Second 52 months
1	10.9886	11.334	11.6874
2	8.2898	8.1375	8.9312
3	2.944		

Fig. 3. Comparison of eigenvector components, largest eigenvector (grey), eigenvector from the bulk (black).
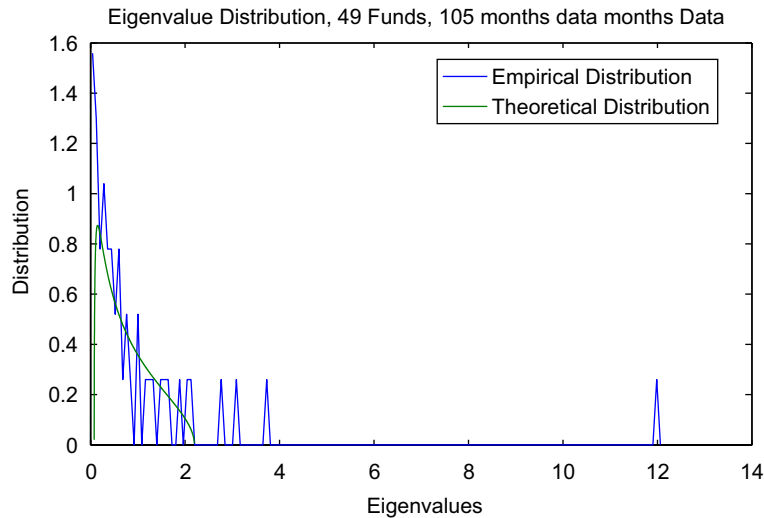
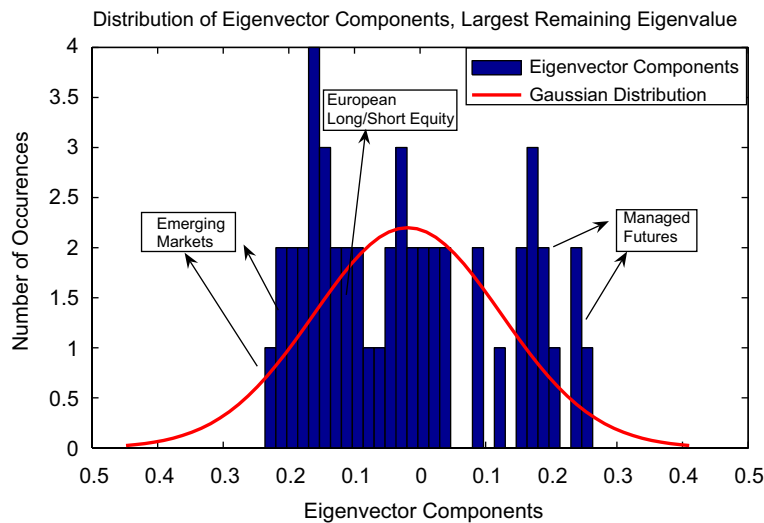Fig. 4. Eigenvalue spectrum after the removal of the effects of the largest eigenvalue.

Fig. 5. Distribution of eigenvector components, largest remaining eigenvalue.

3.3. Eigenvector analysis

Fig. 3 shows the distribution of the components of the largest eigenvector and also the components of a typical eigenvector from the region predicted by RMT. As can be seen from this graph, the distribution of the components of the largest eigenvector are significantly different from that of an eigenvector chosen from the random region. The average value is much larger and the variance of the components much smaller than for the largest eigenvector, which is in agreement with Refs. [11,16] (the largest eigenvector can be interpreted as the 'market'). In this case the 'market' is the set of external stimuli that affect most hedge funds. (e.g., Interest rate changes, large market (i.e., S&P 500, etc.) moves, margin changes, etc.)

We then remove the effects of the largest eigenvalue using the techniques described in Section 2.2. This changes the value of λ_{max} from 2.8329 to 2.1975 which means that 4 of the remaining largest eigenvalues are now outside the RMT region (Fig. 4).

