Can Random Matrix Theory resolve Markowitz Optimization Enigma?

The impact of "noise" filtered covariance matrix on portfolio selection.

by

Kim Wah Ng

A Dissertation submitted to the

Graduate School-Newark

Rutgers, The State University of New Jersey

in partial fulfillment of the requirements for the degree of

Doctor of Management

Graduate Program in Finance

written under the direction of

Professor Michael S. Long

and approved by

Professor Ben Sopranzetti

Professor Yangru Wu

Professor Ren-Raw Chen (Fordham University)

Newark, New Jersey

May, 2014

ABSTRACT OF THE DISSERTATION

Can Random Matrix Theory resolve Markowitz Optimization Enigma?

The impact of "noise" filtered covariance matrix on portfolio selection.

by Kim Wah Ng

Dissertation Director:

Professor Michael Long

Modern finance theory is based on the simple concept of risk and return trade-off. Risk is based upon one holding a diversified portfolio to get the lowest level of risk for a given expected return. This is the foundation of Markowitz's mean-variance (MV) efficient portfolio.

For nearly six decades since Markowitz's pioneering work, it is still a puzzle as to why there are persistent doubts about the performance of MV approach to portfolio selection and the lack of acceptance as a viable tool in the investment community. This puzzle is coined as the "Markowitz optimization enigma". The major problem with MV optimization is its tendency to maximize the effects of estimation errors in the risk and return estimates.

The latest attempt to reduce the noise in covariance estimates is a branch from physics that uses Random Matrix Theory (RMT) prediction. The prediction is that when the number of securities is large relative to the number of observations, the eigenvalues of the covariance matrix within a predicted band closely resemble the distribution as if they were generated from purely random returns. These studies believe that by modifying the eigenvalues within the predicted band, the "filtered" covariance matrix would contain better information than the raw sample matrix.

One proprietary commercial product, called the Neutron QuantumApp which was released in mid-2013, based its filtration technique on RMT prediction. The motivation of this dissertation is to examine the effectiveness of the Neutron product in terms of risk measurement, mean-variance efficiency and portfolio performance. More specifically, does the filtered covariance contain superior information as compared to the raw covariance?

The evidence shows that the efficient frontier, generated from filtered covariance, indeed dominates the raw efficient frontier for the unconstrained case. When short-sale constraint is imposed, the result is similar except for the minimum variance portfolio (MVP). The MVP from the raw matrix dominates the MVP from the filtered matrix. In general, the filtered covariance appears to be better for the purpose of risk measurement and risk management. The filtered correlation structure is considerably higher.

However, more efficient portfolios do not translate into better performers. For the period 2006 to 2013, one cannot reject the null hypothesis that the filtered portfolios perform similarly to the raw portfolios. Therefore, my conclusion is that the Neutron product cannot resolve Markowitz optimization enigma.

PREFACE

This Ph.D. dissertation started while I was at the last stage of my consulting contract with Teknavo Group Ltd. in 2013. They partnered with Market Memory Trading, L.L.C. in developing the Neuron QuantumApp. Surprisingly, the claims of this application did not arouse the interest of many portfolio managers. Thus, my quest and curiosity regarding whether this particular application could resolve Markowitz's long-standing enigma began.

When I started researching the topic on Random Matrix Theory (RMT), I had very little knowledge of this subject. Fortunately, several studies on RMT had already been published since 1999 and they all pointed to the random nature of estimation noise in correlation matrix. In the past, there have been many examples of successful application of physics theory to solving financial problems. Could RMT be the answer to better portfolio selection and better risk management?

Although I was not able to get clear answers for many questions regarding the Neutron QuantumApp, I was able to reverse engineer the eigen system of the filtered correlation matrix. The main clue that mattered most was found to be what one does to the random portion or noisy part of the eigen system. There are clear implications and impact to the correlation and variance structure of the transformed covariance matrix and these will impact the ultimate portfolio selection decision.

I believe that RMT can be useful in Finance, but one has to be cautious of the limitations. It cannot be a one-size-fits-all solution. In fact, my research shows that the simple naïve strategy of portfolio selection is more powerful during a financial crisis in terms of risk diversification and portfolio performance. Portfolio managers would do more good to their investors by comparing their strategies to the performance of random naïve portfolios.

ACKNOWLEDGEMENTS

I would like to thank my Ph.D. committee for their kind consideration and acceptance of this thesis topic in portfolio optimization, which many in Finance would regard as either uninteresting or unchallenging. Special thanks to Professor Michael Long for his patience with me. I changed my thesis topic three times, and each time he patiently read my papers and made good comments. I truly deserve the title of “Rip Van Winkle” student as suggested by Professor Long. I thank Professor Ben Sopranzetti, Professor Yangru Wu and Professor Ren-Raw Chen (Fordham University) for being on my committee and for their valuable feedback.

The most patient person of all is Professor Ivan Brick. He has constantly encouraged me to finish my Ph.D. dissertation. Without his help and patience, this thesis would not be possible. I also thank Tom Cox from Teknavo Group for letting me use their application and the Bloomberg data for the empirical research.

Special thanks to my father Nam-Onn Ng and my mother Hoong-Lin Yip, who are unfortunately not among us. They encouraged and supported me throughout my Bachelor and Master programs. Most importantly, my special thanks and love to my wife, Joyce, who stuck with me for twenty-six years, and together we have four wonderful children – Eric, Lianne, Jonathan and Gabriel. She also edited my entire dissertation. I

hope that finishing this Ph.D. will inspire in my children the importance of finishing what they start, even if it takes years to do so.

Table of Contents

List of table titles and page numbers

List of illustration titles and page numbers

Can Random Matrix Theory resolve Markowitz Optimization Enigma?

The impact of "noise" filtered covariance matrix on portfolio selection.

Introduction

Modern finance theory is based on the simple concept of risk and return trade-off. Risk is based upon one holding a diversified portfolio to get the lowest level of risk for a given expected return. Now, the problem comes in estimating this risk when many different securities exist, and when the returns of these securities often move in similar fashion. Estimating these risks can be tricky and often computational intensive. In light of new evidences, that these risk estimates are subject to estimation "noise", the issue becomes further complicated. This study examines a new approach that claims to "filter" out the noise of the risk estimate which may in turn, produce better portfolio allocation decisions

Chapter 1

Markowitz Optimization Enigma

Markowitz's (1952, 1956) seminal idea, that investors should hold mean-variance (MV) efficient portfolios, is the milestone of modern finance theory for optimal portfolio construction, asset allocation and investment diversification. Not only did he highlight the benefit of naive diversification of unsystematic risk or idiosyncratic risk by merely increasing the number of securities in a portfolio, but also one can do better by forming "efficient" portfolios based on optimizing risk and expected return trade-off. Markowitz defines "efficient" portfolios in the following sense: (a) any portfolio possessing a higher expected yield[1] than an efficient portfolio also has a higher variance of yield (i.e., it is riskier); and (b) any portfolio that is less risky than an efficient portfolio must have a lower expected yield. The identification of riskiness with the variance of the yield results from the assumption that the investor's utility function is concave, with the form

$$\text{Utility} = \text{Yield} - \theta(\text{Yield})^2$$

where θ is some constant, or else from the assumption that yields are distributed according to a distribution that can be described by second-order moments. From the optimal set of efficient portfolios available, the investor selects that efficient portfolio that yields him the maximum utility. That is at the point where his utility curve is tangential to the efficient frontier. In other words, portfolio optimizers respond to the

[1] Markowitz defines yield as "(the closing price for the year) minus (the closing price for the previous year)

uncertainty of an investment by selecting portfolios that maximize profit subject to achieving a specified level of calculated risk or, equivalently, minimize variance subject to obtaining a predetermined level of expected gain, (See Kroll, Levy and Markowitz, (1984)).

Implementing Markowitz's technique involves a critical-path method of quadratic programming, a vector of estimated expected yields for the securities under consideration and a matrix of estimated variances and covariances[2]. Roughly speaking, optimal portfolios are achieved by reducing the covariance term in the portfolio variance expression through diversification. The main problem is that of estimating the expected yields and covariance matrix. This must come either from (a) historical data (objective estimation of the parameters of the multivariate distribution), (b) subjective estimation by an expert in the securities market, or else (c) via estimates of the correlation of individual yields with an index or several independent indexes.

For nearly six decades since Markowitz pioneering work in MV portfolio framework, there have been persistent doubts about the performance of MV approach to portfolio selection, despite the fact that several procedures for computing the corresponding risk-return estimates were developed. [3] Michaud (1981) notes that MV optimization is one of the outstanding puzzles in modern finance and that it has yet to

[2] An alternative measure is to use semi-variance estimates as mentioned by Markowitz (1956).
[3] Sharpe 1967, 1971, Stone 1973, Elton, Gruber, and Padberg 1976, 1978, Markowitz and Perold 1981, Perold 1984).

meet with widespread acceptance by the investment community, particularly as a practical tool for active equity investment management. He coins this puzzle the "**Markowitz optimization enigma**" and calls the MV optimizers "**estimation-error maximizers".** More specifically, Michaud (1981) states

> "The major problem with MV optimization is its tendency to maximize the effects of errors in the input assumptions. Unconstrained MV optimization can yield results that are inferior to those of simple equal-weighting schemes[4]… Risk and return estimates are inevitably subject to estimation error. MV optimization significantly overweights (underweights) those securities that have large (small) estimated returns, negative (positive) correlations and small (large) variances".

The latest attempt to reduce the noise in the risk estimates is a branch from physics that uses Random Matrix Theory (RMT) prediction. The prediction is that when the number of securities is large relative to the number of observations, the eigenvalues of the covariance matrix within a predicted band closely resemble the distribution as if they were generated from purely random returns. These studies believe that by modifying the eigenvalues within the predicted band, the "filtered" covariance matrix would contain better information than the raw sample matrix.

The motivation of this dissertation is to examine the effectiveness of the "filtered" covariance estimates in terms of risk measurement, mean-variance efficiency and portfolio performance. Specifically, the research is focused on a proprietary commercial product, called the Neutron QuantumApp which was released in mid-2013. This application directly applies RMT prediction to its filtration technique. The goal of this

[4] See Jobson and Korkie (1981)

dissertation is to examine the effectiveness of the Neutron product in terms of risk measurement, mean-variance efficiency and portfolio performance. More specifically, does the filtered covariance contain superior information as compared to the raw covariance?

Mathematical Framework

This section presents a mathematical framework of the evolution of portfolio selection from Markowitz's MV optimization to issues relating to estimation. The aim is to better understand the nature of noise relating to RMT and the competing techniques with respect to the estimation of the covariance matrix.

The starting point is MV portfolio selection theory. First, define the following[5]

$\boldsymbol{\mu} = \{E[r_1], E[r_2], \dots, E[r_N]\}^T$ as the vector of expected returns

$\boldsymbol{C}$ = covariance matrix of the returns where $\sigma_{ii} = \sigma_i^2 = Var(r_i)$ and $\sigma_{ij} = Cov[r_i, r_j]$

$\boldsymbol{w} = [w_1, w_2, \dots, w_N]^T$ is the vector of weights where w_i is the fraction of the total amount of invested capital in asset i. Given these definitions, the expected return and variance of the portfolio can be written as:

$$E[r_p] = \boldsymbol{w}^T\boldsymbol{\mu} \tag{1}$$

$$Var[r_p] = \boldsymbol{w}^T\boldsymbol{C}\boldsymbol{w} \tag{2}$$

[5] Bold letter denotes both matrix and vector

Excluding any risk free asset and allowing for unlimited short selling, the MV optimization problem can be written as:

$$\max_w \{\, \theta \boldsymbol{w}^{\boldsymbol{T}} \boldsymbol{\mu} - \frac{1}{2} \boldsymbol{w}^{\boldsymbol{T}} \boldsymbol{C} \boldsymbol{w} \;\;|\;\; \boldsymbol{w}^{\boldsymbol{T}} \mathbf{1} = 1 \,\} \tag{3}$$

where **1** denotes a *N x 1* vector of ones and the parameter θ is the MV investor's risk tolerance parameter which relates expected return to risk trade-off. The optimal portfolio to the MV optimization problem can be written as:

$$\boldsymbol{w}^* = \frac{\boldsymbol{C}^{-1}\mathbf{1}}{b} + \theta[\boldsymbol{C}^{-\mathbf{1}}\left(\boldsymbol{\mu} - \frac{\mathbf{1}a}{b}\right)] \tag{4}$$

where

$$a = \mathbf{1}^{\boldsymbol{T}} \boldsymbol{C}^{-\mathbf{1}} \boldsymbol{\mu}$$

$$b = \mathbf{1}^{\boldsymbol{T}} \boldsymbol{C}^{-\mathbf{1}} \mathbf{1}$$

The minimum variance portfolio (MVP) is thus given by:

$$\boldsymbol{w}^* = \frac{\boldsymbol{C}^{-1}\mathbf{1}}{b} \tag{5}$$

which is independent of the expected return vector $\boldsymbol{\mu}$. It is easy to see that the solution requires the inversion of the covariance matrix $\boldsymbol{C}$. Equation (4) states that the optimal portfolio to the investor is the linear combination of the MVP and another "risky" portfolio (governed by expected return) scaled by the risk tolerance of the investor.

The solution in (4) depends on both the vector of expected return estimates and the sample covariance estimates. Under the assumption of normal distribution in returns, the sample covariance is the best unbiased maximum likelihood (ML) estimator. However, the ML estimator is desirable and useful property if there is enough data for the estimation process. In a small sample size, there is the danger of over-fitting the data. This implies that the sample covariance matrix performs the best in-sample but may perform poorly out-of-sample.

For *N* assets, there are *N(N+1)/2* parameters to be estimated in a covariance matrix. Given 150 assets, one has to estimate 11,325 parameters. This requires a lot of data for estimation. In order to comprehend the nature of this large-scale problem, first define $\mathbf{R} = [\boldsymbol{r_1}, \boldsymbol{r_2}, \ldots., \boldsymbol{r_T}]$ as an *N X T* matrix with

$$\bar{\boldsymbol{r}} = \frac{1}{T}\sum_{t=1}^{T} r_t = \frac{1}{T}\ \mathbf{R1}$$

and the **sample covariance matrix** as

$$\boldsymbol{S} = \frac{1}{T-1}\sum_{t=1}^{T}(\boldsymbol{r_t} - \bar{\boldsymbol{r}})(\boldsymbol{r_t} - \bar{\boldsymbol{r}})' = \frac{1}{T-1}\ \mathbf{R}(\mathbf{I} - \frac{1}{T}\mathbf{11}')\mathbf{R}' \qquad (6)$$

where **I** is the identity matrix and T is the number of historical observations on each of the assets. As quoted by Ledoit and Wolf (2003)

> "Equation (6) shows why the sample covariance matrix is not invertible when N ≥ T: the rank of **S** is at most equal to the rank of the matrix **I – 11'**/T, which is T-1. Therefore, when the dimension N exceeds T-1, the sample covariance matrix is rank-deficient. Intuitively, the data do not contain enough information to estimate the unrestricted covariance matrix."[6]

[6] See page 608 of Ledoit and Wolf (2003) and also page 4 of Bengtsson and Holst (2002)

Clearly when one multiplies the matrices fully in (6), S becomes a *NxN* matrix of rank N.

The second stage of the evolution in resolving the aforementioned scale problem is the class of “factor” models. These models assume that the returns are generated by specific exogenous factors with some underlying economic interpretation. The most famous of these is Sharpe’s (1963) single-index market model. For the ith asset, the single index model can be written as:

$$r_{it} = \alpha_i + \beta_i r_{Mt} + \varepsilon_{it} \tag{7}$$

where r_{Mt} is the return on the market. For all N assets, the matrix form is given by

$$\mathbf{r}_{\boldsymbol{t}} = \boldsymbol{\alpha} + \boldsymbol{\beta}\mathbf{r}_{\mathbf{Mt}} + \boldsymbol{\varepsilon}_{\boldsymbol{t}} \tag{8}$$

where $\boldsymbol{\varepsilon}_{\boldsymbol{t}}$ is a *N X 1* vector containing the zero mean uncorrelated residuals ε_{it}. The covariance matrix for the asset returns, as implied by the market model, is simply

$$\boldsymbol{\Omega} = \sigma_M^2 \boldsymbol{\beta}\boldsymbol{\beta}^{\boldsymbol{T}} + \sum\nolimits_{\boldsymbol{\varepsilon}} \tag{9}$$

where σ_M^2 is the variance of the market portfolio and β’s are estimated using (8).

The main advantage of the single index model is that it only requires 2N + 1 parameter estimates. This is a major reduction in parameters as compared to the full sample covariance estimator. Since there are more data per estimated parameter, one would expect a substantial reduction in estimation error. However, this is at the expense of introducing ***specification error*** because of the restrictive assumptions that the asset returns are generated by a linear function of the market.

In general, all factor models impose some form of structural assumptions and the fewer the factors, the stronger the structure. The single index model follows the well-known capital asset pricing model (CAPM), and the strong structure can introduce specification error. Therefore, one encounters two types of error in the sample covariance estimation: estimation error versus specification error. One way around this issue is by introducing more "factors" into the structure. By doing so, one would hope to reduce the specification error. An alternative class of "factor" model arises from the Arbitrage Pricing Theory (APT), which is an extension of the single index model. Thus, in addition to the market factor, there may be other uncorrelated factors that could explain the returns of assets. Using (8), the matrix form can be written as:

$$\mathbf{r}_{\boldsymbol{t}} = \boldsymbol{\alpha} + \boldsymbol{\beta}_{\mathbf{1}}\mathbf{r}_{\mathbf{Mt}} + \boldsymbol{\beta}_{\mathbf{2}}\mathbf{r}_{\mathbf{F2t}} +, \dots, + \boldsymbol{\beta}_{\mathbf{k}}\mathbf{r}_{\mathbf{Fk}} + \boldsymbol{\varepsilon}_{\boldsymbol{t}} \tag{10}$$

with the covariance matrix as:

$$\boldsymbol{\Omega} = \sigma_M^2 \boldsymbol{\beta}_{\mathbf{1}}\boldsymbol{\beta}_{\mathbf{1}}^{\boldsymbol{T}} + \sigma_{F2}^2 \boldsymbol{\beta}_{\mathbf{2}}\boldsymbol{\beta}_{\mathbf{2}}^{\boldsymbol{T}} + \dots + \sigma_{Fk}^2 \boldsymbol{\beta}_{\boldsymbol{k}}\boldsymbol{\beta}_{\boldsymbol{k}}^{\boldsymbol{T}} + \sum\nolimits_{\boldsymbol{\varepsilon}} \tag{11}$$

where σ_{Fk}^2 is the variance of the k^{th} factor and $\sigma_{Fk,Fj} = 0$ for all factors. The factors are orthogonal to each other.

Clearly, the more factors there are, the less restrictive the structure becomes. However, APT does not give the identity of each factor. Instead, many researchers rely on factor decomposition techniques on the sample covariance matrix in order to infer the hidden factors. Principal Component analysis (PCA) is probably the most famous tool used for this purpose. The goal of PCA is to explain covariance structures using only a few linear combinations of the original stochastic variables. For an *N X N* covariance matrix, *N* principal components are needed to reproduce all variability in the system. In other words, PCA uses orthogonal transformation to convert a set of correlated variables into a set of linearly uncorrelated components. However, most variability in the system can be explained by a lesser number of *P*, $P < N$, principal components without losing much information. What PCA accomplishes is both data reduction and interpretability. In fact, PCA often proves to reveal relationships that otherwise would have been hard to detect.

Let's define the sample principal components as linear combinations of returns

$$f_i = \boldsymbol{\zeta}_i^T \mathbf{r} = \sum_{n=1}^{N} \zeta_{ni} r_n \quad i = 1,2, \dots, N \tag{12}$$

For which the variances and covariances are given by:

$$Var[f_i] = \boldsymbol{\zeta_i^T S \zeta_i} \quad i = 1,2, \dots, N \tag{13}$$

$$Cov[f_i, f_j] = \boldsymbol{\zeta_i^T S \zeta_j} \quad i,j = 1,2, \dots, N$$

$$\boldsymbol{\zeta_i^T \zeta_i = 1}, \quad i = 1,2, \dots, N$$

where $\boldsymbol{\zeta_i}$ is the loading for the i[th] component, $\boldsymbol{S}$ is the sample covariance matrix.

The optimal weight function to the i^th^ Principal Component is given by:

$$\max_{\zeta_i} \{ \frac{\zeta_i^T S \zeta_i}{\zeta_i^T \zeta_i} \quad | \quad Cov[f_i, f_j] = 0, \ \ \forall\, 0 < j < i \ \} \tag{14}$$

Note that the optimal weight functions $\boldsymbol{\zeta_i}$ are solely dependent on the covariance matrix $\boldsymbol{S}$. The sample covariance matrix **S** can be decomposed into its eigenvalues, $\lambda_1 \geq \lambda_2 \geq \lambda_3 \ldots.. \geq \lambda_N \geq 0$ and **eigenvalue-eigenvector** pairs $(e_1, \lambda_1), (e_2, \lambda_2), .., (e_N, \lambda_N)$ with eigenvectors $\mathbf{e_j} = [e_{1j}, e_{2j}, \ldots, e_{nj}]^T$.

Using spectral decomposition of $\boldsymbol{S}$, the solution for the i^th^ (sample) Principal Component is given as:[7]

$$f_i = e_i^T \mathbf{r} = \sum_{n=1}^{N} e_{ni} r_n \qquad i = 1,2, \ldots., N \tag{15}$$

and $$Var[f_i] = \boldsymbol{e_i^T S e_i} = \lambda_i \quad i = 1,2, \ldots, N \tag{16}$$

$$Cov[f_i, f_j) = \boldsymbol{e_i^T S e_j} = 0 \quad i \neq j$$

Therefore, each principal component is determined by its eigenvector**.** The variance of each component is its corresponding eigenvalue. The principal components are uncorrelated with each other. Furthermore, the (sample) correlation between the i^th^ asset and the j^th^ principal component is given by[8]

$$\rho[r_i, f_j] = \frac{e_{ij}\sqrt{\lambda_j}}{s_{ij}} \tag{17}$$

[7] See Johnson and Wichern (1992) for the proof.
[8] See Johnson and Wichern (1992) for the proof.

When analyzing financial assets, it is more common to apply PCA on the correlation matrix rather than the covariance matrix. The reason is that variables are measured on different scales with differing ranges. Using correlation, assets with differing magnitude of volatility are standardized to avoid an unreasonable large impact on the principal components by only a few variables. In this way, the first principal component for asset returns usually resembles a market factor, and other principal components often mirror industry specific effects (see King 1966).

An alternative method to PCA is the Factor Analysis model. The factor analysis model is based on the assumption that the variables all depend on a number of underlying and unobservable stochastic factors, denoted by $F_1, F_2, \ldots, F_k,$ as well as the variable specific errors / variations $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_N$. Denote the ***sample mean*** of the N dimensional stochastic vector **r** by **m.** In matrix notation, the factor model is written as:

$$\mathbf{r} - \mathbf{m} = \mathbf{LF} + \boldsymbol{\varepsilon} \tag{18}$$

where **L** contains the loading coefficients Υ_{ij} on variable i by factor j. **F** is an unobservable vector of the factors. The following assumptions are made:

$E[\mathbf{F}] = 0, \quad Var[\mathbf{F}] = E[\mathbf{FF'}] = \mathbf{I}$ an identity matrix

$E[\boldsymbol{\varepsilon}] = 0, \quad Var[\boldsymbol{\varepsilon}] = E[\boldsymbol{\varepsilon\varepsilon'}] = \boldsymbol{\Psi}$ a diagonal matrix of residual variances,

$Cov[\mathbf{F}, \boldsymbol{\varepsilon}] = 0$

The implied covariance structure for **r** from the factor model is

$$\mathrm{Var}[\,\mathbf{r}] \;=\; \mathrm{E}[(\mathbf{r} - \mathbf{m})(\mathbf{r} - \mathbf{m})^{\mathrm{T}}] \;=\; \mathrm{E}[\,(\mathbf{LF} \,+\, \boldsymbol{\varepsilon})(\mathbf{LF} \,+\, \boldsymbol{\varepsilon})^{\mathrm{T}}]$$

$$=\; \mathbf{L}\mathrm{E}[\mathbf{FF^T}]\mathbf{L^T} \;+\; \mathrm{E}[\varepsilon\mathbf{F^T}]\mathbf{L^T} \;+\; \mathbf{L}\mathrm{E}[\mathbf{F}\boldsymbol{\varepsilon}^{\mathbf{T}}] \;+\; \mathrm{E}\,\boldsymbol{\varepsilon\varepsilon}^{\mathbf{T}}]$$

$$=\; \mathbf{LL^T} \;+\; \boldsymbol{\Psi} \tag{19}$$

It then follows that

$$Var[r_i] = \gamma_{i1}^2 + \gamma_{i2}^2 + \cdots . + \gamma_{iK}^2 + \varphi_i = h_i^2 + \varphi_i$$

$$Cov[r_i, r_j] = \gamma_{i1}\,\gamma_{j1} + \gamma_{i2}\,\gamma_{j2} + \cdots . + \gamma_{iK}\,\gamma_{jK} \tag{20}$$

$$Cov[r_i, F_j] = \gamma_{ij} = e_{ij}\sqrt{\lambda_j}$$

Using spectral decomposition of the sample covariance matrix **S** into its eigenvalue-eigenvector pairs, the following holds:

$$\mathbf{S} = \lambda_1 e_1 e_1^T + \lambda_2 e_2 e_2^T + \dots . + \lambda_N e_N e_N^T$$

$$= \left[\,\sqrt{\lambda_1} e_1 \;\; \sqrt{\lambda_2} e_2 \dots . \sqrt{\lambda_N} e_N\,\right]\;[\,\sqrt{\lambda_1}\; e_1^T \;\; \sqrt{\lambda_2}\; e_2^T \;\dots\dots\; \sqrt{\lambda_N}\; e_N^T\;]^{\,T} \tag{21}$$

This decomposition of **S** in terms of its eigenvalue-eigenvector pairs is equivalent to (19) when the residual variance $\boldsymbol{\Psi} = \mathbf{0}$. In terms of factor loadings, **S** can be written as:

$$\mathbf{S} = \mathbf{LL^T} \tag{22}$$

where **L** is a N x N matrix. The loadings on the j^{th} factor are the coefficients in the j^{th} principal component multiplied by a scale factor $\sqrt{\lambda_j}$.

The practical aspect of choosing how many "significant" factors or principal components to describe the covariance matrix is usually based on the cumulative variance that can be explained by K components, (where K < N) as a percentage of total variability. In equation (16), the variance of each component is actually its own eigenvalue. Suppose the cut-off point is 90%. Then, one would select K such that

$$\frac{\sum_i^K \lambda_i}{\sum_i^N \lambda_i} \geq 0.9$$

However, knowing K does not reveal the true nature of the components or factors in the real world. Many practitioners would then attempt to find economic variables, industry or firm specific variables that are highly correlated to the eigenvectors associated with the K components. In this case, these factors can be explained intuitively to investors as the "risk" premiums. This approach is more appealing and **sellable** than simply using MV optimization for asset selection. The factor model becomes the portfolio manager's black-box, and the real-world variables become her secret weapon in generating alphas. Moreover, the portfolio manager appears to understand the portfolio risk better by "managing" and "monitoring" their risk premiums. This would not be possible with the raw covariance matrix.

In recent years, there has been renewed interest in portfolio selection theory based on the relatively new field of **econophysics** and the works of Plerou et al. (1999, 2001), Laloux et al. (1998) and Guhr (2001). The application of Random Matrix Theory (RMT) to large correlation matrix in finance has provided new excitement in the portfolio field. The idea behind this research is to separate the real correlation from the estimation error by comparing the sample correlation matrix with known results for a completely ***random correlation matrix***. Suppose $\tilde{C}$ is defined as a random correlation matrix

$$\tilde{\boldsymbol{C}} = \frac{1}{T-1}\tilde{\boldsymbol{Z}}\widetilde{\boldsymbol{Z}^{T}} \tag{23}$$

where $\tilde{\boldsymbol{Z}}$ contains mutually uncorrelated time series with zero mean and unit variance from an empirical distribution. According to RMT when N and T tend to infinity such that $Q = T/N$ is fixed, the eigenvalue distribution $\rho_{RM}(\lambda)$ of a matrix like $\tilde{\boldsymbol{C}}$ is given by

$$\rho_{RM}(\lambda) = \frac{Q}{2\pi}\frac{\sqrt{(\lambda_{max} - \lambda)(\lambda - \lambda_{min})}}{\lambda} \tag{24}$$

where

$$\lambda_{min} \leq \lambda \leq \lambda_{max}$$

and the bounded region of eigenvalues is given by

$$\lambda_{min}^{max} = 1 + \frac{1}{Q} \pm 2\sqrt{\frac{T}{Q}} \tag{25}$$

In other words, for a purely random matrix such as $\widetilde{\boldsymbol{C}}$, the eigenvalues are expected to be distributed according to equation (24) and the boundary of this distribution is given by (25). If one, then, compares the distribution of eigenvalues from a real-world sample correlation matrix, **S,** to the distribution as described in (24) then the ***overlapping*** part of the eigenvalues in **S** with those eigenvalues in $\widetilde{\boldsymbol{C}}$ is considered "noisy" or random. The real correlation or information contained in S is the non-overlapping eigenvalues outside λ_{max}. There are abundant evidences that a large portion of the sample correlation of asset returns is random noise due to estimation errors. This approach gives a systematic way to identify the seriousness of estimation error in any large scale covariance matrix. Appendix A contains more detail discussion on RMT.

There are several empirical evidences to support the RMT. These findings show a consistent pattern, in that the bulk of the eigenvalue distribution of the cross correlation matrix of a major index [S&P 500 of the NYSE and Tokyo Stock Exchange (TSE)] is found to follow the eigenvalue distribution of the Wishart matrix[9], which is a random correlation matrix constructed from mutually uncorrelated time series. In short, the eigenvalue distribution obeys the RMT prediction in the bulk, but there are some deviations at the larger eigenvalues. Utsugi et al. (2004) further examine the nearest-neighbor spacing, the next-nearest-neighbor spacing, and the rigidity of the eigenvalues and

[9] Laloux et al. (1999), Plerou et al. (1999), (2002) and Utsugi et al. (2004).

> “found that they follow the universality of Gaussian Orthogonal Ensemble (GOE) .... implying that the large eigenvalues are due to the existence of true correlations while the eigenvalues distributed in the bulk are due to randomness”.

When they examine the properties of the eigenvectors associated with the large eigenvalues, they found sector effect – in the S&P data, the electric power sector and oil and gas related sectors play major parts in the correlations. In the TSE, the electric products sector and construction sector play major parts.[10] These results are consistent with the findings of sector effect by King (1966).

The noise impact, due to insufficient number of observations relative to the number of variables, is serious enough in affecting decision making based on large correlation matrix. The distortion can have serious implications in terms of trading, asset allocation and risk management. The question is what can be done to mitigate the randomness. Clearly, one can simply increase the number of data such that the Q ratio is sufficiently large. That is fine when dealing with a small number of variables, but merely increasing the number of observations may not be suitable if there is non-stationary property in the underlying correlation structure. If one believes there is a regime change in the economy, there may be insufficient data to estimate the correlation for the new regime. If an effective method can be found to “clean” the estimation error from the sample correlation matrix, then it is a preferred approach, as it avoids any specification error.

[10] The data period for TSE was from January 1993 to June 2001, and January 1991 to July 2001 for the S&P 500. Daily data were used in the study.

The chapters in this dissertation are organized as follow. Chapter 2 discusses the competing methods for reducing the error in sample covariance matrix. Chapter 3 examines the characteristics of "filtered" correlation matrices that are generated by a commercial product, called the "Neutron QuantumApp", are examined. The data samples and the properties of the filtered correlation matrices are also analyzed. Subsequent tests compare the 'filtered" matrices from Neutron with the "raw" sample correlation matrices. Chapter 4 examines the predictive power of the filtered risk measure as compared to the raw risk measures. Chapter 5 investigates the effectiveness of using the filtered correlation matrix in passive strategy.

Chapter 6 examines the return performance of MV optimization using filtered correlations and excluding the effect from expected return estimates. Chapter 7 further examines the performance using the passive strategy. Chapter 8 includes the expected return estimates in the MV optimization process and examines the effectiveness in enhancing return performance. Chapter 9 implements a modified version of the three fund separation strategy in an attempt to improve portfolio performance. Chapter 10 attempts to explain the puzzle surrounding the RMT filtering method. The summary and final remarks are contained in Chapter 11.

Chapter 2

Reducing Error in Covariance Matrix

The previous chapter laid the mathematical framework around the estimation error in sample covariance matrix and the implication on portfolio selection. This chapter reviews three competing methods of reducing the error in sample covariance matrix. The first type is the class of shrinkage estimators. The second method of error reduction is the portfolio of estimators. The third method deals with modifying the eigenvalues that are associated with noise, as predicted by RMT. In this chapter, these three methods are briefly described along with the empirical findings relating to their estimators.

Like any other estimation process, the estimation of the covariance matrix contains an error. However, there is a distinction between estimation error and specification error. The estimation error occurs when there are not enough degrees of freedom per estimated parameter. This typically occurs when the number of observations in the sample is not big enough compared to the number of the estimated parameters. As stated by Pafka et al. (2004), the simple sample covariance matrix estimator often suffers from the “curse of dimensions”.

The second source of error is specification error. This type of error occurs when some form of structure is imposed on the model that is being used in the estimation process. As a result, the estimator becomes too specific in comparison with reality. Therefore, there exists a trade-off between the estimation error and the specification error. Clearly, an improved estimator must ideally reduce the large estimation error without creating too much specification error.

All shrinkage techniques date back to Stein (1955) seminal work. At the core, Stein estimator is a Bayesian statistical procedure which assumes a “prior” or a structure such that the estimated parameters can shrink toward it. In the context of estimating a covariance matrix, the estimation error in a sample matrix can be reduced by shrinking the sample matrix towards an existing prior covariance matrix. The prior can either be derived from an assumption or from a model.

The shrinkage estimator is often computed as a weighted average between the sample covariance matrix and the prior covariance matrix (or the shrinkage target). The weight assigned to the prior matrix is known as the shrinkage intensity parameter and is usually solved by minimizing a quadratic error function of the combined estimator. If the prior covariance matrix is invertible, then the shrinkage estimator will always be invertible.

Ledoit and Wolf (2003) develop a shrinkage estimator that is a weighted average of the sample covariance matrix and the covariance matrix implied by the market model of Sharpe (1963). In this case, the market model is the imposed structure or the prior and the sample covariance matrix shrinks toward the structure model. The resulting optimally weighted average matrix of the two supposedly minimizes the estimation error. Denoting the estimated covariance matrix implied by the single-index market model as **F** and as $T \rightarrow \infty$, it is assumed to converge to $\mathbf{\Omega}$. Also denote **S** as the sample covariance and as $T \rightarrow \infty$, it is assumed to converge to $\sum$ for which $\mathbf{\Omega} \neq \sum$. In other words, **F** is an asymptotically biased estimator. The shrinkage estimator can be written as:

$$\alpha\mathbf{F} + (1-\alpha)\mathbf{S} \tag{26}$$

The assumption made is that asset returns are independent and identically distributed (iid) and that they have finite fourth moments. For a fixed N, **S** will be consistent, while **F** is not. Thus, the shrinkage intensity should vanish asymptotically (as $\rightarrow \infty$). In order to solve for the optimal α, Ledoit and Wolf employs the quadratic loss function

$$L(\alpha) = ||\alpha\mathbf{F} + (1-\alpha)\mathbf{S} - \sum||_F^2 \tag{27}$$

The key aspect of this approach is that while the optimal solution α* is complex the procedure does not depend on inverting the covariance matrix. Therefore, this shrinkage estimator does not break down when $N \geq T$.

Jagannathan and Ma (2000) use the concept of a portfolio of covariance estimators. A ***portfolio of estimators*** is an estimator consisting of an equally weighted average of the sample matrix and several other estimators of the covariance matrix whose diagonal elements are the sample variance and at least one of them is invertible.[1] The portfolio approach also builds on the tradeoff between estimation and specification error. By averaging the sample matrix with other estimators whose primary error is specification error, an improved covariance matrix can be obtained. Jagannathan and Ma (2000) basically extend the Ledoit-Wolf's model by adding a third matrix that is the diagonal part of the sample covariance matrix but using simple **equal weights** among the three matrices. In this case, the technique is less complex than that of the shrinkage estimator method because it does not require solving for optimal shrinkage parameters. Their model can be represented by:

$$\frac{1}{3}\mathbf{F} + \frac{1}{3}\mathbf{S} + \frac{\mathbf{1}}{\mathbf{3}}\mathbf{D} \tag{28}$$

where **F** and **S** are identical in (26) and

$$\mathbf{D} = \begin{bmatrix} S_{11} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & S_{NN} \end{bmatrix}$$

In essence, their model contains a one-factor model (**F**), an N-factor model (**S**) and a **zero** factor model (**D**). They argue that the equal weight is the safest bet because little is known about the covariance structure of the estimation errors of the different estimators.

[1] A weighted average of two matrices, one of which is invertible, is also invertible.

In a later paper, Jagannathan and Ma (2003) show that imposing constraints on the portfolio weights such as no short selling and upper bounds, may sometimes prove to reduce out-of-sample volatility of the investor's optimal portfolio. No short selling constraint is realistic in that it is often very hard for an ordinary investor to short sell an asset. In fact, no short selling constraint can be interpreted as shrinking the largest variances and covariances, which cause the negative weights, towards more standard values, since it can be argued that these extreme estimates are those most likely caused by estimation error. In order to appreciate their argument, consider the case of minimizing a portfolio's risk with no short selling constraint:

$$\min_w \{\tfrac{1}{2}\, \boldsymbol{w^T \widehat{S}\, w} \mid \boldsymbol{w^T 1} = 1,\ \boldsymbol{w} \geq 0\} \qquad (29)$$

whereby the solution for the MVP is given by $\boldsymbol{w}^*(\widehat{\boldsymbol{S}})$**.** Denoting δ_i as the Lagrangian multiplier associated with the non-negative constraint, the authors derived the unconstraint covariance matrix as:

$$\breve{\boldsymbol{S}} = \widehat{\boldsymbol{S}} - \left[\begin{matrix} 2\delta_1 & \delta_1 + \delta_2 \cdots & \delta_1 + \delta_N \\ \vdots & \ddots & \vdots \\ \delta_N + \delta_1 & \delta_N + \delta_2 \cdots & 2\delta_N \end{matrix} \right] \qquad (30)$$

and the solution for the MVP is given by $\boldsymbol{w}^*(\breve{\boldsymbol{S}})$**.** In this case, the variance of **S** is reduced by $2\delta_i$ and the off-diagonal is reduced by $\delta_i + \delta_j$. Thus, the new covariance matrix $\breve{\boldsymbol{S}}$ is constructed by shrinking the largest elements in the initial covariance matrix estimate towards more standard values.

Bengtsson and Holst (2002) recognize that estimators based on different assumptions make errors in different directions. They combine the shrinkage method (Ledoit-Wolf) and the predictions from random matrix theory (Pleoru et al. 2001). They essentially replace the covariance matrix implied by the market model with a ***K* factor** principal component covariance matrix, $\mathbf{P_K}$, where the choice of the number of factors K is based on random matrix theory. In fact, K is the number of eigenvalues of the sample correlation matrix that deviates significantly from the maximum eigenvalue prediction of λ_{max} as defined in (25).[2] The principal component matrix $\mathbf{P_K}$ is the shrinkage target and the shrinkage estimator is given by:

$$\alpha \mathbf{P_K} + (1-\alpha)\mathbf{S} \tag{31}$$

where the optimal α is solved using the Ledoit-Wolf (2003) procedure.

The last methods of reducing estimation error (or noise) are those proposed by the studies in covariance matrix using RMT. In this section, I explore two methods in the literature that deal with "cleaning" the noise from the sample correlation matrix $C_{original}$.

---

[2] $\lambda_{min}^{max} = 1 + \frac{1}{Q} \pm 2\sqrt{\frac{T}{Q}}$ is in Chapter 1, equation (25).

The first method treats the noisy eigenvalues as containing no useful information. As a result of the assumption, the eigenvalues are either ignored or "flattened" out. The second method examines the eigenvectors associated with the noisy eigenvalues. In particular, the overlapping part of eigenvector may contain useful information and may affect the stability of the transformed covariance matrix.

The first method of filtering or cleaning the noise of the correlation matrix is that proposed by Bouchaud and Potters (2000). They first separate out the "noisy" part from the "non-noisy" parts of the correlation matrix *C*. The noisy part is comprised of the eigenvalues that conform to the properties of randomness as predicted by RMT. The non-noisy part or the "information" part is that set of eigenvalues that deviates from RMT predictions.[3] The whole idea is to obtain a background measure of the noise element while retaining the information trace; based on the fact that the eigenvalues corresponding to the noise band are not expected to contain real information, they are all equally useless.

One obvious implication is that to ignore all eigenvalues within the noisy band, i.e., set them to zeros. However, many studies are unwilling to do so because the noise band may still contain some correlation information. Bouchaud and Potters (2000) suggestion is to keep the non-noisy eigenvalues the same and to flattening each eigenvalue associated with the noisy part by the average of those eigenvalues. Finally,

[3] As shown previously in Chapter 1, the RMT prediction is that segment of the distribution of eigenvalues that is bounded by the theoretical minimum and maximum eigenvalues.

they reconstruct the correlation matrix, C_{clean}, from the cleaned eigenvalues and the original eigenvectors using the following

$$C_{clean} = VD_{clean}V \tag{32}$$

where V is the matrix of eigenvectors and D_{clean} is the reconstructed diagonal matrix of eigenvalues by replacing the noisy eigenvalues with the average eigenvalues.

The authors used 600 data points for 200 stocks from S&P 500 intraday data to test whether the cleaned correlation matrix is better in predicting risk. They divided the period into two sub-periods. They calculate the efficient frontier for the first sub period using the ***actual return*** on the second sub-period and the correlation matrix from the first sub-period. In other words, they use the actual return from the second-period as the first-period expected return in the MV optimization. The assumption made is that the investor has "perfect" foresight in predicting the future average returns. The efficient frontier is called the *prediction* of the portfolio and the associated risk as *predicted risk*.

The *realized* risk and return are computed using the second-period correlation matrix and returns with the weights of the same family of portfolios as the predicted ones. Bouchaud and Potters (2000) argued that the predicted and realized risks are closer when the cleaned matrix is used in delineating the efficient frontier. According to the authors, the closeness of the predicted and realized curves is due to the power of C_{clean} in predicting the future risk. Hence, they conclude that the stability of C_{clean} is higher than the stability of $C_{original}$.

Jolliffe (1986) argued that eigenvectors and principal components can only be confidently interpreted if they are stable. Therefore, the issue of **stable** correlation matrix is important following this argument. Sharifi et. al (2004) argue that the cleaning method as suggested by Bouchaud and Potters (2000) does not improve the stability of the cleaned correlation matrix. In fact, the opposite effect of reducing stability is created by the cleaning method. The issue is how to determine the stability of C_{clean} after cleaning it. The goal clearly is to remove noisy elements from $C_{original}$ in such as way that maximum stability is conserved.

Fortunately, the works done by Laloux et. al (2000) and Lee (2001) indicate that the overlap of the eigenvectors of two consecutive time sub-periods determines the consistency (or convergence) of the eigenvectors. This follows from the fact that the dot product of two normalized vectors represents the cosine of the angle between them and gives a measure of overlap. The cosine value should be large if the directions of the eigenvectors remain similar over the two sub-periods. Small cosine values indicate less overlap and thus less stability.