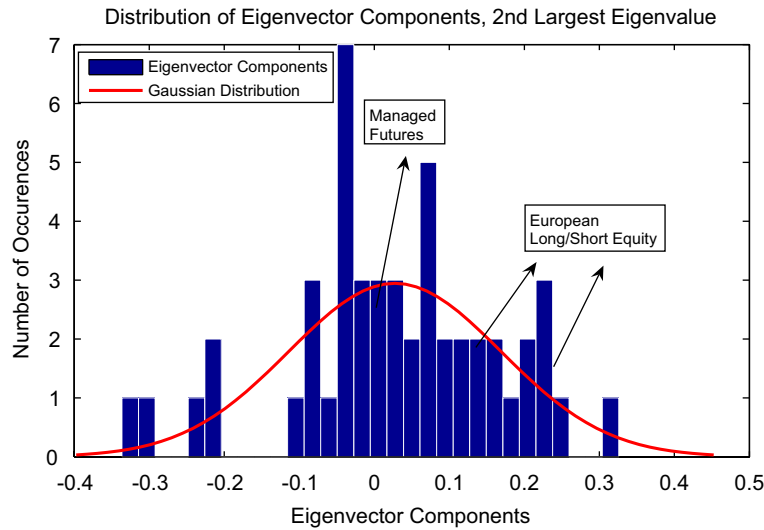Distribution of Eigenvector Components, 2nd Largest Eigenvalue

Fig. 6. Distribution of eigenvector components, second largest remaining eigenvalue.
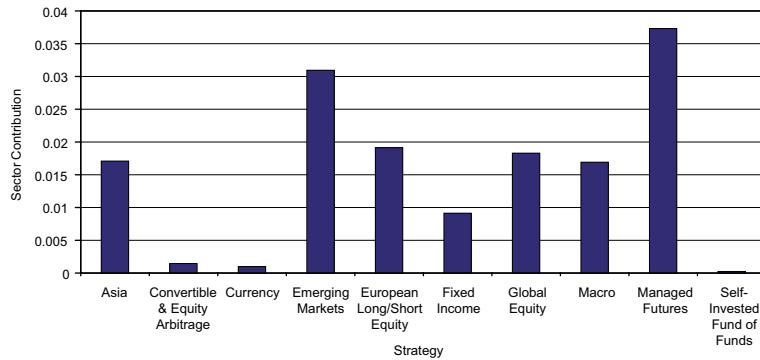

Fig. 7. Strategy contribution, largest eigenvalue.

The distribution of the components of largest remaining deviating eigenvector shows some distinctive clustering (Fig. 5). In particular, the managed futures, emerging markets and European long/short equity strategies are the major contributors here.

A similarly clustered distribution also emerges for the other deviating eigenvalues. An analysis of the eigenvector components for the second largest remaining eigenvector, after removing the effects of the market eigenvector (Fig. 6), shows distinctive clusters emerging, especially for the managed futures and long/short equity sectors. However, the components corresponding to the managed futures strategy do not deviate much from zero and hence make little contribution. These findings for clustering of the deviating eigenvalues agree with Ref. [19] where the authors show that, in addition to the largest, the other deviating eigenvalues of the correlation matrix of asset returns also contain information about the risk associated with the assets.

3.4. Strategy identification

From the analysis described in Section 3.3, we look at the contribution, $X_l^k \equiv \sum_{i=1}^{N} P_{li}^k \left[u_i^k \right]^2$, to each of the deviating eigenvalues from the different strategies employed. Fig. 7 shows X_l^k for the largest remaining
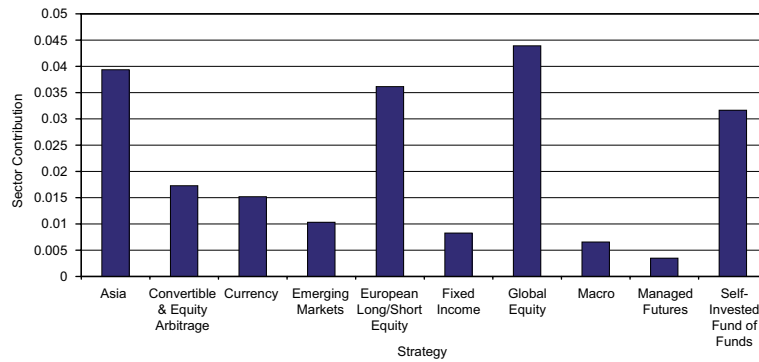
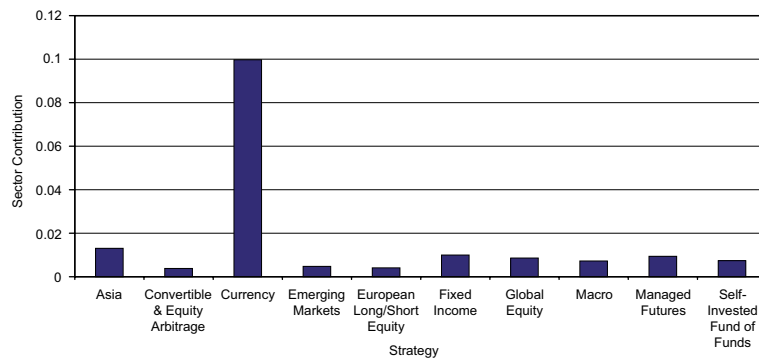Fig. 8. Strategy contribution, second largest eigenvalue.

Fig. 9. Strategy contribution, third largest eigenvalue.

eigenvector once the effects of the market eigenvalue are removed. The largest contributors are clearly managed futures and emerging markets. However, the strategy contribution for managed futures is only around twice that for many of the other sectors. Hence managed futures and emerging markets are the dominant strategies but care in interpretation is needed, since neither contributor is dominant overall.

Fig. 8 shows the strategy contributions for the second largest eigenvalue. These are interesting, since three of the four dominant strategies (Asia, global equity & European long/short equity) are *equity strategies* and would all be affected by events in world equity markets. Also the fourth strategy, self-invested fund of funds, may well also consist of equity funds. However, there is limited information available on exactly what type of funds the managers were invested, although there is reason to believe that a majority of them would be equity based. This implies that this eigenvalue seems to contain information just on equity funds.

Fig. 9 contains the strategy contributions for the third largest remaining eigenvalue. Clearly the dominant strategy here is currency. The final deviating eigenvalue also has one dominant strategy which is self-invested fund of funds. It is notable in the above that analysis of the eigenvalues from within the random matrix region revealed no dominant strategies. The evidence of strategy information in the deviating eigenvalues, coupled with a lack of dominant strategies within the RMT region, supports the idea that information in the correlation matrix is chiefly contained within the deviating eigenvalues. We show how this information can be used to create a portfolio with an improved risk–return profile in Section 3.6.
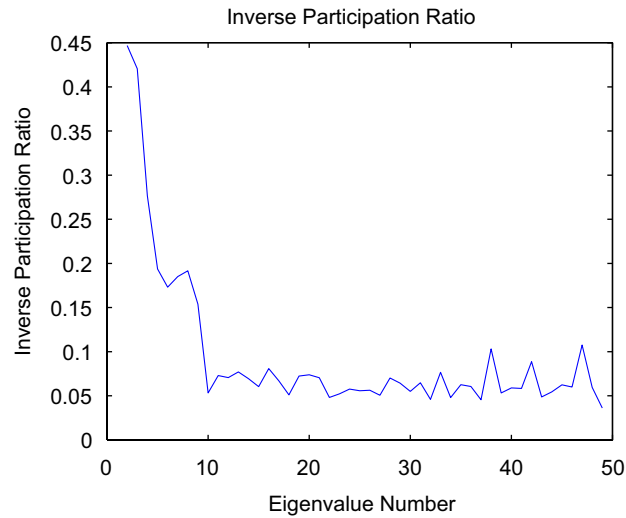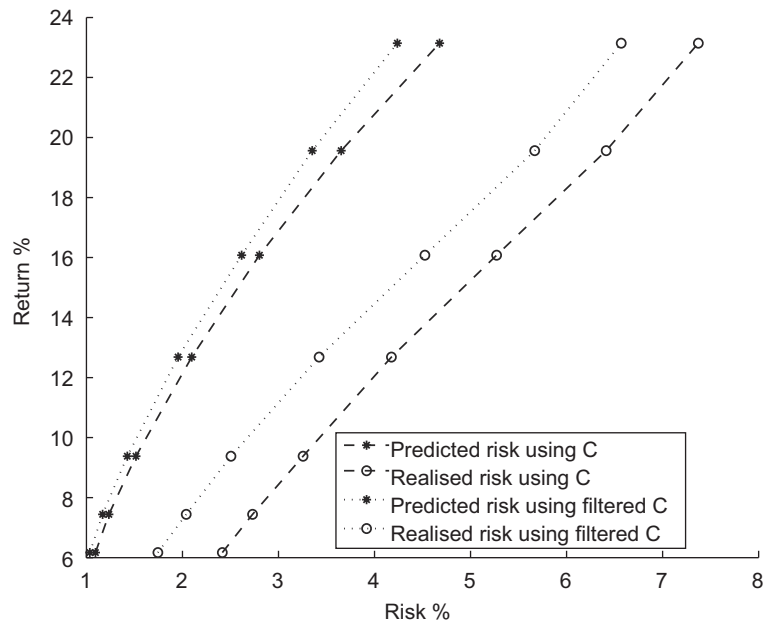
Fig. 10. Inverse participation ratio.