Sharifi et. al (2004) introduce the second method of cleaning the original correlation matrix which is based on the Krzanowski (1984) technique. According to their argument, the stability does not depend on the absolute size of eigenvalues alone, but on the separation or distance between eigenvalues' size. They employ a principal component technique to measure the stability of the matrix and its eigenvectors. The information on the stability of the principal components can be seen from the effect on the kth eigenvector v_k of small changes in the associated eigenvalue $\lambda_{k.}$ By examining the perturbation of an eigenvector derived for a small increase / reduction, ϵ, in the corresponding eigenvalues, they can find the component $v(i)$ that diverges the most from the ith eigenvector, v_i, but still has an eigenvalue which is at most ϵ greater / less than that of v_i. The angle θ between $v(i)$ and v_i can be calculated by

$$\cos\theta = \begin{cases} \left(1 + \frac{\epsilon}{\lambda_i - \lambda_{i+1}}\right)^{-1/2} & \epsilon \; decreased \; from \; \lambda_i \\ \left(1 + \frac{\epsilon}{\lambda_{i-1} - \lambda_i}\right)^{-1/2} & \epsilon \; increased \; to \; \lambda_i \end{cases} \tag{33}$$

where $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq ..... \geq \lambda_n$. Clearly, the effect on v_i of an ϵ change in λ_i is an inverse function of $\lambda_i - \lambda_{i+1}$. Therefore, it is the relative separation of eigenvalue size from the next component that determines the overlap and stability of the matrix. The cleaning method proposed by Sharifi et. al is to replace the noisy eigenvalues with components that have maximal separation from each other while maintaining a fixed sum.

The next section discusses the findings on these error reducing covariance estimators. In one way or another, each study claims that their estimators are better. The first study is that of Ledoit and Wolf (2003). Their shrinkage estimator to the Single-Index model can be seen as a way to account for extra-market covariance without having to specify an arbitrary multi-factor structure. They found that for NYSE and AMEX stock returns from 1972 to 1995, their shrinkage estimator can be used to select portfolios with significantly lower out-of-sample variance than a set of existing estimators, including multi-factor models.

In a follow up study, Ledoit and Wolf (2004) use a shrinkage estimator to a constant correlation model – in which the covariance matrix estimator is obtained by assuming that each pair of stocks has the same correlation joins the sample matrix in the weighted average. They find the performance of this estimator comparable to the performance of the shrinkage to the single-index model estimator.

The portfolio approach as suggested by Jagannathan and Ma (2000), which is an equally weighted average of the sample matrix, the single index matrix and the diagonal matrix, was found to be one of the best performers in Bengtsson and Holst's (2002) study of the Swedish stock market.

Jagannathan and Ma (2003) argue that when no-short-sale constraints are imposed in MV optimization, the sample covariance matrix performs as well as covariance matrix estimates based on factor models, shrinkage estimators, and daily data. This conclusion is in direct contrast to Green and Hollifield (1992), who argue that the presence of a dominant factor would result in extreme negative weights in the MV efficient portfolios even in the absence of estimation errors. Another result from Jagannathan and Ma (2003) shows that tangency portfolios, whether constrained or not, do not perform as well as the global MVPs in terms of the out-of-sample Sharpe ratio, implying that the estimates of mean returns are so noisy as to distort the expected results. They also found that when short sales are allowed, MVP and minimum tracking error portfolios constructed using daily return covariance matrix performs the best.

Disatnik and Benninga (2007) ran a “horse race” between various shrinkage estimators and portfolios of estimators. Using the ex-post standard deviation of the global MVP as their betterment criterion, they found no statistically significant gain from using more sophisticated shrinkage methods and thus recommend the simpler portfolio of estimator. This is when no short sale constraint is imposed. But a big drawback is that the global MVP consists of large short sale positions, which is clearly an undesirable feature of portfolio optimization. When short sale constraint is imposed, both the shrinkage estimator and the portfolio of estimators perform statistically significantly better than the sample matrix.[1]

Over the years, there has been a big flaw in many of the empirical studies on covariance estimation and MV optimization. Most of the existing studies, including those mentioned above, use a less frequent time series, such as monthly return data, T, with a much larger number of stocks, N, resulting in a, Q=T/N, ratio being typically less than 1. Cohen and Pogue (1967) use ten annual return data to estimate the covariance of 75 and 150 stocks. Elton and Gruber (1973) use 60 monthly returns to estimate a covariance matrix of rank 76. Eun and Resnick (1992) use 84 monthly data to estimate a covariance matrix of 140 stocks. Chan et al. (1999) use 60 monthly returns for 250 stocks; Ledoit and Wolf (2003) use 120 monthly return data for 1,000 stocks; Jagannathan and Ma (2003) use 60 monthly data to cover 500 stocks. The Q ratios range from 0.07 to 0.79 in these studies.

Is the flaw really that serious so as to invalidate the findings in these studies? There are four reasons that many of their conclusions may be suspect. First, when N is of the same order of magnitude as the number of historical returns per stock T, the total number of parameters to estimate is of the same order as the total size of the data set. There is zero degree of freedom. This can lead to degenerate solutions in the MV optimization. How is it possible that almost all the studies that have insufficient data is able to construct a global MVP?[4]

[4] Both Shrinkage and Portfolio of estimator went around the singularity problem by combining the sample matrix with a non-singular matrix.

Second, under normality assumption of returns, the sample covariance matrix has the desirable maximum likelihood property. It means letting the data speak for itself as to the most likely parameter values. Maximum likelihood is justified asymptotically as the number of observations per variable goes to infinity. It is a general disadvantage of the maximum likelihood method as it performs badly in small sample. For the covariance matrix, small sample problems occur unless T > N.

Third, Pafka and Kondor (2003), (2004) use a simulation-based approach to systematically compare the relative performance of different correlation matrix estimators. They find that T/N is indeed an important factor that influences the relative performance of alternative correlation estimation methods. Liu and Lin (2010) provide empirical support to their findings.

Finally, in the most extensive test of 14 estimation models and across seven datasets, DeMiguel et.al (2009) conclude the following:

> "Based on parameters calibrated to the US equity market, our analytical results and simulations show that the estimation window needed for the sample-based mean-variance strategy and its extensions to outperform the 1/N benchmark is around 3000 months for a portfolio of 25 assets and about 6000 months for a portfolio with 50 assets"

Chapter 3

Characteristics of “Filtered Correlation Matrix”

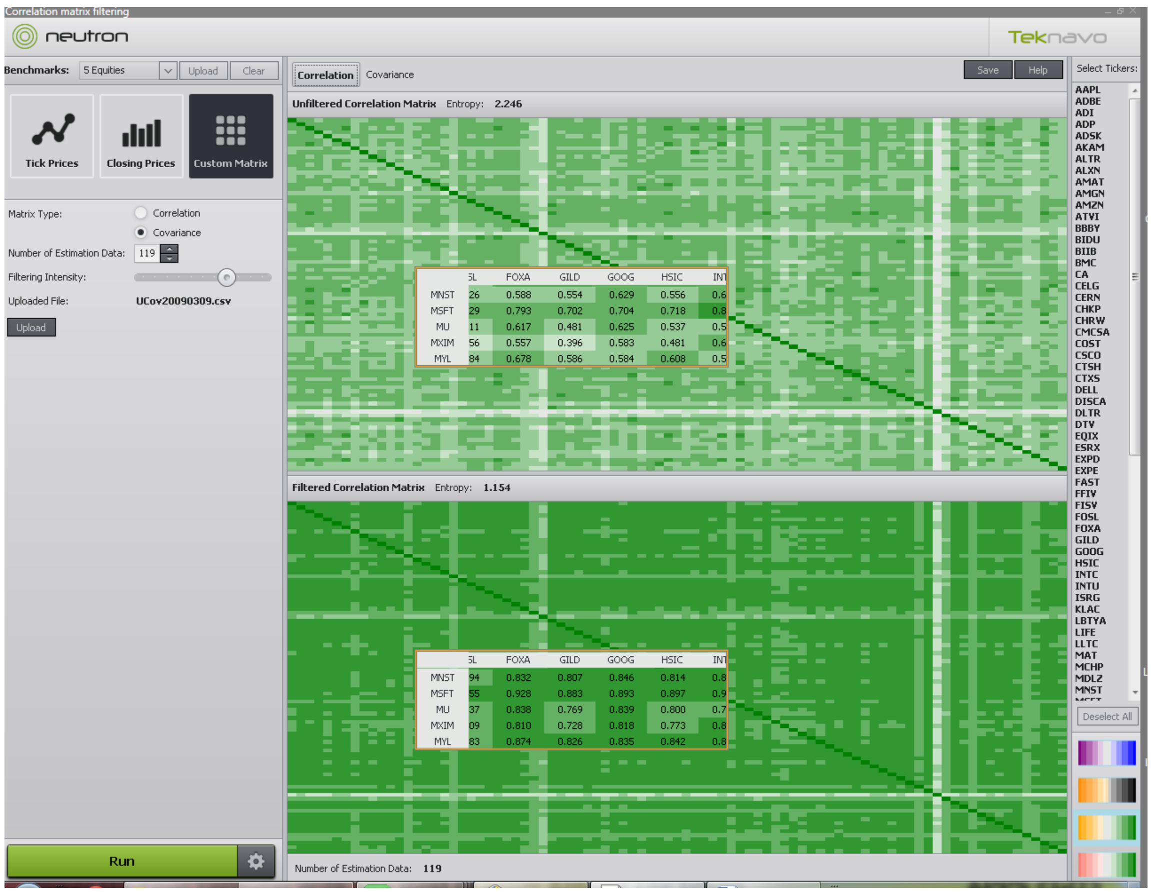
The primary aim of this chapter is to introduce a commercially available product that claims to filter the noise contained in a covariance or correlation matrix, along the lines of research in Random Matrix Theory. The secondary aim is to examine the characteristics of the cleaned matrix vis-à-vis the raw sample correlation matrix. This will, hopefully, lead to a better understanding of how it may affect the portfolio selection procedure.

The first commercially available product that is using the studies of RMT is called “Neutron QunatumApp”[1]. This product was built and released in June of 2013 to Bloomberg’s users on a monthly subscription basis. It is a third-party product that utilizes Bloomberg’s infrastructure and data. The user can access this application by typing “APPS Neutron <GO>” in the Bloomberg terminal. The product taps into the Bloomberg data source either through a real-time basis or through historical data. The user can search for ticker symbol of stock, bond, index or economic variable using the tool, save the portfolio to a file, and reload the file containing the tickers. Alternatively, a user can upload a custom generated covariance or correlation matrix into the Neutron application

[1] As far as this author is aware of.

and define the N and T parameters. There is one other parameter that is required – the filtering intensity parameter that ranges from 0 to 1. A value of zero means no filtering of estimation noise, 1 means maximum filtering, and value in between measures the intensity required. Figure 1 illustrates an example of the Neutron Application. The top half is the raw sample correlation matrix while the lower half is the filtered correlation matrix.

Figure 1 Neutron Screen Shot



The green color represents positive correlation; yellow color indicates negative correlation and white color indicates low correlation. Darker color shade implies higher correlation values.

The brochure and white papers relating to the Neutron product make many claims. Some of these claims include:[2]

1. "Neutron provides stable estimates (clean statistical inputs) for increased Profitability/ Sharpe Ratio (by a factor of 2-3)"

2. "Dynamically hedge short or long positions with securities reliably and highly, positively or negatively, correlated with such positions".

3. "Diversify alpha portfolios by adding positions in those names that are reliably and highly correlated with securities in such portfolios".

4. "Trade option strategies more reliably given the high sensitivity to volatility of the underlying securities' returns".

5. "Reveals up to 30% increased volatility post noise-filtering"

6. "Efficiently allocate capital for substantially higher returns, at a fixed risk, or substantially lower risk, at a fixed return".

7. "Reliably hedge any portfolio by identifying those ETF(s) and indices that are now reliably and highly correlated or anti-correlated with a majority of names in the portfolio".

8. "Reveals up to 200-300% higher returns (at fixed risk) and Lower Volatility (at fixed return) post noise-filtering"

9. "Assess VaR and Shortfall with substantially increased accuracy using Noise-filtered covariances".

---

[2] See "Introducing Neutron from Teknavo Quantum Apps Product Suite", 2013, by Teknavo / Market Memory Trading Joint Venture. Refer to www.teknavo.com/en/applications/neutronbrochure.pdf The White papers can be obtained via www.teknavo.com/en/applications/whitepapers.php

10. "Rebalance departmental portfolios in order to significantly reduce VaR and optimize shortfall"

The exact methodology used by the Neutron product in filtering the estimation error (or noise) is not reveal because of its proprietary nature. Nevertheless, the literature that was discussed in previous chapters has pointed to three methods of dealing with the eigenvalues that conform to the predictions of RMT. One of the earliest studies simply assigns these smaller eigenvalues zero values in the belief that there is no useful information associated with these eigenvalues. Only the larger eigenvalues explain most of the variability of the correlation matrix and thus contain the real information in the correlation structure. Another study suggests setting these eigenvalues to their average value, thus flattening the correlation structure for the smaller eigenvalues. Another study suggests using a stability measure to reset the eigenvalues.

It is not certain as to which of the three methods Neutron is adopting. Perhaps it may adopt an entirely different approach. Due to the complexity of the method, my purpose here is to examine the validity, the usefulness and practicality of this commercial product. Since it is an easy to use application and it has the advantage of working with custom generated matrices, my goal is to adopt this complex and advanced technique in order to examine the performance and usefulness of "cleaned" correlation matrix vis-à-vis the raw sample matrix.

The sample data used throughout this dissertation comes directly from Bloomberg. The sample interval is from 3-Jan-2006 to 30-Aug-2013 and consists of 1,929 daily closing prices. This period contains the financial crisis of 2008. It is crucial to examine portfolio behavior during this period. The first data set consists of 87 stocks listed under the NSDAQ 100 Index. This sample is further divided into 4 random portfolios. The first three portfolios contain 20 randomly selected stocks from the 87 total samples; the fourth portfolio has the remaining 27 stocks. In the second sample set, 80 stocks were randomly selected from the S&P 500 Index universe. These 80 stocks in the S&P sample do not overlap with the stocks in the NASDAQ sample. Again, the sample is divided into 4 random portfolios, each containing 20 stocks. Therefore, my study consists of 10 portfolios – 2 large portfolios (with $N \geq 80$) and 8 smaller portfolio (with $N \geq 20$). It is important to see if there is consistency in the performance and behavior of these correlation matrices irrespective of sample size. The number 20 was chosen because this usually corresponds to the minimum number of stocks required in portfolio diversification. The aim is not simply to include very large numbers of securities, as many studies have done, but to analyze the consistency of the correlation given a relatively fixed Q=N/T ratio.

The log return $\ln(P_t / P_{t-1})$ is used to generate the return series of each stock. For every 10 trading days (approximately 2 weeks), the sample covariance matrix is estimated for the entire NASDAQ sample and for the S&P stock sample.[3] The number of

[3] The NASDAQ 100 Index is included in the covariance matrix generated. Thus there are 88 stocks in the matrix. Similarly, the S&P 500 Index is included in the other sample covariance matrix (81 stocks).

log returns used to compute the covariance matrix is fixed at T=119 (6-month of data points). Thus, a total of 180 sample covariance matrices were created for each market. The raw covariance matrix is then loaded into Neutron using the "Custom Matrix" feature and the intensity parameter is set to 0.7. After the cleansing process, both raw and filtered covariance and correlation matrices were downloaded in Excel files. In total, I am working with 180 raw and 'filtered" pairs of covariance matrices.

The full sample period (June-23-2006 to Aug-19-2013) of 180 correlation matrices is also divided into three sub-periods of non-overlapping 60 correlation matrices. Period 1 is from June-23-2006 to Oct-27-2008. Period 2 is from Nov-10-2008 to Mar-16-2011. Finally, period 3 is from Mar-30-2011 to Aug-19-2013. The first analysis is to examine the histogram of the raw correlations vs. the filtered correlations. The elements in the half matrix of the correlation matrix are divided into 12 buckets – from the lowest correlation value of -0.02 to the highest 0.9. Each bucket represents the number of stocks with correlation ≥ the assigned bucket value and less than the next bucket value. For example, a correlation coefficient of 0.05 falls into the 0 bucket, 0.601 falls into the 0.6 bucket and 0.99 falls into the highest bucket of 0.9. Ignoring the diagonal values of 1.0, the total number of elements in the half diagonal correlation matrix is N(N—1)/2. The number in each bucket is then converted into frequency or the percentage of total elements.

Table 3.1 shows the frequency table of the correlations for the NASDAQ sample. The full period and the three sub-period results are reported along with the average entropy values for the "raw" and "filtered" cases. The frequency plots corresponding to the full period and sub-periods are shown in Figure 3.1A, 3.1B, 3.1C and 3.1D.

Table 3.1 – Frequency Table of Correlation for NASDAQ sample

	NASDAQ Sample											
Full Period	**0.9**	**0.8**	**0.7**	**0.6**	**0.5**	**0.4**	**0.3**	**0.2**	**0.1**	**0**	**-0.1**	**Entropy**
Raw	0.0%	0.6%	3.3%	8.8%	13.6%	18.2%	21.5%	18.6%	10.7%	3.8%	0.7%	**2.978**
Filtered	2.9%	21.1%	26.4%	20.7%	13.4%	7.8%	4.3%	2.0%	0.8%	0.3%	0.1%	**1.742**
2006 to 2008												
Raw	0.0%	0.0%	0.5%	2.1%	6.2%	15.2%	25.4%	26.8%	16.6%	5.8%	0.0%	**3.280**
Filtered	0.6%	9.2%	25.0%	25.3%	18.1%	11.3%	6.3%	2.6%	1.1%	0.4%	0.0%	**1.969**
2008 to 2011												
Raw	0.0%	0.8%	4.8%	14.4%	21.0%	21.9%	18.2%	11.3%	5.4%	1.9%	0.1%	**2.728**
Filtered	4.0%	30.9%	29.2%	17.0%	9.3%	4.8%	2.6%	1.3%	0.5%	0.3%	0.0%	**1.547**
2011 to 2013												
Raw	0.0%	0.9%	4.7%	9.8%	13.7%	17.6%	20.8%	17.7%	10.2%	3.6%	0.0%	**2.926**
Filtered	4.1%	23.2%	25.0%	19.8%	12.9%	7.4%	4.1%	2.1%	0.9%	0.4%	0.3%	**1.710**

Figure 3.1A (NASDAQ sample – Full Period)

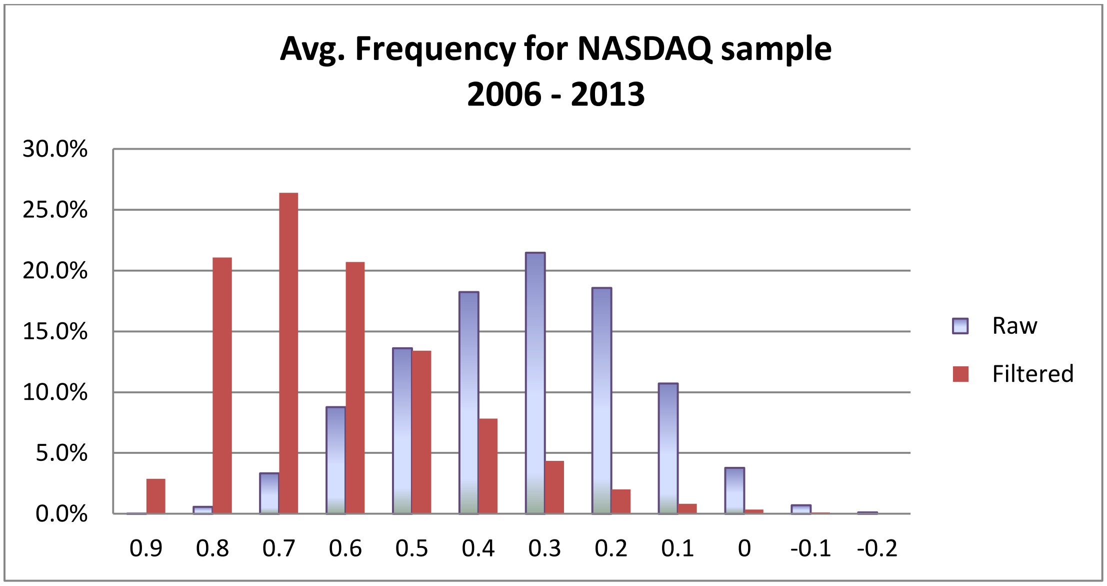


Figure 3.1B (NASDAQ Sample - Period 1)

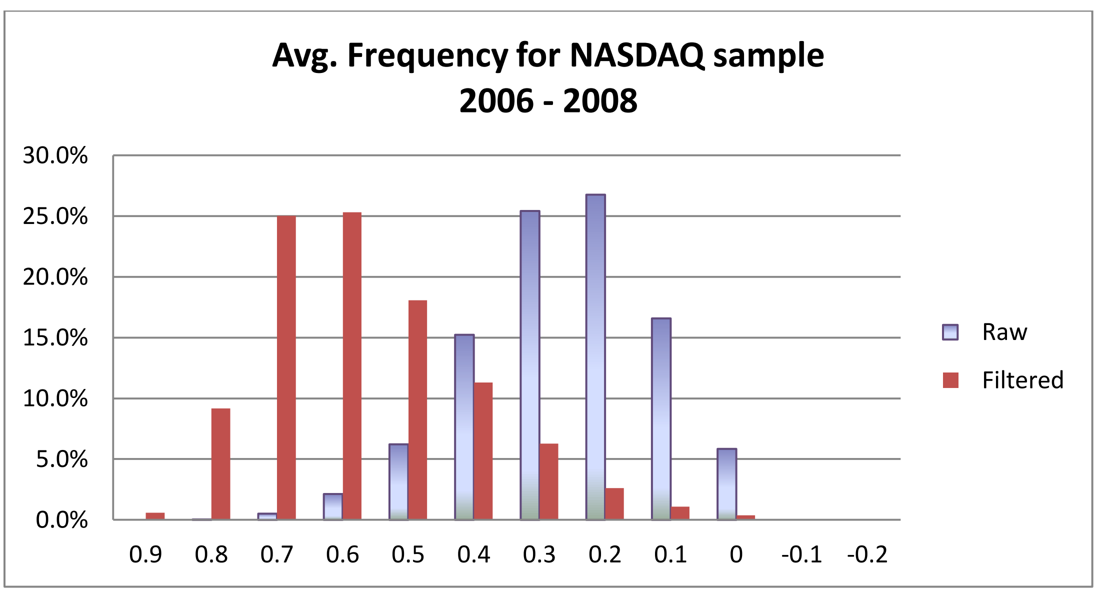


Figure 3.1C (NASDAQ Sample - Period 2)

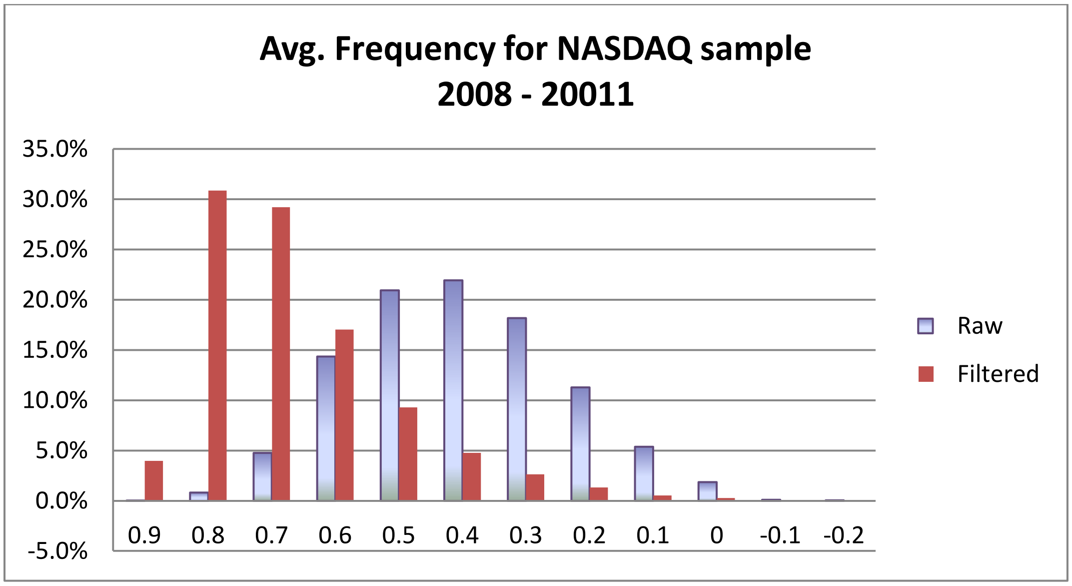
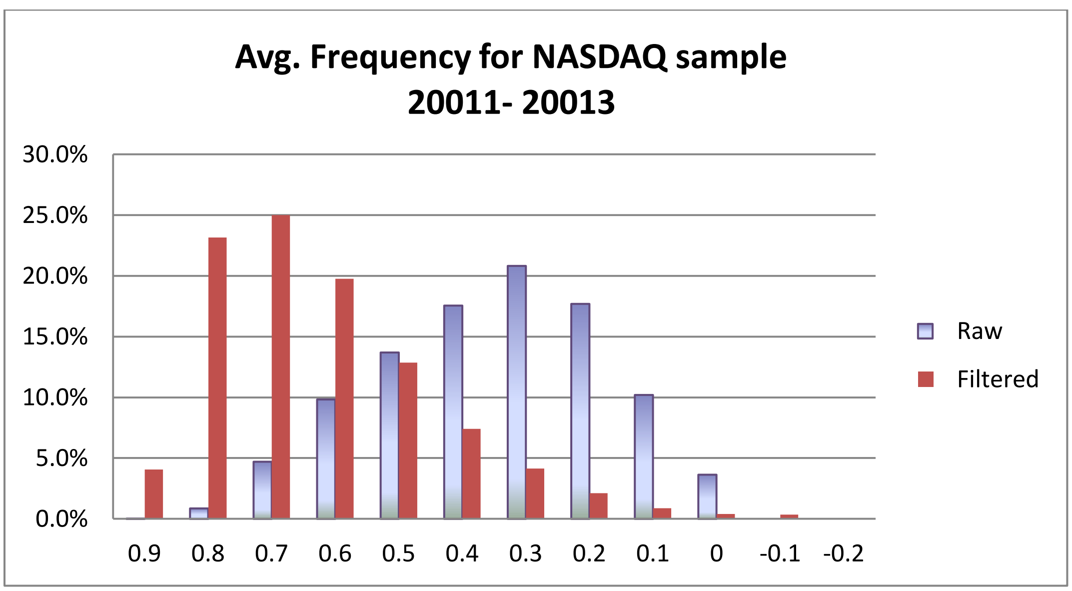


Figure 3.1D (NASDAQ Sample - Period 3)



The most noticeable difference between the distributions of the raw correlation and the filtered correlation is that the former is more centered between 0.3 and 0.4 whereas the latter is more centered between 0.6 and 0.7. The filtered correlations are skewed more towards the higher correlation values than the raw correlation. In fact, less than 1% of the raw correlations ever exceed 0.8. In comparison, more than 20% of the filtered correlations exceed 0.8. During the second period, where the majority of the correlations are estimated using prices that straddle the financial crisis period, the average frequency of correlation exceeding 0.8 is around 35% for the filtered case, whereas it is only 0.8% for the raw case. The filtering process was able to reduce the impact of estimation noise, thus resulting in more realistic and stronger correlated relationship among stocks; especially during periods of high systemic risk. In addition, the reduction in the Shannon Entropy, which is a measure of noise, ranges between 66% and 76%. (See Appendix A for a brief discussion of Shannon Entropy).

Table 3.2 presents the frequency table of correlation for the S&P sample. The full period and the three sub-period results are reported along with the average entropy values for the “raw” and “filtered” cases. The frequency plots corresponding to the full period and sub-periods are shown in Figure 3.2A, 3.2B, 3.2C and 3.2D.

Table 3.2 Frequency Table of Correlation for S&P sample

Full Period	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1	0	-0.1	Entropy
Raw	0.0%	1.2%	4.8%	9.7%	14.3%	17.5%	18.9%	16.7%	10.8%	4.5%	1.2%	**2.860**
Filtered	3.8%	20.8%	23.2%	18.4%	13.1%	8.7%	5.4%	3.1%	1.7%	1.0%	0.5%	**1.742**
2006 to 2008												
Raw	0.0%	0.3%	1.3%	3.6%	8.5%	15.9%	21.6%	22.2%	16.2%	7.5%	0.0%	**3.124**
Filtered	1.0%	7.8%	17.8%	21.3%	17.9%	13.1%	8.6%	5.2%	3.2%	2.0%	0.0%	**2.026**
2008 to 2011												
Raw	0.0%	1.3%	6.0%	14.9%	21.2%	20.9%	16.8%	11.1%	5.5%	1.9%	1.2%	**2.635**
Filtered	4.3%	30.7%	29.2%	16.4%	9.1%	5.2%	2.7%	1.3%	0.7%	0.3%	0.0%	**1.521**
2011 to 2013												
Raw	0.1%	1.9%	7.0%	10.7%	13.3%	15.7%	18.3%	16.9%	10.9%	4.3%	0.0%	**2.822**
Filtered	6.0%	23.8%	22.6%	17.4%	12.3%	7.8%	5.1%	2.7%	1.3%	0.5%	0.4%	**1.680**

Figure 3.2A (S&P Sample - Full period)

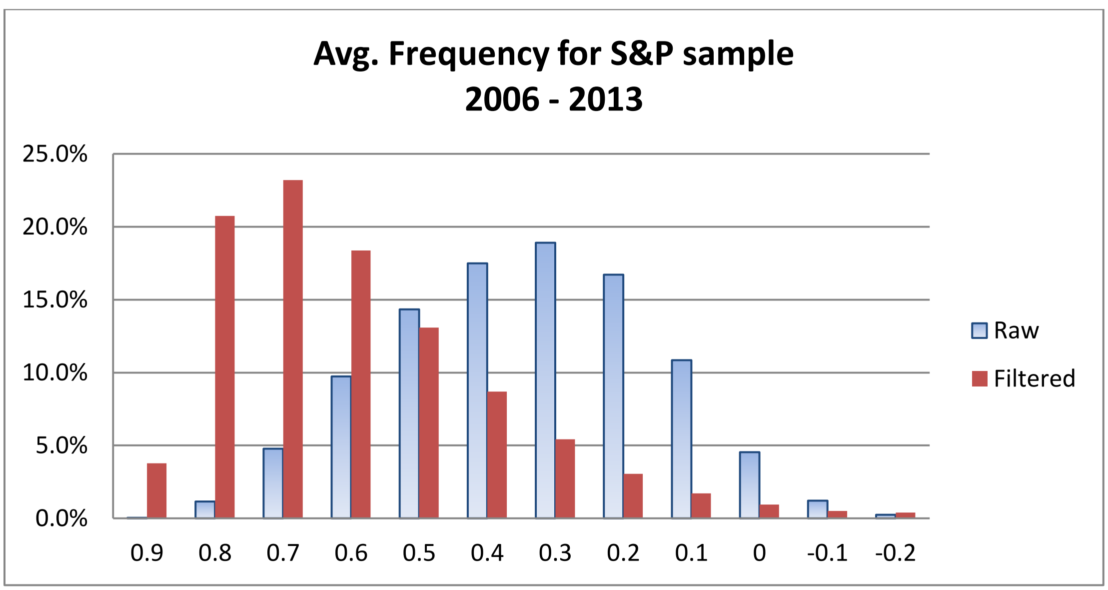


Figure 3.2B (S&P Sample - Period 1)

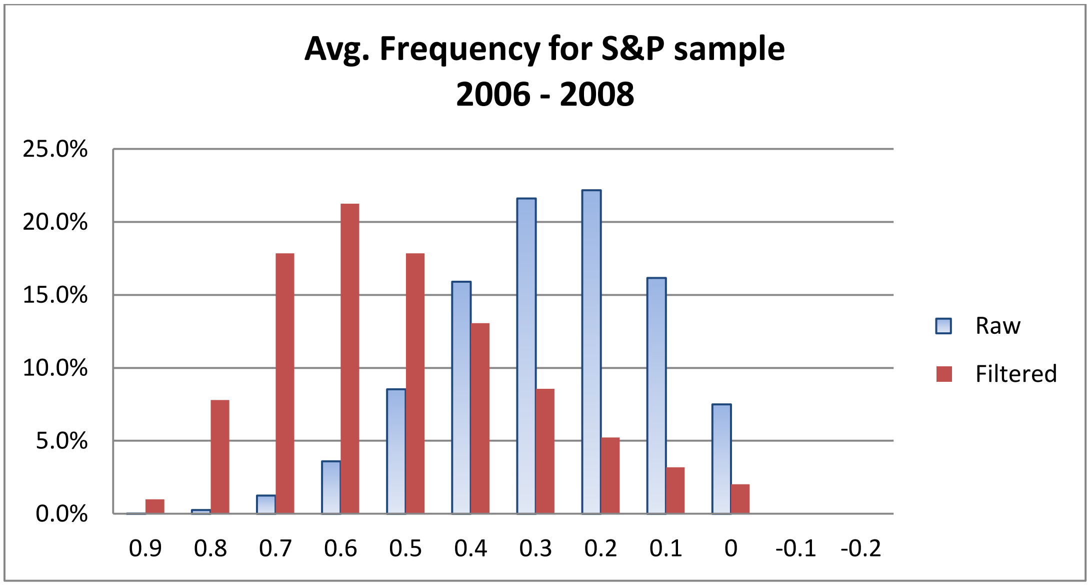


Figure 3.2C (S&P Sample - Period 2)

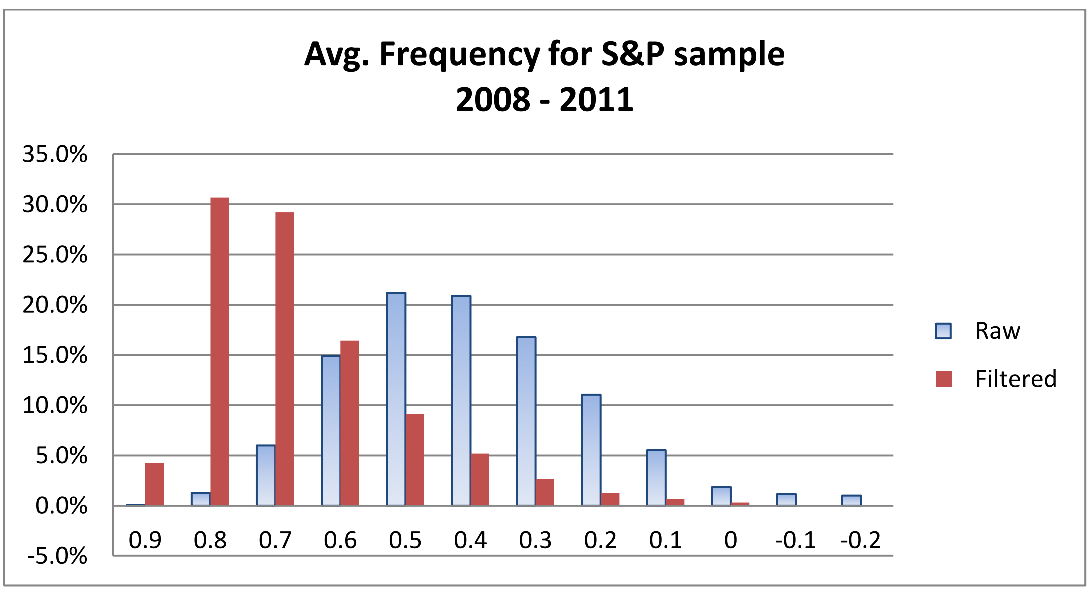
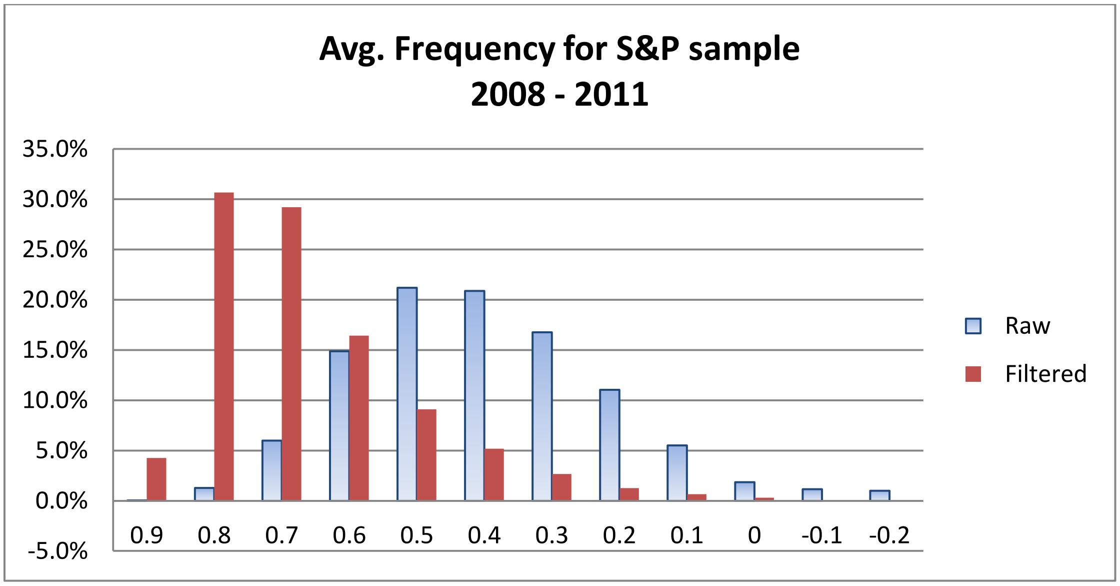


Figure 3.2D (S&P Sample - Period 3)



The results reported for the S&P sample are very similar to those for the NSADAQ sample. Again the filtered correlations are skewed towards the higher values whereas the raw correlations appear to be centered between 0.3 and 0.4. During the high volatility second period, the average frequency of correlation exceeding 0.8 is around 35% for the filtered case whereas it is only 1.3% for the raw case. The reduction in the Shannon entropy ranges from 54% to 73%.

The implication of these results indicate that estimated raw correlations, in general, are more evenly distributed and lower in value. By modifying the eigenvalues in the predicted random band, the filtered correlations are "stretched" towards higher values and more realistic correlation structure that is consistent to volatility in the market. This is consistent with the fact that a positive real eigenvalue is in fact a scalar or a stretching factor.

The next analysis is to compare the variance of each stock prior to and after the filtering process. First, the variance of each stock is extracted from the raw sample covariance matrix. This is denoted as $\sigma^2_{j,Raw.}$ Next, this variance is then compared to the variance extracted from the filtered covariance which is denoted by $\sigma^2_{j,Filter.}$ If the Neutron's filtering process was to produce lower risk measure (in terms of volatility), then one would expect overwhelmingly lower $\sigma^2_{j,Filter}$ in comparison to the raw variance. Define the following change in variance as a percentage of the filtered variance as:

$$\Delta\sigma_j^2 = (\sigma^2_{j,Raw} - \sigma^2_{j,Filter})/\sigma^2_{j,Filter} \quad \text{for j = 1,…., N securities}$$

If risk is reduced, one expects positive value for $\Delta\sigma_j^2$, otherwise it will be negative. After completing the above change in variance, I then average these changes across stocks in the sample portfolio. The median is also computed for comparison purposes. The results for the NASDAQ sample of 87 stocks are reported here.

Figure 3.3 shows the time series **average** of $\Delta\sigma_j^2$ across the 87 stocks. That is

$\Delta\bar{\sigma}_t^2 = \sum_j^N \Delta\,\sigma_{jt}^2/N$ *for t = 1,…,T.* The average values are all positive falling between 2.334% and 13.554%. It appears that the filtering process reduces the volatility across stocks. However, this is not necessarily true for the median values as displayed in Figure 3.4. There are periods when the post-filtered variance actually increased.

Figure 3.3 Average change in Variance

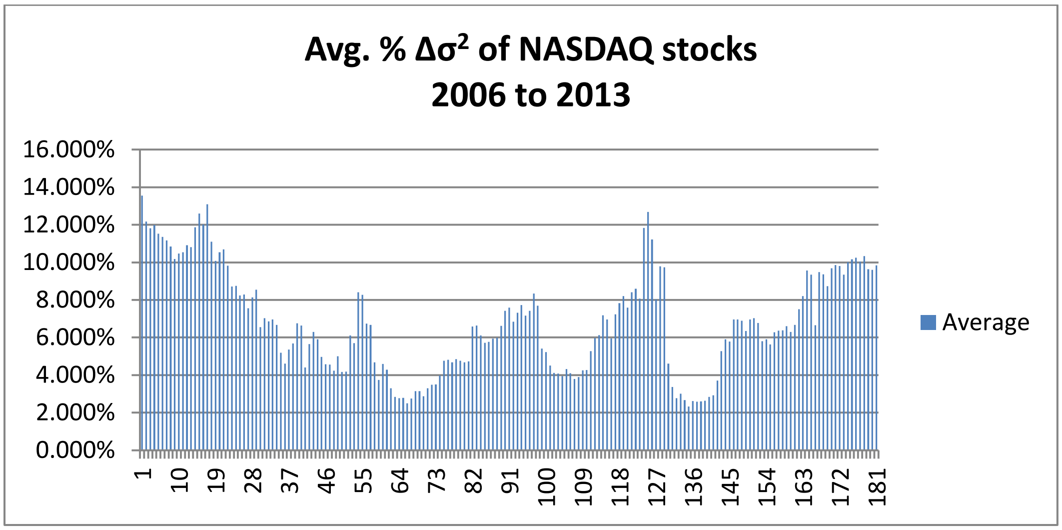
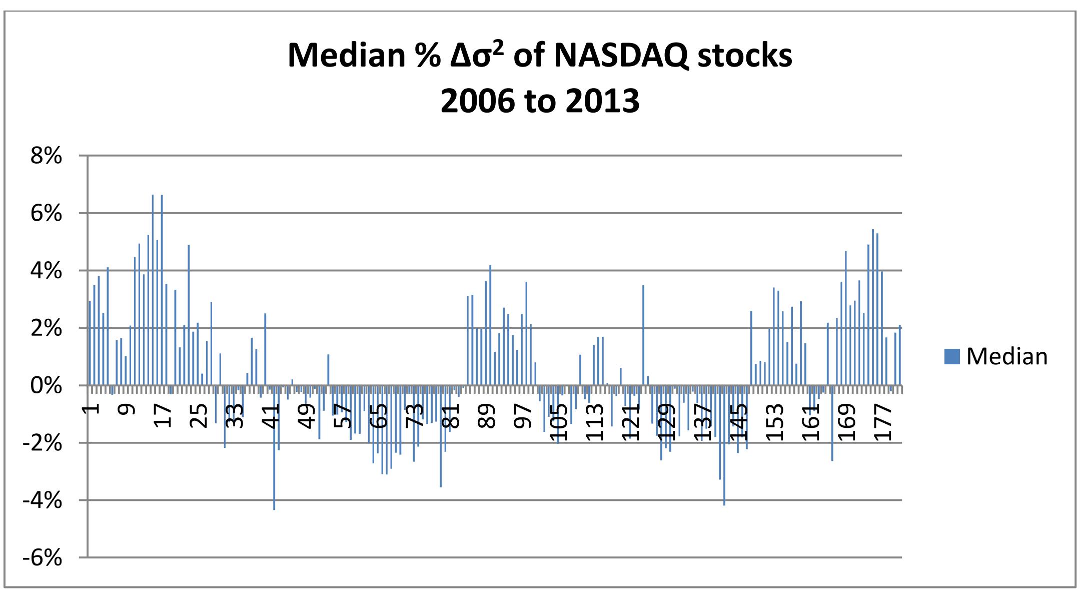
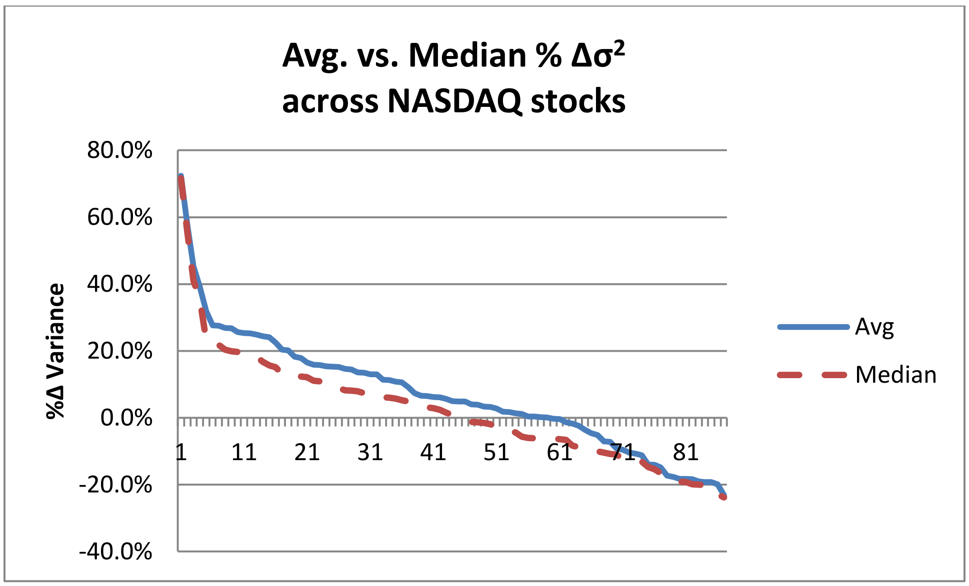


Figure 3.4 Median change in Variance



I further examine the average and median $\Delta\sigma_j^2$ for each individual stock across the entire sample period. Figure 3.5 shows the ranking order from positive value to negative value for all 87 stocks.