Fig. 11. Efficient frontiers using original and cleaned correlation matrices.

3.5. Inverse participation ratio

Fig. 10 shows the IPR calculated for the eigenvectors of the hedge fund cross-correlation matrix studied. The average IPR value is around 0.06, larger than would be reasonably expected ($1/N \approx 0.02$) if all components contributed to each eigenvector. We would also expect that the largest eigenvectors contributed much more markedly. However, they appear to have a similar IPR to eigenvectors corresponding to the random section. Part of the reason may be that the sample size is small, so the IPR is not particularly effective in terms of assessing by how much larger eigenvectors deviate from the random region, since the average value of the IPR relies on a sample size that tends to infinity.

A significant deviation from the average IPR value is found for the first few eigenvalues. This seems to have been caused by the inclusion of four or five funds, identical apart from being either in different base currencies or leveraged versions of each other with correspondingly high correlation (≈ 1). As mentioned in Ref. [11] funds with a correlation coefficient much greater than the average effectively decouple from the other funds.

3.6. Noise removal and portfolio optimisation

It was noted earlier that where the time series available to estimate cross-correlation matrices are of finite length this leads to measurement noise. This problem is particularly prevalent with hedge fund data, since only monthly returns are available. As seen in Fig. 11 the realised risk is, on average, 292% of the predicted risk. This has obvious consequences for risk management. However, by cleaning the correlation matrix as described in Section 2.3, the realised risk is 190% of the predicted risk. This huge improvement is brought about by limiting the correlation matrix to the information band prescribed by RMT.

It can be seen in Fig. 11 that, for some return values, the predicted risk (using the filtered correlation matrix) is actually less than that of the original correlation matrix. This is due to the constraints imposed on the portfolio, in particular the restriction of no 'short-selling' (Section 2.3). The use of the cleaned correlation matrix leads to a 35% improvement in the difference between the realised risk and the predicted risk for the optimal portfolio. So we have shown that, in the case of hedge fund returns data, the cleaned correlation matrix is a good choice for portfolio optimisation.