Figure 3.5



The number of stocks that have positive average $\Delta\sigma_j^2$ is 59 stocks (68% of the sample) and only 45 stocks (51%) have positive median value. Comparing the absolute changes in variance, the magnitude is higher for the positive $\Delta\sigma_j^2$ than the negative ones. Therefore it is not conclusive that the filtering method indeed reduces volatility risk across most stocks.

The results for the S&P sample of 80 stocks are reported in Figures 3.6, 3.7 and 3.8. The results for the S&P sample are in line with the results for the NASDAQ sample. Only around 50% of stocks show reduced variance while the other half show increased in variance. The conclusion here is that while Neutron appears to enhance the correlation effect in most cases, the corresponding effect on the variance is not conclusive. It appears that for those stocks whose variance has been reduced, the reduction magnitude can reach as high as 70%. For others, the variance increase can go as high as 23%.

Figure 3.6

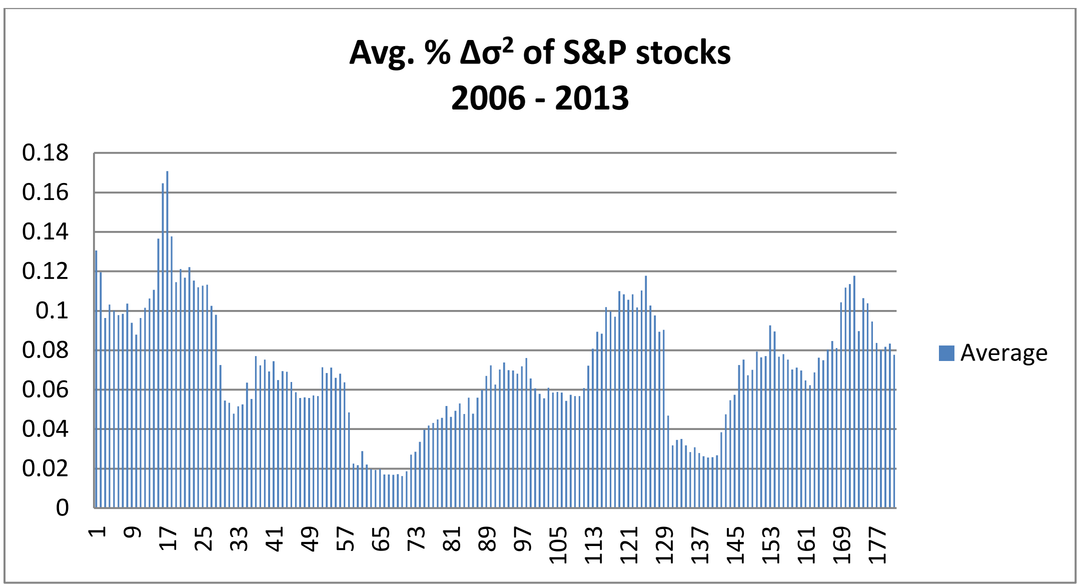


Figure 3.7

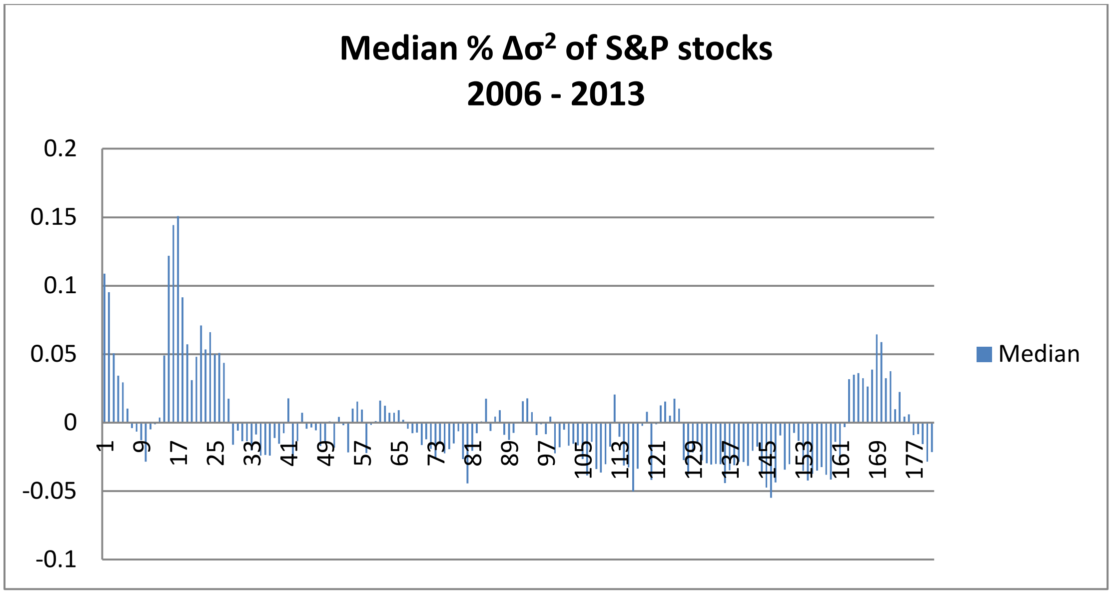
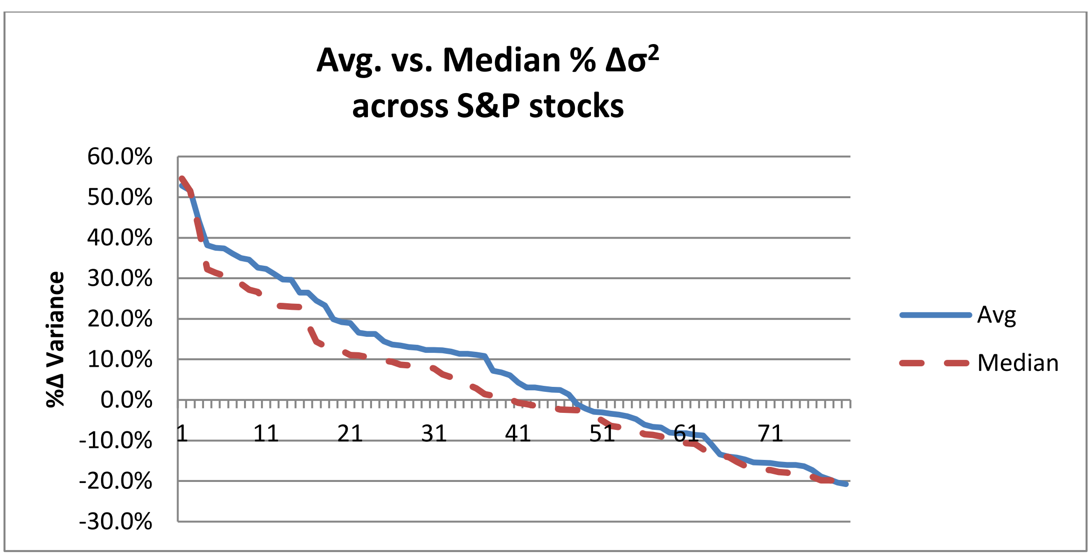


Figure 3.8



The above impact analysis on correlation and variance may be interesting, however, it is more interesting to analyze the implication of filtering on efficient portfolios. In this section, I analyze the efficient frontier by minimizing the portfolio's variance for various levels of expected return. In particular, the two sets of efficient frontiers are traced out – one based on the raw sample covariance matrix; the other based on the filtered covariance matrix.

Two cases are under examination. The first case (Case A) is imposing the "No Short Sale" constraint. The second case (Case B) is when "Short Selling Allow". Three dates from the sample period are chosen, based on the level of variance of the Minimum Variance Portfolio (MVP) under the "No Short Sale" case and using the raw covariance matrix for the NASDAQ sample. The first date is February-13-2014. The MVP shows a variance of 0.439%. The second date is March-9-2009 where the MVP's variance is 2.419%. This data set contains the returns during the height of the financial crisis. The third date is January-13-2012 where MVP's variance is 1.184%. The results for the NASDAQ sample of 87 stocks are reported here. Only the graphs for the NASDAQ sample are reported. The results for the S&P are consistent and similar in nature and magnitude to the NASDAQ sample. Figure 3.9A shows the efficient frontier for Case A on 2/13/2014. Figure 3.9B shows the Case B efficient frontier on the same date.

Figure 3.9A

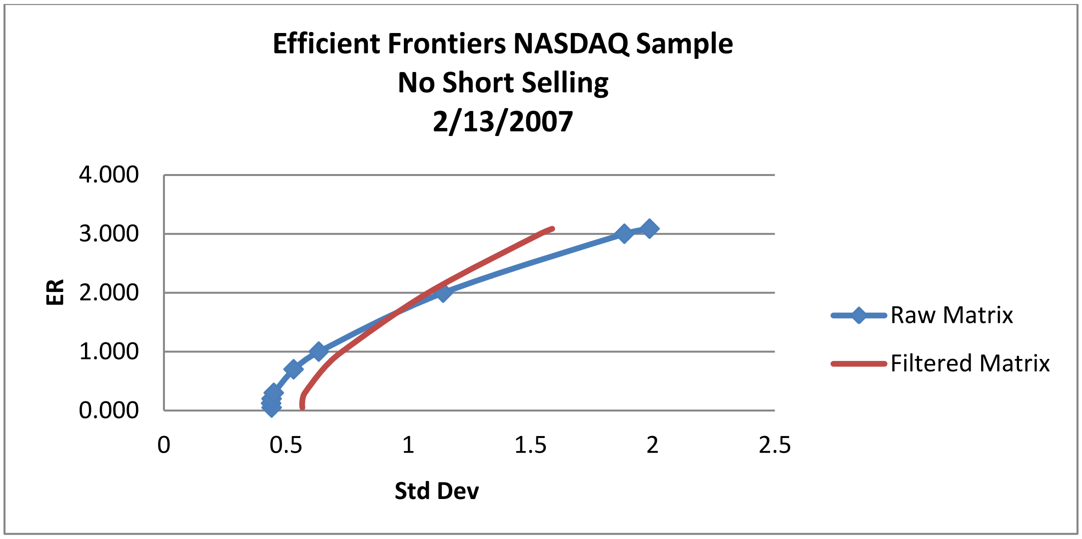


Figure 3.9B

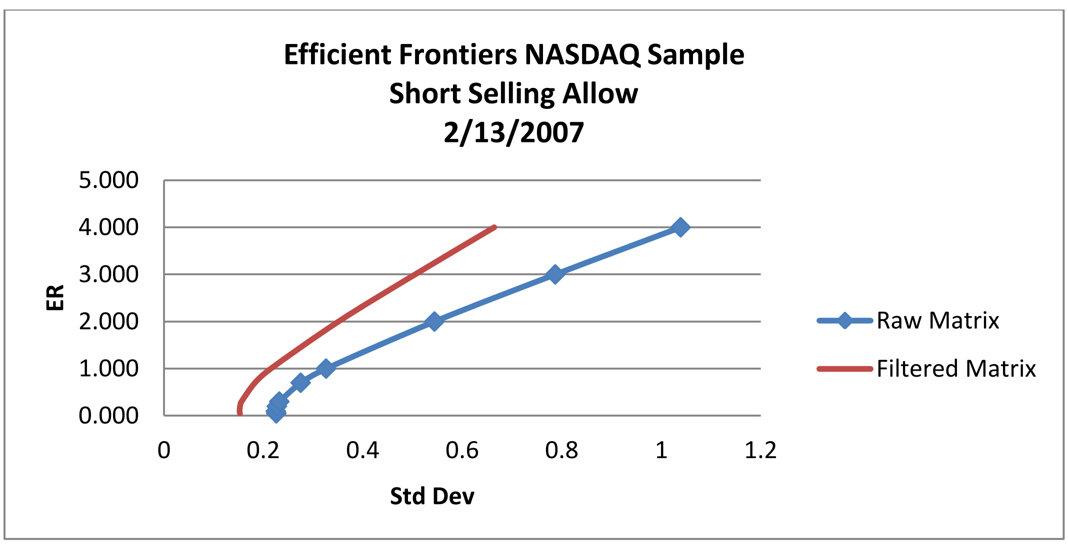


Figure 3.10A

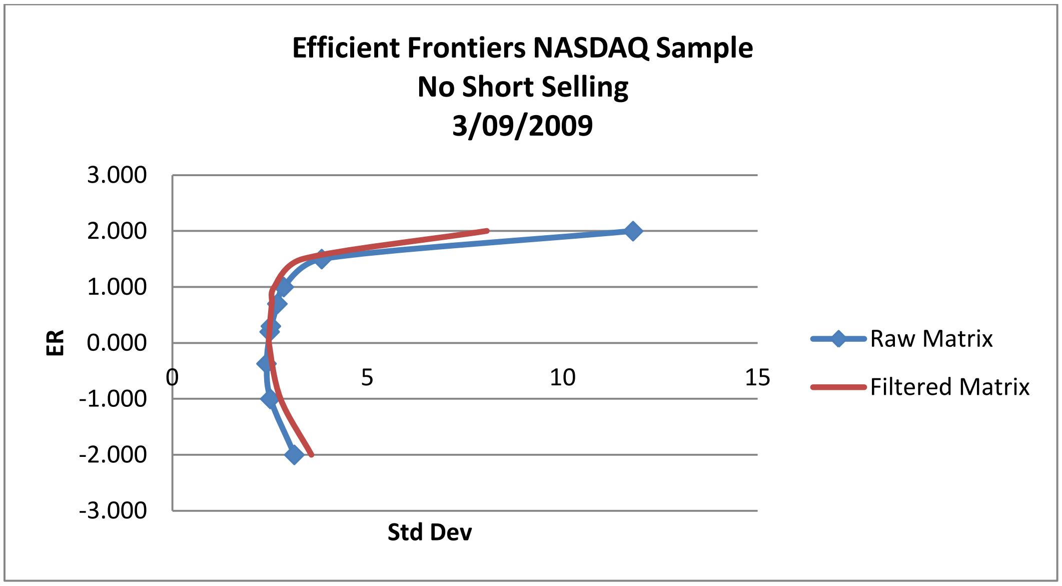


Figure 3.10B

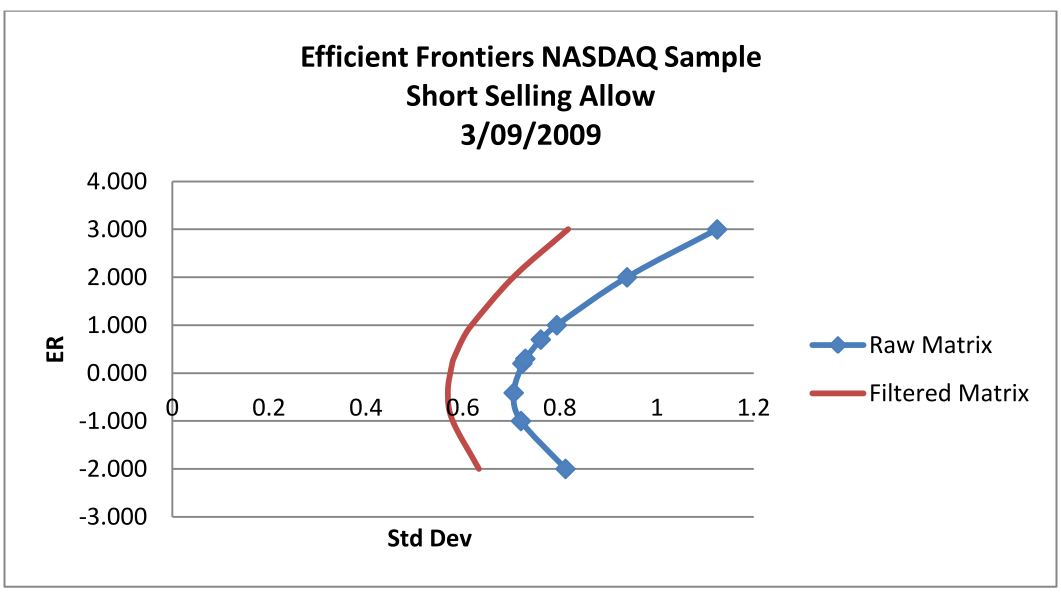


Figure 3.11A

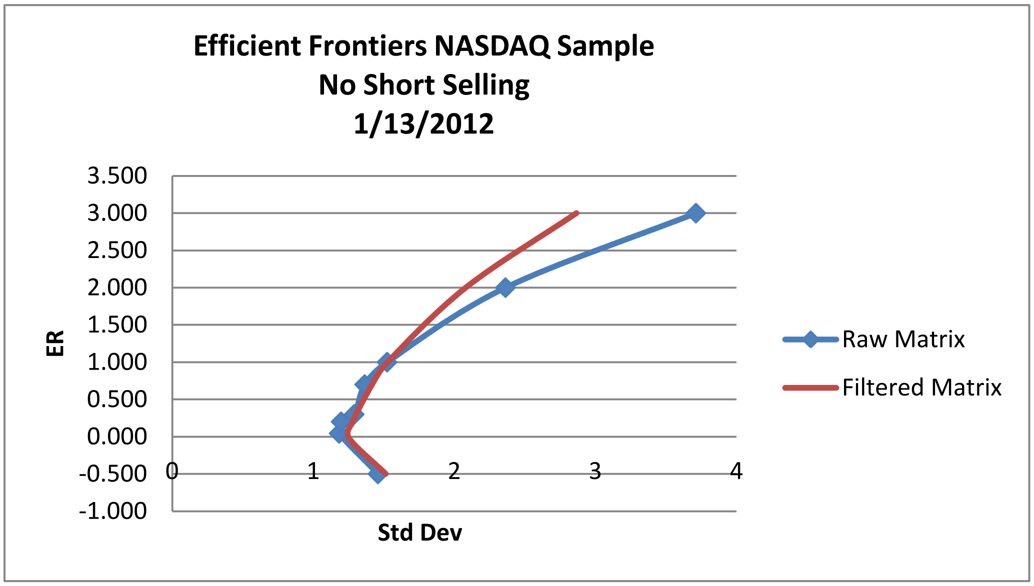
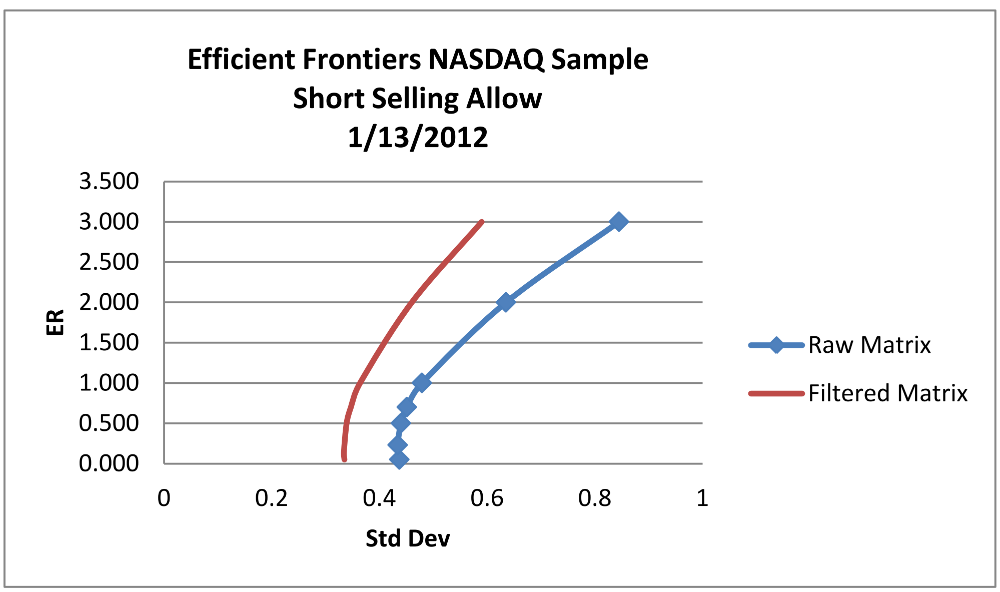


Figure 3.11B



Under the **short sale restriction**, the MVP for the raw covariance actually dominates the MVP for the filtered covariance. But as the level of expected return increases, the filtered frontier dominates the raw frontier. This can be seen clearly in Figures 3.9A, 3.10A and 3.11A. At the same time, the number of securities selected in the optimal portfolios declines as expected return increases. This reflects the fact that as the expected return and risk trade-off is further away from the MVP, the expected return input dominates the risk input. Recall that the solution for the MVP is solely dependent on the covariance risk.

When the short sale restriction is removed, the filtered efficient frontier dominates the raw efficient frontier in all cases. Figures 3.9B, 3.10B and 3.11B clearly show the result. The question is whether this ex ante dominance can translate into ex post superior performance. The answer to this important question will be tested in later chapters.

Chapter 4

Testing the noise of sample portfolios and predictive risk.

The main objectives of this chapter are twofold. First, it is important to measure and access how “noisy” the empirical covariance matrices are with respect to the “filtered” covariance matrices. Second, it is more important to compare the accuracy of “predicted risks” that are generated from the filtered matrices versus those predicted risks from the raw empirical matrices. Of particular interest is the period surrounding the financial crisis of 2008.

This chapter follows the experiment put forth by Pafka and Kondor (2002, 2003). Consistent to those explained elsewhere in prior chapters, they find that the effect of noise strongly depends on the ratio $\boldsymbol{r = N/T}$, where N is the size of the portfolio and T the length of the available time series.[1] Their simulation results show that for larger $\boldsymbol{r}$ (e.g. 0.6) noise does have the pronounced effect as suggested by Galluccio et al. (1998) and Laloux et al. (2000)

Pafka and Kondor argue that in addition to noise coming from finite length of time series, real data always contain additional sources of error. Examples of such error

[1] In Chapter 1, the ratio Q=T/N is the inverse of ratio r

can arise from non-stationarity in risks and returns, changes in the composition of the portfolio, changes in regulation and changes in fundamental market conditions. In order to isolate their experiments from such additional errors, they based their analyses on artificially generated data from some “toy” models. The advantage is that the “true” parameters of the underlying stochastic process and the statistics of the covariance matrix are exactly known.

In both papers, Pakfa and Kondor (2002, 2003) investigated the impact of noisy covariance matrices on the portfolio optimization problem. Specifically, the objective function is to

Minimize $\sum_{i,j=1}^{n} w_i \sigma_{ij} w_j$

subject to a linear budget constraint:

$$\sum_{i=1}^{n} w_i = 1$$

w_i denotes the weight of asset i in the portfolio while σ_{ij} represents the covariance matrix of returns. Short sale is allowed in the optimization. Only the minimal risk portfolio (MVP) is under investigation. Using the method of Lagrange multipliers, the solution to the optimization problem is simply

$$w_i^* = \frac{\sum_{j=1}^{n} \sigma_{ij}^{-1}}{\sum_{j,k=1}^{n} \sigma_{jk}^{-1}} \tag{34}$$

Let the "noiseless" covariance matrix be denoted by $\sigma_{ij}^{(0)}$ and the "noisy" covariance matrices by $\sigma_{ij}^{(1)}$ such that

$$\sigma_{ij}^{(1)} = \frac{1}{T}\sum_{t=1}^{T} y_{it}y_{jt} \tag{35}$$

where $y_{it} = \sum_{j=1}^{n} L_{ij}x_{jt}$ with x_{jt} ~ i.i.d. N(0,1) and L_{ij} is the Cholesky decomposition of the matrix $\sigma_{ij}^{(0)}$ such that $\sum_{k=1}^{n} L_{ik}L_{jk} = \sigma_{ij}^{(0)}$. In other words, the "noisy" matrix is randomly generated off from the "noiseless" matrix by a standardized Gaussian distribution. Thus, $\sigma_{ij}^{(1)}$ can be seen as representing the empirical noisy covariance matrix while $\sigma_{ij}^{(0)}$ represents the "true" covariance matrix. As $T \to \infty$ the noise disappears and $\sigma_{ij}^{(1)} \to \sigma_{ij}^{(0)}$.

Their experiments employ two toy models. The first toy model (Model I) is to simply use the identity matrix for $\sigma_{ij}^{(0)}$ (unit variance for all assets and zero correlations elsewhere). The second toy model (Model II) has one distinct eigenvalue set to be 25 times larger than the rest of the eigenvalues and with corresponding eigenvector (representing the "whole market").[2]

---

[2] In Appendix 2, I find the largest eigenvalue to be around 16 times larger than the maximum predicted eigenvalue.

Define the set of risk for the MVP as

1) $V_0 = \sum_{i,j=1}^{n} w_i^{(0)*}\sigma_{ij}^{(0)}w_j^{(0)*}$ as the ***"true" risk of MVP without noise***, where $w_i^{(0)*}$ is the optimal weight without noise.
2) $S_0 = \sum_{i,j=1}^{n} w_i^{(1)*}\sigma_{ij}^{(0)}w_j^{(1)*}$ as the ***"true" risk of MVP with noise***, where $w_i^{(1)*}$ is the optimal weight in the presence of noise.
3) $S_1 = \sum_{i,j=1}^{n} w_i^{(1)*}\sigma_{ij}^{(1)}w_j^{(1)*}$ as the ***"predicted" risk of MVP***, that is the risk that can be observed if the optimization is based on a return series of length T.
4) $S_2 = \sum_{i,j=1}^{n} w_i^{(1)*}\sigma_{ij}^{(2)}w_j^{(1)*}$ as the "realized" risk of MVP, that is the risk that would be observed if the portfolio were held one more period of length T, where $\sigma_{ij}^{(2)}$ is the covariance matrix calculated from the returns in the second period.

In addition, denote the following three ratios as:

$q_0 = \frac{\sqrt{S_0}}{\sqrt{V_0}}$ = "true" risk with noise per unit of the "true" risk in the absence of noise.

$q_1 = \frac{\sqrt{S_1}}{\sqrt{V_0}}$ ="predicted" risk with noise per unit of the "true" risk in the absence of noise.

$q_2 = \frac{\sqrt{S_2}}{\sqrt{V_0}}$ ="realized" risk with noise per unit of the "true" risk in the absence of noise.

One would expect that $q_0 > 1$ for all values of N and T since the "optimal" portfolio obtained from the 'noisy' covariance matrix must be less efficient than the one obtained from the "true" covariance matrix.

Pakfa and Kondor (2003) find that q_0 is indeed greater than 1 for all simulated cases and that q_2 is also very close to q_0. This suggests that "realized" risk can be a good proxy for the "true" risk when the "true" covariance matrix is not known. Unfortunately, this is not true for the "predicted" risk. They find that q_1 is always smaller than q_0 and q_2. This implies that **optimization in the presence of noise will bias risk measurement and lead to the underestimation of the risk** of the optimal portfolio. These results are in perfect qualitative agreement with those in Laloux et al. (2000) and Plerou et al. (1999).

In fact both ratios q_0 and q_1 can be calculated analytically. For $T, N \to \infty$ and for a fixed $= \frac{N}{T}$, the eigenvalue density of the covariance matrix is given as:[3]

$$\rho(\lambda) = \frac{1}{2\pi \mathrm{r}} \frac{\sqrt{(\lambda_+ - \lambda)(\lambda - \lambda_-)}}{\lambda} \tag{36}$$

where $\lambda\pm = (1 \pm \sqrt{\mathrm{r}}\,)^2$ are the upper(+) and lower (-) bounds of the eigenvalues.

Accordingly, q_0 can be written as

$$q_0 = \frac{\sqrt{\int \rho(\lambda)/\lambda^2 d\lambda}}{\int \rho(\lambda)/\lambda\, d\lambda} \tag{37}$$

thus yielding $q_0 = 1/\sqrt{1-r}$ by simple integration. By the same token, $q_1 = \sqrt{1-r}$. These asymptotic formulae are useful in checking the real empirical covariance matrices below.

[3] See Crisanti and Sompolinsky (1987)

Finally, as quoted by the Pakfa and Kondor (2003),

> "Our main finding was that for parameter values typically encountered in practice the 'true' risk of the ***minimum-risk portfolio*** determined in the presence of noise (i.e., based on the covariance matrix deduced from finite time series) is usually no more than 10-15% higher than that of the portfolio determined from the 'true' covariance matrix".

It appears that noise has relatively small effect on the portfolio variance under linear constraints. This seems to be in contradiction with the reported conclusion by Laloux et al. (1999) that about 94% of the spectrum of these matrices can be fitted by that of a completely random matrix. The large discrepancy between "predicted" and "realized" risk in previous studies can be explained by the relatively high values in N/T used in these studies.[4] However, the authors caution that in the presence of non-linear constraints of the type: $\sum_{i=1}^{n} \gamma_i |w_i| = 1$, the presence of noise may create high degree of instability and degenerate solutions.

At this juncture, one can easily apply the formulae above for both q_0 and q_1 to my current studies. For the NASDAQ sample N = 87 and T = 119 which means r = 0.7311 which is relatively high. Most of the random noise is expected to be in the range of the eigenvalue between λ- = 0.021 and λ+ = 3.441. The calculated values for q_0 and q_1 are 1.928 and 0.518, respectively. As for the S&P sample, N =80 and T=119 with r = 0.6722 is slightly lower than the NASDAQ sample. The eigenvalue with the most random noise is between λ- = 0.0324 and λ+ = 3.312. The calculated values for q_0 and q_1 are 1.747 and 0.572, respectively.

[4] See Laloux et al. (2000) and Plerou et al. (1999)

These values imply that one would expect a lot of differences between predicted risk and realized risk on a MVP. Clearly, in my studies, the "true" covariance matrix is not known. Neither can I isolate other sources of data error. However, one can judge how well the Neutron filtering process is by analyzing the $\sigma_{MVP,Raw}$ to $\sigma_{MVP,Filter}$. Define the following:

1) $S_{raw} = \sum_{i,j=1}^{n} w_i^{(raw)*} \sigma_{ij}^{(raw)} w_j^{(raw)*}$ as the "predicted" risk of MVP with noise, where $w_i^{(raw)*}$ is the optimal weight in the presence of noise in the empirical covariance matrix.

2) $S_{filter} = \sum_{i,j=1}^{n} w_i^{(filter)*} \sigma_{ij}^{(filter)} w_j^{(filter)*}$ as the proxy of "true" risk for MVP using the filtered covariance matrix.

3) $S_{filter}^{*} = \sum_{i,j=1}^{n} w_i^{(raw)*} \sigma_{ij}^{(filter)} w_j^{(raw)*}$ as the "predicted" true risk of MVP with noise, where $w_i^{(raw)*}$ is the optimal weight in the presence of noise.

4) Define $Q_1 = \frac{\sqrt{S_{raw}}}{\sqrt{S_{filter}}}$ to be the "predicted" risk with noise per unit of the "predicted" risk in the filtered-out noise.

5) Define $Q_3 = \frac{\sqrt{S_{raw}}}{\sqrt{S_{filter}^{*}}}$ to be the "predicted" risk with noise per unit of the "predicted" risk in the filtered-out noise, but using the optimal weights generated from the raw matrices.

This section presents the results for the Q ratios. In addition to reporting the results for the full period (2006 to 2013), I divide the period into three sub-samples. The first sub-period is from 6/23/2006 to 10/27/2008. The second sub-period is from 11/10/2008 to 3/16/2011. The last sub-period is 3/30/2011 to 8/19/2013. The results of the $\boldsymbol{Q_1}$ ratios using the MVP from both raw and filtered matrices

Table 4.1 Q1 ratios of MVPs

			--- No Short Sale --		*----- Short Sale -----*	
		N	**Avg Q(1)**	**Stdev**	**Avg Q(1)**	**Stdev**
NASDAQ	Full Period	180	0.926	0.078	1.381	0.097
	Sub-period 1	60	0.882	0.073	1.416	0.113
	Sub-period 2	60	0.945	0.046	1.355	0.083
	Sub-period 3	60	0.950	0.090	1.372	0.084
S&P	Full Period	180	0.933	0.075	1.338	0.090
	Sub-period 1	60	0.892	0.068	1.327	0.087
	Sub-period 2	60	0.961	0.050	1.337	0.092
	Sub-period 3	60	0.945	0.086	1.351	0.091

In the case where there is short sale constraint, the average Q_1 for both NASDAQ and S&P samples is consistently less than 1.0 but much higher than the expected values of 0.5 to 0.6. The average value of Q_1 falls between 0.88 and 0.96. On the other hand, when the short sales constraint is removed, the average value of $Q_1 > 1$. Under this scenario, it appears that the risk of the MVP, generated using the raw correlation matrices, is significantly higher than the risk of the MVP using the filtered matrices.

Table 4.2 compares the average standard deviation of both MVP $_{Raw}$ and MVP $_{Filter}$ under short sale constraint. In every period, the risk of the MVP$_{Filter}$ exceeds the risk of MVP $_{Raw}$ (4% to 11% higher). The coefficient of variation (Average/Std deviation) is also higher for the filtered case. But there are less stocks being selected in the MVP $_{Filter}$ than those contained in the MVP $_{Raw}$ -- at least two to three times more stocks in MVP $_{Raw}$. This can be explained by the ***efficiency*** in the filtered covariance case whereby fewer stocks are required to generate the minimum variance portfolio. In other words, since there is less noise in the filtered correlation matrix, less number of securities are required to "efficiently" diversify the risk to a minimum level.

Table 4.2 Average MVP Risk under No Short Sale

				No Short Sale					
		Raw Covariance				***Filtered Covariance***			
NASDAQ	**N**	**Avg**	**Stdev**	**Coef Var**	**W > 0**	**Avg**	**Stdev**	**Coef Var**	**W > 0**
Full Period	180	0.847%	0.44%	1.91	16.3	0.914 %	0.46 %	1.98	6.4
Sub-period 1	60	0.745%	0.25%	3.03	21.9	0.842 %	0.26 %	3.28	8.7
Sub-period 2	60	1.115%	0.62%	1.81	13.0	1.179 %	0.65 %	1.82	5.7
Sub-period 3	60	0.683%	0.22%	3.13	14.1	0.725 %	0.23 %	3.09	4.8
S&P									
Full Period	180	0.716%	0.41%	1.75	15.0	0.761 %	0.42 %	1.83	6.7
Sub-period 1	60	0.605%	0.22%	2.72	17.2	0.671 %	0.22 %	3.03	7.7
Sub-period 2	60	0.957%	0.58%	1.65	12.2	0.995 %	0.61 %	1.64	5.3
Sub-period 3	60	0.589%	0.18%	3.23	15.7	0.620 %	0.17 %	3.75	7.3

In Table 4.3, when short sale restriction is lifted, the MVP $_{Filter}$ is showing less risk than its MVP $_{Raw}$ counterpart. The latter's average risk is at least 31% higher than the average risk for MVP $_{Filter}$. The coefficient of variation is also higher for the MVP $_{Raw}$. The number of long securities and the number of short securities are very close for both cases. Therefore, one can safely conclude that the portfolio risk in the filtered case, as measured by the MVP, is significantly lower than the corresponding portfolio risk in the unfiltered case. The optimization process reduces risk more efficiently when noise is reduced.

Table 4.3 Average MVP Risk under Short Sale Allowed

		Short Sale Allowed							
		Raw Covariance				***Filtered Covariance***			
NASDAQ	**N**	**Avg**	**Stdev**	**Coef Var**	**W > 0**	**Avg**	**Stdev**	**Coef Var**	**W > 0**
Full Period	180	0.340%	0.13%	2.64	45.6	0.249%	0.10%	2.44	45.7
Sub-period 1	60	0.325%	0.09%	3.60	45.6	0.233%	0.08%	2.92	46.0
Sub-period 2	60	0.407%	0.18%	2.26	44.9	0.302%	0.14%	2.19	45.5
Sub-period 3	60	0.291%	0.05%	5.40	46.3	0.213%	0.05%	4.67	45.7
S&P									
Full Period	180	0.322%	0.13%	2.49	43.7	0.243%	0.11%	2.30	43.0
Sub-period 1	60	0.285%	0.07%	3.89	44.2	0.215%	0.06%	3.75	43.0
Sub-period 2	60	0.415%	0.17%	2.42	42.9	0.317%	0.14%	2.19	42.9
Sub-period 3	60	0.267%	0.05%	5.13	44.0	0.197%	0.04%	5.48	43.2

The next step is to compare the risk for any **constant weighted** portfolio. The natural candidate for this study is an equally weighted portfolio. This is an interesting comparison because it shows the portfolio risk due to naïve diversification strategy. In Chapter 3, I examined the impact on the correlation structure after the filtration process.

The immediate effect is higher correlated values in general for the filtered matrix. If that is the case, one expects the $\sigma_{Filter}^{EW} > \sigma_{Raw}^{EW}$. The result is borne out in Table 4.4 and the risks are computed for 15 non-overlapping periods. The implication of this is that given any existing portfolio, the risk measure (as defined by the standard deviation of the portfolio) is always higher using the filtered covariance matrix than the risk measure derived from the raw covariance matrix. In terms of risk management and VaR analysis, this is critical because the raw covariance leads to **underestimation** of the actual portfolio risk. The filtered covariance gives a better (higher) risk measure for the portfolio.

Table 4.4 Risk Comparison for Equally Weighted Portfolio

	Equally Weighted Portfolio			
	NASDAQ (87 stocks)		S&P (80 stocks)	
Date	σ_{Raw}^{EW}	σ_{Filter}^{EW}	σ_{Raw}^{EW}	σ_{Filter}^{EW}
6/23/2006	0.950	1.557	0.842	1.245
12/13/2006	1.028	1.527	0.747	1.049
6/8/2007	0.864	1.262	0.730	1.026
11/28/2007	1.262	1.747	1.243	1.613
5/21/2008	1.514	1.983	1.374	1.709
11/10/2008	2.765	3.376	3.048	3.552
5/5/2009	3.006	3.616	3.543	4.163
10/23/2009	1.350	1.780	1.541	1.989
4/19/2010	1.014	1.365	1.082	1.449
10/7/2010	1.608	1.946	1.560	1.883
3/30/2011	0.894	1.254	0.801	1.144
9/20/2011	1.692	2.000	1.693	1.975
3/13/2012	1.487	1.867	1.574	1.913
8/31/2012	1.123	1.470	1.027	1.338
2/27/2013	0.839	1.205	0.831	1.210

At this juncture, I also compare the predictive risk of the MVP to the realized or actual risk. The sample period is divided into 15 predictive non-overlapping dates. The dates are shown below:

Predictive date	Realized date
6/23/2006	12/13/2006
12/13/2006	6/8/2007
6/8/2007	11/28/2007
11/28/2007	5/21/2008
5/21/2008	11/10/2008
11/10/2008	5/5/2009
5/5/2009	10/23/2009
10/23/2009	4/19/2010
4/19/2010	12/3/2010
12/3/2010	7/25/2011
7/25/2011	3/13/2012
3/13/2012	10/31/2012
10/31/2012	6/21/2013

The root mean square error (RMSE) is chosen as the criterion to measure the effectiveness of risk prediction. The mean square error (MSE) is defined as

$$MSE = \sum_{t=1}^{T}(\sigma_{Realize}^{MVP} - \sigma_{Raw}^{MVP})^{2}/T \quad \text{and}$$

$$RMSE = \sqrt{MSE}$$

The detail results are contained in various tables (Tables B-1 to B-14) that are delegated to Appendix B. Instead the summary of the RMSE results are shown in Table 4.5

Table 4.5

Summary of RMSE for Predictive Risk

	No Short Sale		**Short Sale Allow**		**Equally Weighted**	
	MVP(Raw)	MVP(Filter)	MVP(Raw)	MVP(Filter)	Raw	Filter
NSADAQ	**0.571**	0.674	1.221	**0.893**	**0.658**	0.708
S&P	**0.477**	0.508	1.103	**0.716**	**0.796**	0.855
	Use MVP w*(raw)		**Use MVP w*(raw)**			
NSADAQ	**0.571**	0.650	1.221	**0.957**		
S&P	**0.477**	0.504	1.103	**0.790**		
	Use MVP w*(filter)		**Use MVP w*(filter)**			
NSADAQ	**0.559**	0.674	0.995	**0.893**		
S&P	**0.465**	0.508	0.844	**0.716**		

Both “No Short Sale” and “Short Sale Allow” cases are examined. The first test is to examine how the predicted risk of each MVP is measured against realized risk for non-overlapping periods. Using the NASDAQ results as the example, under short sale restriction, MVP_{Raw} has a smaller predictive error (0.571) than MVP_{Filter} (0.650). But the reverse holds when the short sale constraint is removed (1.221 for MVP_{Raw} vs. 0.893 for MVP_{Filter}). When I replaced the weights, w*, from MVP_{Raw}, the results are still consistent (0.571 for MVP_{Raw} vs. 0.650 for MVP_{Filter}). Likewise, when the weights are replaced by MVP_{Filter}, the same results hold. It becomes clear that when short sale constraint is imposed, the filtered risk is less effective in its predictive power than the raw sample risk. But for unconstrained optimization, there is a big advantage using the filtered data. For the equally weighted portfolio, the predictive risk for the raw sample data is better than the filtered data (0.658 for MVP $_{Raw}$ vs. 0.708 for MVP $_{Filter}$).