4. Conclusions

We have illustrated that, even with limited data (105 months of returns data for 49 hedge funds), useful information can be extracted from a cross-correlation matrix constructed from hedge fund returns. Significant deviations from random matrix theory predictions are observed, with further analysis showing that there is real strategy information contained within the deviating eigenvalues. Eigenvector analysis revealed distinct *strategy clustering* in the deviating eigenvectors. These included emerging markets and managed futures in the largest eigenvector, equity funds in the second, currency and fund of funds in the final two deviating eigenvectors. The strategy information in the deviating eigenvalues was then used to clean the correlation matrix, by flattening the eigenvalues from the bulk to their average and holding the deviating eigenvalues the same. A 35% improvement between the risk of the predicted and realised portfolios was found using this filtering technique.

Appendix A. Hedge fund strategies

Strategies employed by the managers in the sample considered are given in Table A1.

Table A1
Hedge fund strategies

Strategies	Number of funds
Asia excluding Japan long/short equities	2
Convertible and equity arbitrage	2
Currency	7
Emerging markets	6
European long/short equity	10
Fixed income	1
Global equity	5
Japan market neutral	1
Macro	3
Managed futures	11
Self-invested fund of funds	1

References

[1] The Barclay Group, ⟨www.barclaygrp.com⟩, accessed on 1st August 2006.
[2] F.S. L'habitant, Hedge Funds, Wiley, New York, 2002.
[3] F.S. L'habitant, Hedge Funds—Quantitative Insights, Wiley, New York, 2004.
[4] M.A. Miceli, G. Susinno, Ultrametricity in fund of funds diversification, Physica A 344 (2004) 94–99.
[5] M.L. Mehta, Random Matrices, third ed., Elsevier, Amsterdam, 2004.
[6] J.P. Bouchaud, M. Potters, Theory of Financial Risk and Derivative Pricing, Cambridge University Press, Cambridge, 2003.
[7] L. Laloux, P. Cizeau, J. Bouchaud, M. Potters, Noise dressing of financial correlation matrices, Phys. Rev. Lett. 83 (7) (1999) 1467–1470.
[8] L. Laloux, P. Cizeau, M. Potters, J. Bouchaud, Random matrix theory and financial correlations, Int. J. Theor. Appl. Finance 3 (3) (2000) 391–397.
[9] V. Plerou, P. Gopikrishnan, B. Rosenow, L. Amaral, H.E. Stanley, Universal and nonuniversal properties of cross correlations in financial time series, Phys. Rev. Lett. 83 (7) (1999) 1471–1474.
[10] P. Gopikrishnan, B. Rosenow, V. Plerou, H.E. Stanley, Quantifying and interpreting collective behaviour in financial markets, Phys. Rev. E 64 (2001) 035106R.
[11] V. Plerou, P. Gopikrishnan, B. Rosenow, L. Amaral, T. Guhr, H.E. Stanley, A random matrix approach to cross-correlations in financial data, Phys. Rev. E 65 (2002) 066126.
[12] B. Rosenow, V. Plerou, P. Gopikrishnan, H.E. Stanley, Portfolio optimization and the random magnet problem, Europhys. Lett. 59 (2002) 500–506.
[13] Z. Burda, A. Görlich, A. Jarosz, J. Jurkiewicz, Signal and noise in correlation matrix, Physica A 343 (2004) 295–310.
[14] Z. Burda, J. Jurkiewicz, Signal and noise in financial correlation matrices, Physica A (1–2) (2004) 295–310.
[15] D. Wilcox, T. Gebbie, On the analysis of cross-correlations in South African market data, Physica A 344 (1–2) (2004) 294–298.
[16] S. Sharifi, M. Crane, A. Shamaie, H.J. Ruskin, Random matrix theory for portfolio optimization: a stability approach, Physica A 335 (3–4) (2004) 629–643.
[17] H. Markowitz, Portfolio Selection: Efficient Diversification of Investments, Wiley, New York, 1959.
[18] E.J. Elton, M.J. Gruber, Modern Portfolio Theory and Investment Analysis, sixth ed., Wiley, New York, 2002.
[19] A. Sharkasi, M. Crane, H.J. Ruskin, J.A. Matos, The reaction of stock markets to crashes and events: a comparison study between emerging and mature markets using wavelet transforms, Physica A 368 (2) (2006) 511–521.