Finally, the question is whether there is any improvement in the predictive risk for the raw sample if one were to use the "filtered" weights of MVP $_{Filter.}$ Indeed, there is improvement using the filtered weights. Under the no short sale case, the RMSE for the raw sample is reduced from 0.571 to 0.559 for the NASDAQ stocks and from 0.477 to 0.465 for the S&P stocks. On average, the improvement in RMSE is around 2.3%. The improvement is more pronounced when the short sale restriction is removed. The RMSE is reduced from 1.221 to 0.995 for NASDAQ and from 1.103 to 8.44 for the S&P. The average improvement in RMSE is around 19%. The implication is that optimal weights that are derived from filtered covariance have better risk prediction. Furthermore, under the short sale restriction, the MVP $_{Filter}$ contains an average of 6 stocks versus an average of 15 stocks for MVP $_{Raw.}$ This implies that optimization using the filtered data requires less number of stocks in reducing risk to a minimal.

Chapter 5

Index Tracking

Many investors manage portfolios (or parts of portfolios) to match index returns. Active managers may fall back to passive index tracking in times when they have no definite views. Many of the Electronic Traded Funds (ETFs) are simply index tracking funds. The simplest form of tracking the performance of an index, for example the S&P 500 Index, is the straightforward replication technique. This involves duplicating the target index precisely, holding all its securities in their exact proportions. Once replication is achieved, trading in the indexed portfolio becomes necessary only when the composition of the portfolio changes or as a way of reinvesting cash flows.

While the straightforward index replication method may be applicable to the largest of funds, it becomes overwhelming costly for smaller funds. These funds require a smaller set of securities in order to duplicate the index. This is especially true in the bond fund world, where many of the bonds in the Bond Index are thinly traded and bonds often mature, thereby changing the composition of the Index.

Alternative methods such as stratified sampling, tracking error minimization and factor-based replication were invented in order to reproduce the overall attributes of the Index with a limited number of securities. As Martellini et al. (2003) quote,

> "While this may sound simple in theory, it is difficult to achieve in practice. Passive does not mean inactive. In fact, it takes a very active portfolio management process to deliver reliable index performance with low tracking error. It requires extensive portfolio modeling and monitoring, together with very disciplined and cost-conscious trading capabilities."[1]

The stratified sampling method is cell-matching the Index's attributes (example Sector matching) while the factor-based replication attempts to match the exposure of the replicating portfolio with respect to a set of common factors with that of the index (example BARRA Equity model[2]). The tracking error optimization method tries to replicate the index return directly. It is the latter method that is of interest in this chapter.

Some active equity funds may choose to trade only a smaller number of stocks (10 to 20) in the hope of beating the index (or benchmark). A neutral view for these funds is to replicate the benchmark's return. An active view for the fund may consist of over-allocation of some stocks while under-allocation of other stocks within the same portfolio. In any case, tracking error minimization approach may be their starting point of active portfolio allocation.

[1] Page 214.
[2] www.msci.com/products/portfolio_management_analytics/equity_models/

The aim of this chapter is to compare the effectiveness of using Neutron's filtered covariance matrices vis-à-vis the raw covariance matrices in tracking the stock benchmarks. The NSADAQ 100 and the S&P 500 Index are the two benchmarks chosen for the experiment. As mentioned before, there are 5 randomly chosen portfolios for each benchmark index. In addition, the larger portfolios, consisting of all the securities in the 5 random subsets, are also included in the test. The null hypothesis is that filtered covariance matrix produces lower actual tracking error than the tracking error using the corresponding unfiltered matrix.

According to Roll (1992), the mean-variance optimization method can be applied for tracking error optimization. Stocks with higher correlation with the benchmark are more effective than stocks with lower correlation. The problem is to replicate as closely as possible the stock index return with a portfolio invested in M individual stocks. Denote R_B as the return on the benchmark (i.e., the index), $\{w_1, \ldots\ldots, w_M\}$ the weights of M stocks in the replicating portfolio, σ_{ij} $_{i,j=1,\ldots,M}$ the variance-covariance matrix of these stocks, and R_p the return on the replicating portfolio. Then

$$R_p = \sum_{i=1}^{M} w_i R_i$$

The objective function for the tracking error optimization is given by:

$$\text{Min}_{w_1..w_M} Var(R_p - R_B) = \sum_{i=1}^{M} \sum_{j=i}^{M} w_i w_j \sigma_{ij} - 2 \sum_{i=1}^{M} w_i \sigma_{iB} + \sigma_B^2 \qquad (38)$$

subject to the constraint

$$\sum_{i=1}^{M} w_i = 1$$

and with or without the short sale constraint

$$w_i \geq 0 \; for \; all \; i = 1, \dots, M$$

Note that σ_{iB} is the covariance between the return on the *i*th stock in the replicating portfolio and the benchmark return, and σ_B^2 the variance of the benchmark.

The quality of replication is measured by the tracking error (denoted by TE) which is the standard deviation of the difference between the actual return on the replicating portfolio and that of the benchmark:

$$TE = \sqrt{Var(R_p - R_B)} \qquad (39)$$

Note that the tracking error optimization does not require the expected return of stock as an input. It behaves like a minimum variance portfolio (MVP) except that the variance of the portfolio is minimized with respect to the benchmark.

The efficient frontiers, under both constrained and unconstrained short-sale, are first evaluated. The results for the NASDAQ sample, at three separate dates: 2/13/2007, 3/9/2009 and 1/13/2012, are presented in Figures 5.1, 5.2 and 5.3. The "A" graphs are associated with the constrained optimization, whereas the "B" graphs are unconstrained frontiers.

Figure 5.1A

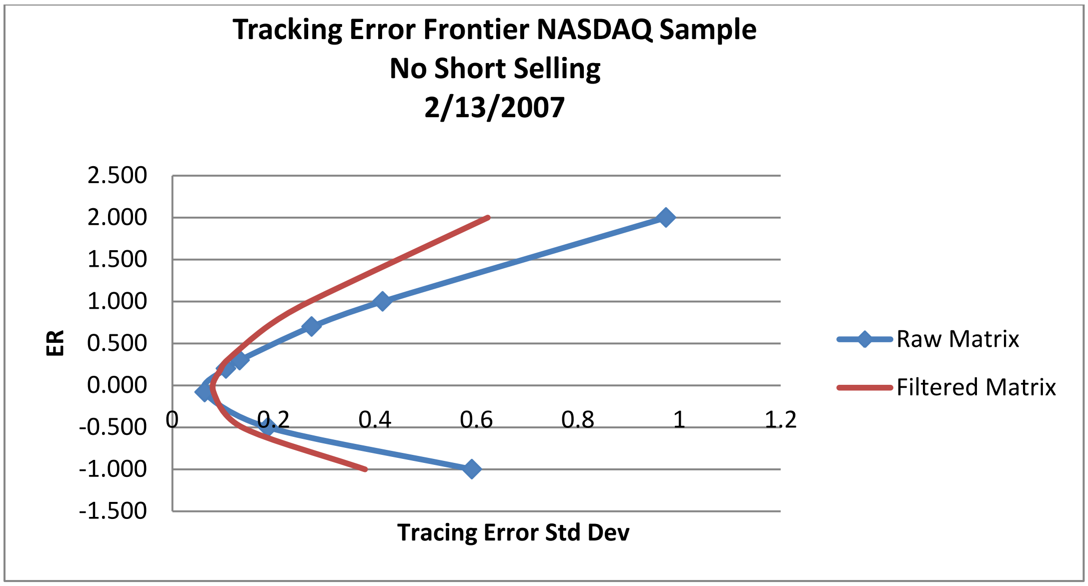


Figure 5.1B

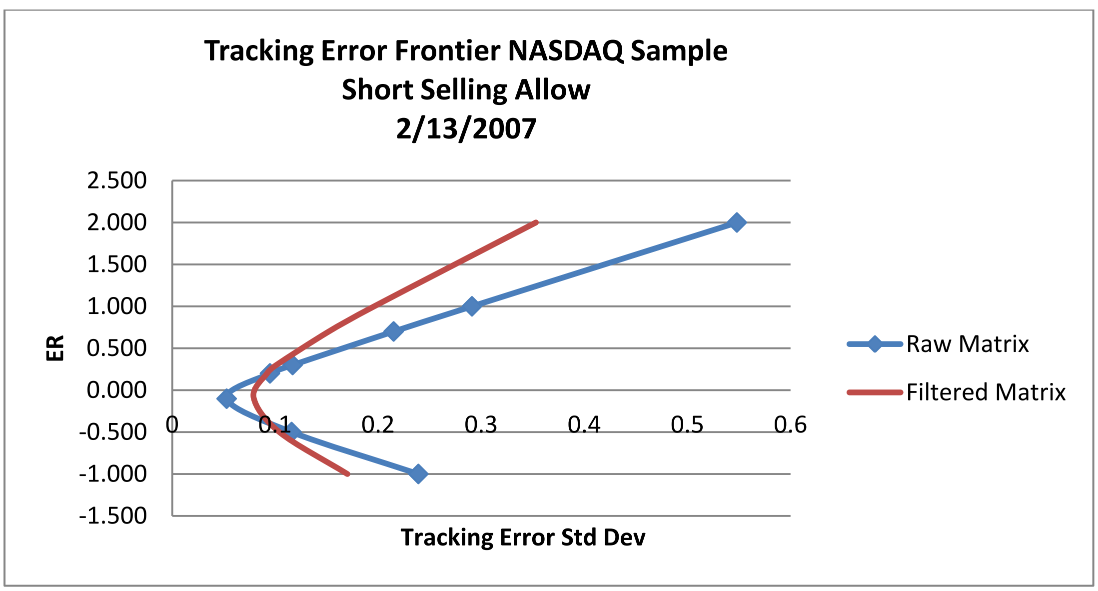


Figure 5.2A

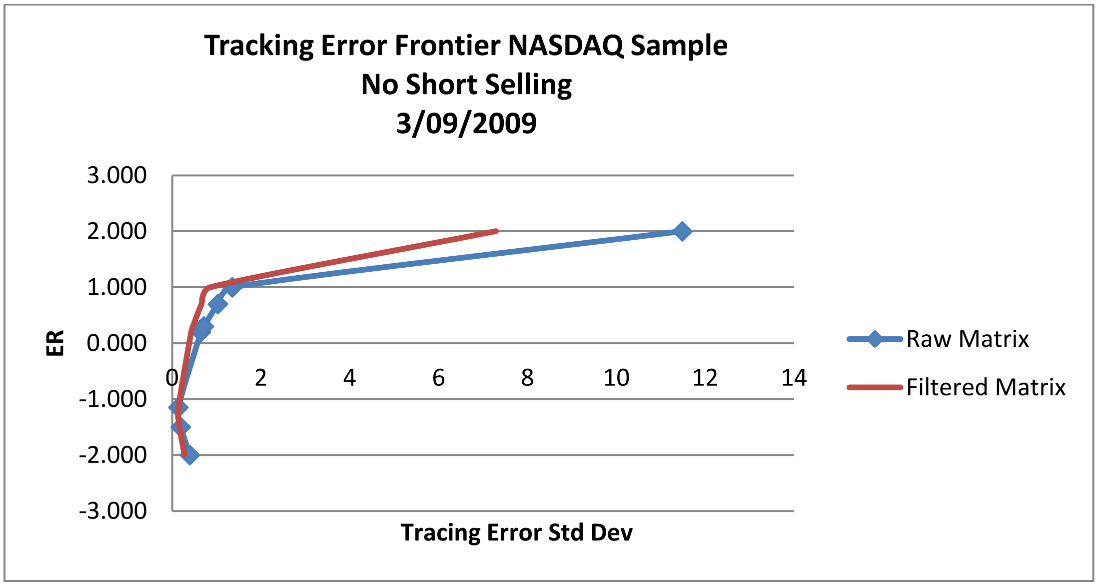


Figure 5.2B

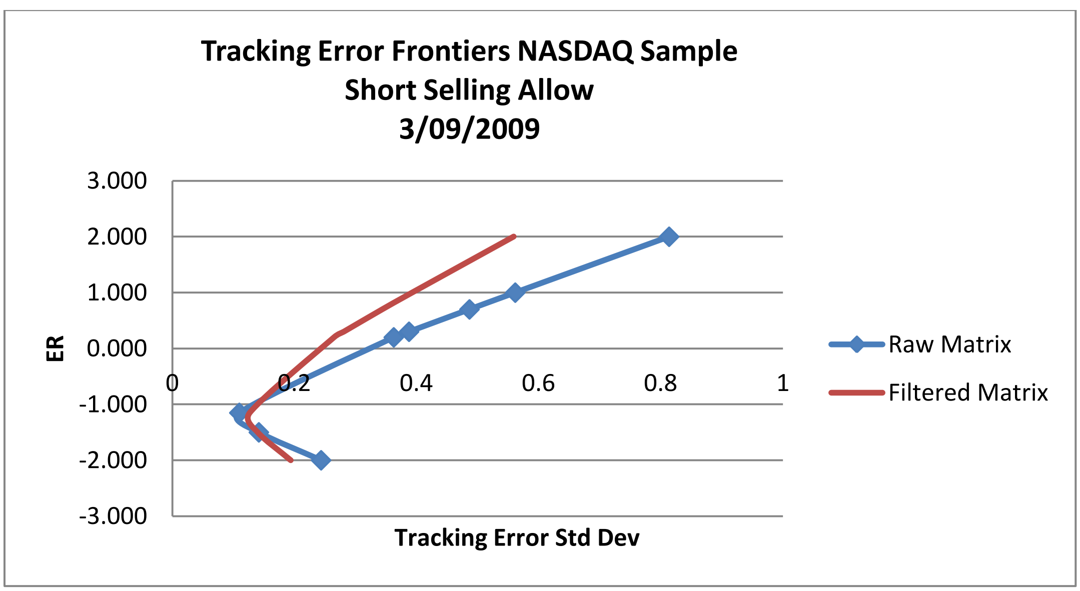


Figure 5.3A

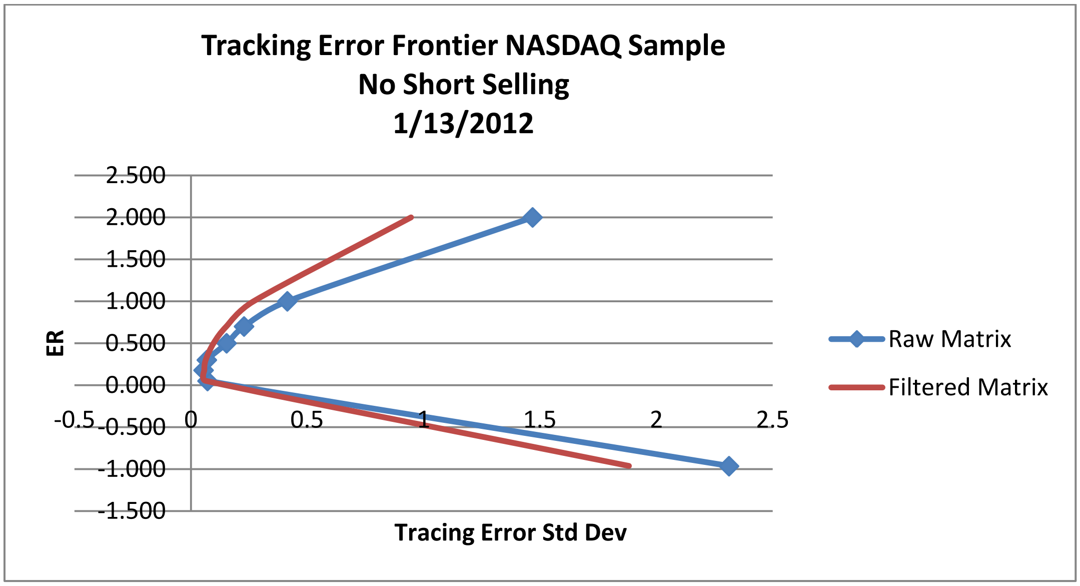
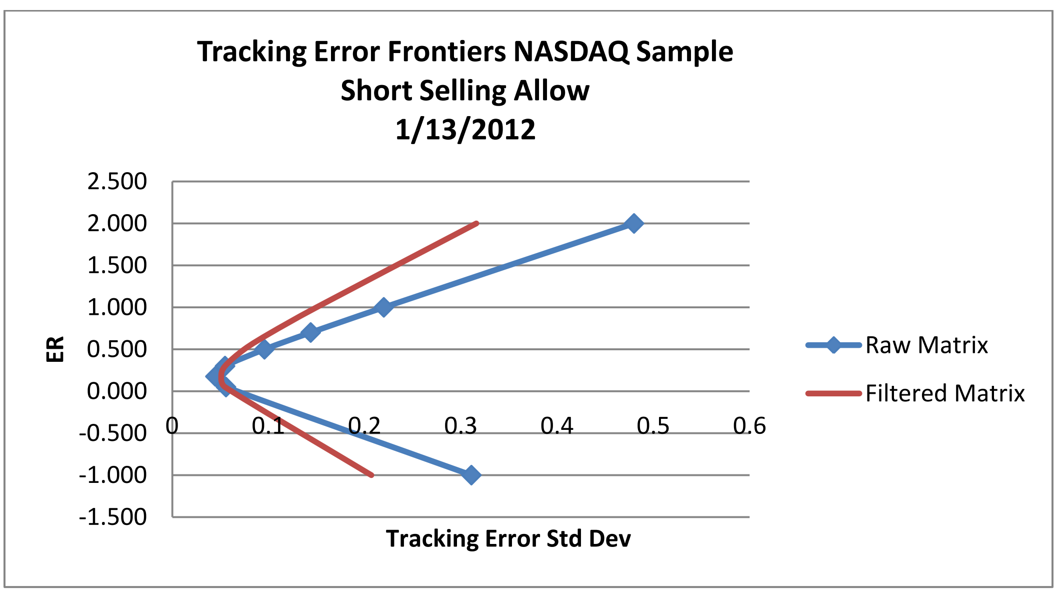


Figure 5.3B



In every case, the minimal tracking risk portfolio (MTP) is slightly more efficient for the raw covariance than the filtered covariance. However, as the tracking error standard deviation increases, the filtered efficient frontier dominates the raw efficient frontier. This is true for both constrained and unconstrained optimization. The number of securities with positive weights is very similar for both constrained and unconstrained results. The implication is that while the raw covariance produces slightly better results in the minimum tracking error than the filtered covariance, the latter produces more dominant efficient frontiers than the former.

The optimal portfolio, as according to equation (38), is first selected based on the raw covariance matrix. Then the actual returns of the benchmark index and the replicating portfolio are measured during the following 10 business days. The portfolio is then rebalanced with the updated raw covariance matrix. The process repeats itself until the entire sample period ends. The tracking error as given in (39) is then computed over the entire sample period and for the three sub-periods. The experiment is repeated for the 'filtered" covariance matrices. The results are shown in Table 5.1 and Table 5.2 below.

Table 5.1 Average MTP Risk under No Short Sale

	No Short Sale								
		Raw Covariance				*Filtered Covariance*			
NASDAQ	**N**	**Avg.**	**Stdev**	**Coef Var**	**W > 0**	**Avg**	**Stdev**	**Coef Var**	**W > 0**
Full Period	180	0.072%	0.029%	2.49	63.3	0.080%	0.028%	2.84	63.5
Sub-period 1	60	0.083%	0.020%	4.10	62.6	0.095%	0.018%	5.40	64.0
Sub-period 2	60	0.082%	0.038%	2.18	62.9	0.087%	0.035%	2.46	63.1
Sub-period 3	60	0.052%	0.009%	5.77	64.4	0.059%	0.010%	5.95	63.5
S&P									
Full Period	180	0.105%	0.036%	2.95	60.0	0.115%	0.033%	3.48	63.5
Sub-period 1	60	0.099%	0.021%	4.62	58.4	0.114%	0.021%	5.47	64.4
Sub-period 2	60	0.124%	0.052%	2.36	61.5	0.132%	0.048%	2.76	63.6
Sub-period 3	60	0.092%	0.009%	10.24	60.1	0.099%	0.008%	11.89	62.5

Table 5.2 Average MTP Risk under Short Sale Allowed

	Short Sale Allow								
		Raw Covariance				*Filtered Covariance*			
NASDAQ	**N**	**Avg**	**Stdev**	**Coef Var**	**W > 0**	**Avg**	**Stdev**	**Coef Var**	**W > 0**
Full Period	180	0.057%	0.021%	2.68	62.6	0.074%	0.025%	2.94	63.5
Sub-period 1	60	0.065%	0.014%	4.59	62.2	0.089%	0.016%	5.60	64.4
Sub-period 2	60	0.064%	0.028%	2.33	62.8	0.079%	0.031%	2.60	63.2
Sub-period 3	60	0.042%	0.008%	5.55	62.8	0.055%	0.010%	5.55	63.1
S&P									
Full Period	180	0.085%	0.026%	3.29	55.8	0.106%	0.027%	3.89	57.9
Sub-period 1	60	0.085%	0.018%	4.83	56.6	0.109%	0.019%	5.65	59.7
Sub-period 2	60	0.098%	0.037%	2.64	55.0	0.118%	0.038%	3.08	56.5
Sub-period 3	60	0.073%	0.007%	11.03	55.8	0.092%	0.009%	10.17	57.7

The results clearly show that the average tracking error of MTP $_{Raw}$ is lower than the tracking error of MTP $_{Filter}$ for all three sample periods as well as the full sample period. However, the information ratio or coefficient variation is higher for MTP $_{Filter}$ in all cases and in all sub periods. This means the expected return per unit of risk is higher for the filtered case than the raw case. These findings are consistent with the earlier observation that while the MTP$_{Raw}$ dominates the MTP $_{Filter}$, the filtered efficient frontier dominates the raw efficient frontier.

It is one thing to desire a more efficient portfolio during asset allocation, but it is another thing as to whether the efficient portfolio can translate into better performance. The next chapter will focus on the performance of both MVP and MTP. More specifically, I investigate whether the MVP $_{Filter}$ and MTP $_{Filter}$ can outperform their rivals MVP $_{Raw}$ and MPT $_{Raw.}$ In addition, one desires to see if these “efficient” portfolios can even outperform their market benchmarks or the naïve portfolio.

Chapter 6

Portfolio Performance of MVP

This chapter examines the return performance of MV optimization using various objective functions and constraints. Specifically, the objective is to uncover whether the “filtered” covariance matrix can lead to better security selection and whether this translates into superior performance rather than the corresponding portfolio selected based on “unfiltered” covariance matrix.

In order to isolate the influence of expected return on the portfolio optimization process, this chapter focuses on portfolio performance strictly based on optimizing risk alone. In particular, the minimum variance portfolio (MVP) and the minimum tracking error (MTP), with and without short sale constraint, are analyzed. The full sample period, from 2006 to 2013, and the three sub-periods are analyzed. In addition, the five randomly selected portfolios plus the entire sample stocks under the two benchmarks -- NSADAQ 100 and S&P 500 indices, were examined in details.

For each randomly selected sample of securities, the MVP $_{\text{Raw}}$ is first generated using the “raw” covariance matrix. The covariance matrix is computed using the entire securities sample under each benchmark using 119 log return series. The optimal weights $\{w_1^*, \ldots, w_N^*\}$ are computed via

$$\min_{w_1 \ldots w_N} \sum_{i,j=1}^{N} w_i w_j \sigma_{ij}$$

subject to

$$\sum_{i=1}^{N} w_i = 1$$

$$w_i \geq 0 \qquad \text{(no short sale constraint)}$$

The optimal portfolio is held for the next 10 business days (proxy for bi-weekly) and its realized return is computed as:

$$R_p^A = \sum_{i=1}^{Np} w_i^* R_i \qquad p = 1, \ldots, 6 \ \ portfolios$$

The portfolio is then re-optimized using the newly generated covariance matrix, based on moving 119 log return series. The process is repeated for every 10 business days, yielding a total of 181 realized returns. The same experiment is repeated using the "filtered" covariance matrices to generate the $\text{MVP}_{\text{Filter.}}$

Both naïve (or equally weighted) portfolio and the benchmark index (value weighted) are included as side-by-side comparison. However, one cannot directly compare the performance of the naïve portfolio to the MVPs because the former has higher ex ante risk. One would expect the naive portfolio to outperform the MVP over time because the latter has lower volatility. In other words, it is not a "horse-race" between the MVPs and the naive strategy. Rather it is a horse-race between the two MVPs.

Each of the four portfolios' realized returns are geometrically compounded into an equivalent "Index" with 100 as the base value on 6/23/2006:

$$Index_{k,t} = 100 * \prod_{t=1}^{T}(1 + R_{kt}^{A}) \quad for\ k = 1, .., 4\ and\ t = 1, \dots, T$$

The compounded annual growth rate (CAGR) is computed as:

$$CAGR_k = 26 * \left(\left(\frac{Index_{k,T}}{100}\right) \wedge \left(\frac{1}{T}\right) - 1\right)$$

The Sharpe ratio is computed as:

$$Sharpe_k = \sum_{t=1}^{T} \frac{\left(R_{kt}^{A} - R_t^{rf}\right)}{T} \Big/ \sigma_k \quad \text{where} \quad \sigma_k = \sqrt{\sum_{t=1}^{T} \frac{\left(R_{kt}^{A} - \bar{R}\right)}{T-1}}$$

Table 6.1 (page 89) shows the performance of the MVP for Portfolio A. The results for all other portfolios are contained in Appendix C. The shaded line represents the best performing portfolio based on the CAGR. The bold number under the Sharpe column indicates the best performance based on Sharpe ratio.

In this table, Port A (Filter) scores 5 wins, EQW scores 2 wins based on CAGR. Port A (Raw) and the NASDAQ 100 benchmark index both underperform using this criteria. Using the Sharpe ratio as the performance measurement, Port A (Filter) scores 5 wins, EQW scores 2 wins and Port A (Raw) also scores 2 wins. The benchmark did not score a win. But this type of analysis is not complete because the focus is only on winners. Therefore, I devise a numerical scoring system. Under each performance measure, CAGR and Sharpe, one assigns a 4 to the winner, 3 to the runner up and 1 to the worst performer for each Portfolio. The average score across all 5 portfolios in each sample are tallied. The highest possible score is 4.0 and the lowest possible score is 1.0. Table 6.2 (page 90) contains the ranking score under the **Short Sale restriction**.

First, looking at the CAGR score, one can see that the naïve portfolio (EQW) outperforms the MVP $_{Raw}$, MVP $_{Filter}$ and the benchmark Indexes for the full sample period. MVP $_{Filter}$ beats the competition during the first period; EQW wins in the second period while MVP $_{Raw}$ wins the last period. The results hold for both NASDAQ and S&P stocks. In general, the benchmark NASDAQ 100 performs the worst.

Second, using the Sharpe scores, MVP $_{Filter}$ beats the others during the first period and the full period. EQW performs best in the second period while MVP $_{Raw}$ wins the last period. The results are for the NASDAQ sample. However, in the S&P sample, the naïve portfolio clearly dominates during the entire period and during the second period. On the other hand, MVP $_{Filter}$ wins the first and last periods. Again, the benchmark NASDAQ 100 performs the worst.

The total scores indicate that the naïve strategy is the overall best performer, followed by MVP $_{Filter}$, MVP $_{Raw}$ and the benchmark index. MVP $_{Filter}$ appears to slightly outperform the MVP $_{\text{Raw}}$. This bears out for the full period and for both NASDAQ and S&P samples. Surprisingly, all three portfolios appear to easily outperform the benchmark indexes.

The experiment is repeated for the case when Short Sale restriction is removed. The results for all the portfolios are contained in Appendix C. Table 6.3 (page 91) presents the results of the scores. Using the CAGR scores, the naive portfolio clearly dominates the full period. The average CAGR is 15.81% as compared to the NASDAQ CAGR of 9.88%. The naive's average CAGR for the S&P sample is 9.13% as compared to 4.77% of MVP $_{\text{Filter}}$. The runner up is clearly MVP $_{\text{Filter.}}$ Using the Sharpe scores, EQW wins both the full and second periods whereas MVP $_{\text{Filter}}$ wins the first and third periods. Basing on the total score, the conclusion for the short sale restriction case is that the naïve strategy is the winner, followed by MVP $_{\text{Filter}}$ (at least for the S&P sample), but both MVPs are neck-to-neck in the NASDAQ sample.

Another simple measure to confirm this conclusion is to compute the excess return of each portfolio return over the benchmark return. In my case, the benchmark is either the NASDAQ 100 Index or the S&P 500 Index. Defining the "Benchmark Excess Return" for portfolio A as:

$$R_{A,t}^{excess} = R_{A,t} - R_{Benchmark,t} \; for \; t = 1, \ldots., T$$

where the average excess return is simply

$$\bar{R}_{A,excess} = \left(\sum_{t=1}^{T} R_{A,t}^{excess}\right)/T \; ,$$

and dividing the mean excess return by its standard deviation gives the "Benchmark Excess Ratio" or *BER*:

$$BER_A = \bar{R}_{A,excess}/\sigma_{A,excess}$$

A positive value for BER_A implies that portfolio A outperforms the benchmark index. A negative value implies underperforming the benchmark index. The higher the value, the more superior is the excess return per unit of volatility of excess return. This is a more stringent performance measure than the Sharpe's ratio as the portfolio must be able to beat the benchmark itself. It is equivalent to a ***risk adjusted alpha*** measure. Table 6.4 presents the result of BER_K for the raw MVP, the filtered MVP and the naïve portfolio.

Table 6.4 Benchmark Excess Ratio (BER_K) for MVP (2006-2013)

$\boldsymbol{BER_K}$

2006 to 2013	No Short Sale			Short Sale Allow		
NASDAQ	MVP Raw	MVP Filter	Naïve	MVP Raw	MVP Filter	Naïve
All stocks	2.47%	4.28%	19.38%	-3.40%	-2.48%	19.38%
Port A	1.86%	3.83%	6.51%	-0.05%	0.21%	6.51%
Port B	-4.00%	-3.56%	15.05%	-9.00%	-11.16%	15.05%
Port C	3.96%	5.52%	10.78%	0.27%	0.50%	10.78%
Port D	9.11%	6.32%	22.31%	0.79%	0.38%	22.31%
S&P						
All stocks	-2.47%	-1.37%	19.88%	0.61%	1.71%	19.88%
Port A	-1.72%	-1.36%	12.77%	-2.33%	-1.41%	12.77%
Port B	-1.02%	0.17%	13.51%	2.77%	3.18%	13.51%
Port C	-3.43%	-0.72%	13.86%	-3.02%	-0.85%	13.86%
Port D	1.20%	0.70%	17.76%	0.80%	0.61%	17.76%

The results clearly show that the naïve portfolio beats the NASDAQ 100 Index for all cases and its BER is far larger than both MVPs can achieve. The MVP for the filter case appears to do better than the raw MVP when there is no short sale constraint. Without the short sale constraint, the results are not conclusive. The following graphs summarize the portfolio performances for the full sample period.

Figure 6.1A   MVP Performance for NASDAQ Full Sample – No Short Sale

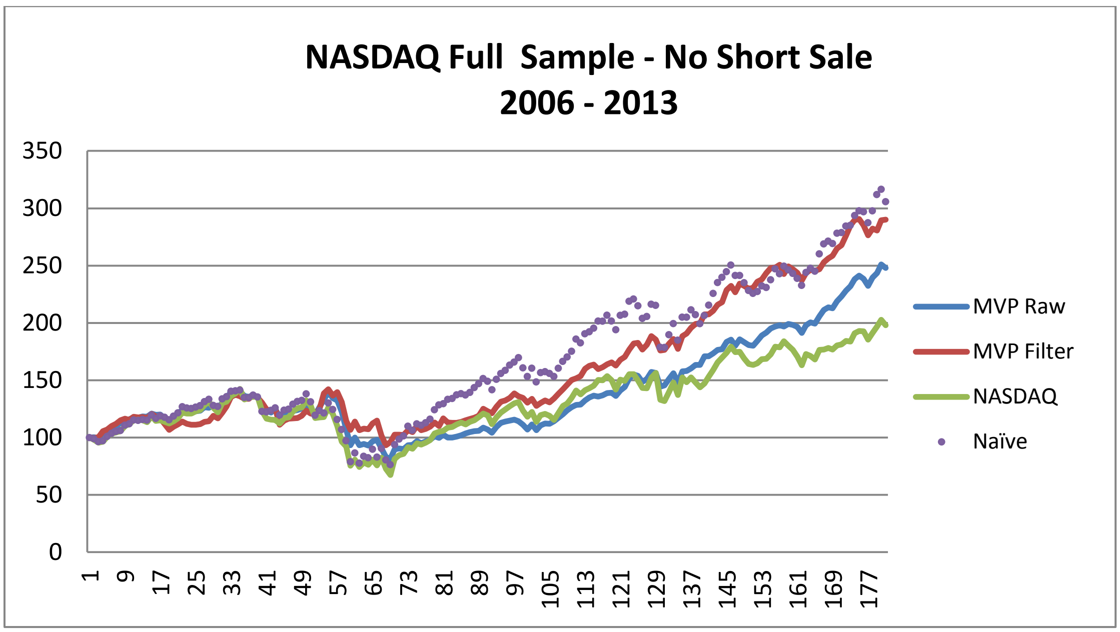


Figure 6.1B MVP Performance for **NASDAQ Full Sample – Short Sale Allow**

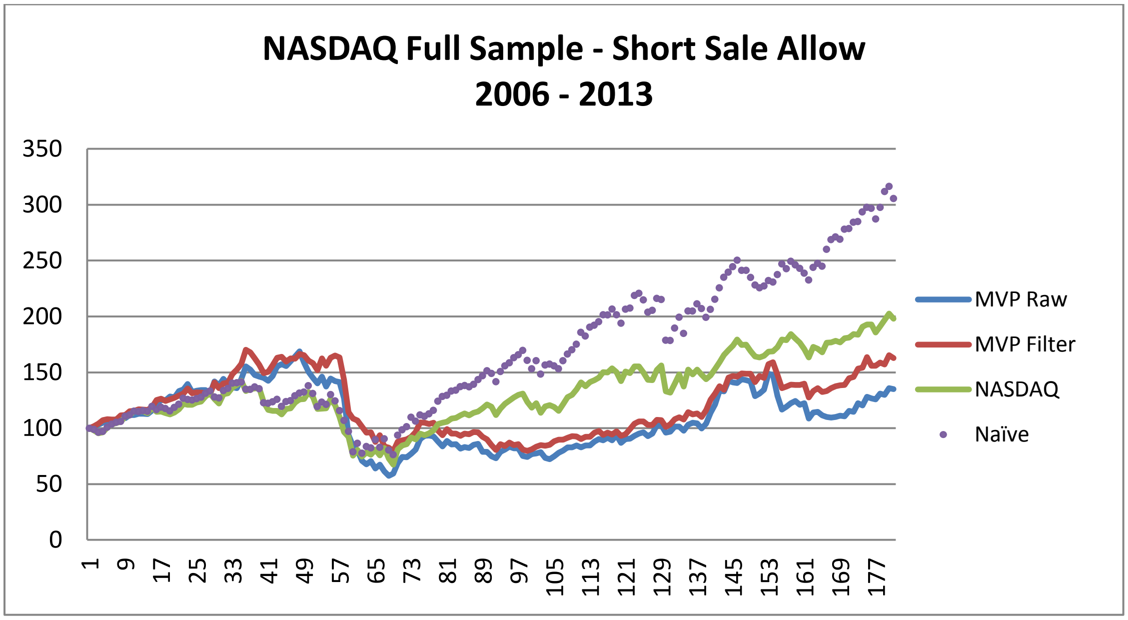


Figure 6.2A MVP Performance for S&P Full Sample – No Short Sale

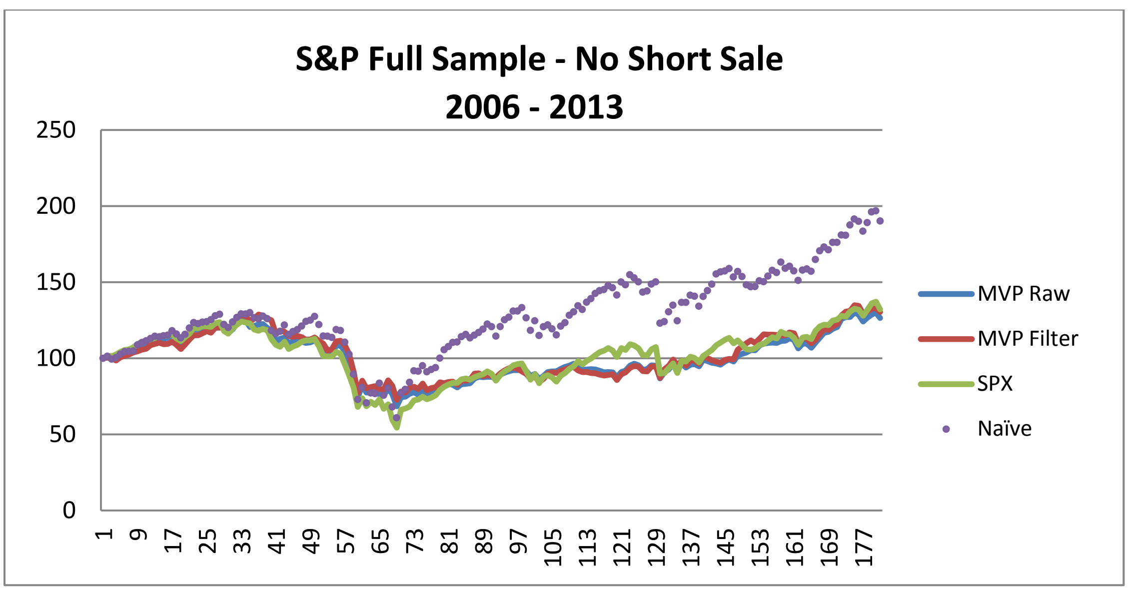


Figure 6.2B **MVP Performance for S&P Full Sample – Short Sale Allow**

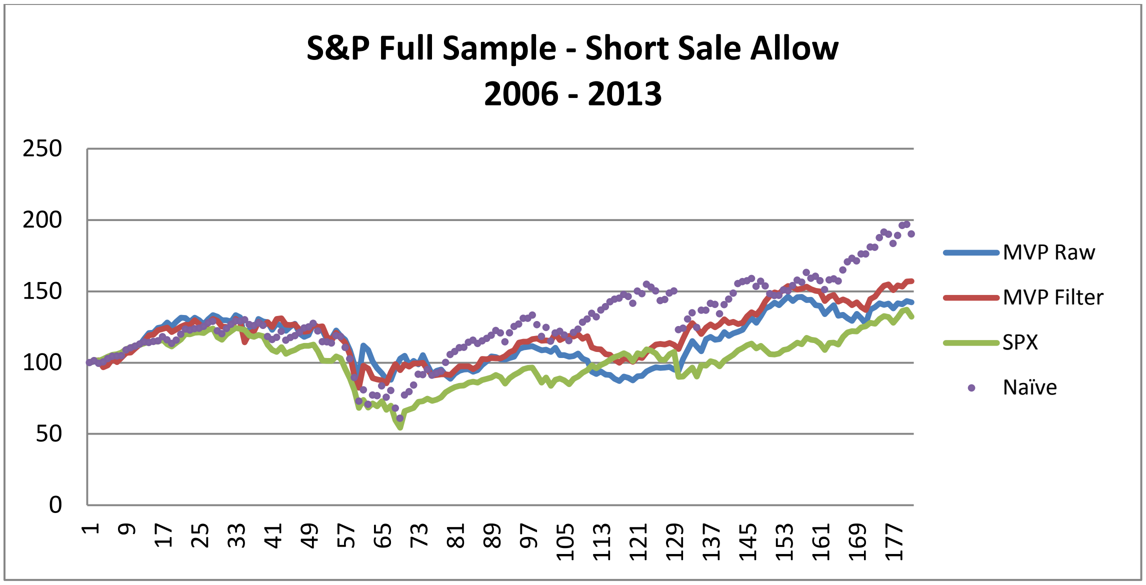


Table 6.1	Performance of MVP					
	No Short Sale			Short Sale Allowed		
	Terminal	**CAGR%**	**Sharpe**	**Terminal**	**CAGR%**	**Sharpe**
Full Period						
NASDAQ 100	197.91	9.88	0.107	197.91	9.88	0.107
EQW	244.11	12.92	0.132	244.11	12.92	0.132
Port A (Raw)	242.81	12.85	0.175	218.29	11.30	**0.159**
Port A (Filter)	284.68	15.16	**0.197**	217.20	11.23	0.141
Period 1						
NASDAQ 100	80.66	-9.30	-0.053	80.66	-9.30	-0.053
EQW	73.61	-13.24	-0.098	73.61	-13.24	-0.098
Port A (Raw)	98.69	-0.57	0.009	106.93	2.90	0.053
Port A (Filter)	102.36	1.01	**0.029**	121.55	8.47	**0.112**
Period 2						
NASDAQ 100	186.77	27.21	0.238	186.77	27.21	0.238
EQW	220.75	34.54	**0.271**	220.75	34.54	**0.271**
Port A (Raw)	137.68	13.89	0.176	109.33	3.87	0.062
Port A (Filter)	154.16	18.82	0.228	89.42	-4.84	-0.031
Period 3						
NASDAQ 100	131.38	11.85	0.136	131.38	11.85	0.136
EQW	150.23	17.70	0.204	150.23	17.70	0.204
Port A (Raw)	178.69	25.28	0.382	186.73	27.20	**0.374**
Port A (Filter)	180.40	25.69	**0.387**	199.83	30.17	0.371

Table 6.2

MVP Performance under No Short Sale

		CAGR Ranking				**Sharpe Ranking**		
	NASDAQ	EQW	Raw	Filter	NASDAQ	EQW	Raw	Filter
Full Period	1.40	**3.60**	2.20	2.80	1.40	2.80	2.80	**3.00**
Period 1	1.20	2.20	2.60	**4.00**	1.20	2.40	2.60	**3.80**
Period 2	2.80	**4.00**	1.80	1.40	2.60	**4.00**	2.00	1.40
Period 3	1.60	2.20	**3.20**	3.00	1.40	2.00	**3.40**	3.20
Total	7.00	12.00	9.80	11.20	6.60	11.20	10.80	11.40
	S&P 500	EQW	Raw	Filter	S&P 500	EQW	Raw	Filter
Full Period	2.00	**4.00**	1.60	2.40	1.80	**4.00**	1.80	2.60
Period 1	1.20	2.20	3.00	**3.60**	1.00	2.80	2.80	**3.40**
Period 2	3.00	**4.00**	2.00	1.00	3.20	**3.80**	2.00	1.00
Period 3	1.60	2.00	2.60	**3.80**	1.60	1.80	2.80	**3.80**
Total	7.80	12.20	9.20	10.80	7.60	12.40	9.40	10.80
		Average CAGR%				**Average Sharpe**		
	NASDAQ	EQW	Raw	Filter	NASDAQ	EQW	Raw	Filter
Full Period	9.88	**15.81**	12.61	13.58	0.107	0.152	0.154	**0.158**
Period 1	(9.30)	(6.71)	(2.49)	**0.09**	(0.053)	(0.017)	(0.001)	**0.026**
Period 2	27.21	**37.51**	17.96	17.64	0.238	**0.298**	0.210	0.198
Period 3	11.85	16.82	22.44	**23.07**	0.136	0.181	0.306	**0.314**
	S&P 500	EQW	Raw	Filter	S&P 500	EQW	Raw	Filter
Full Period	4.04	**9.13**	3.81	4.30	0.059	**0.097**	0.063	0.067
Period 1	(13.10)	(9.38)	(6.49)	**(5.49)**	(0.125)	(0.059)	(0.053)	**(0.041)**
Period 2	16.00	**26.74**	4.97	2.46	0.150	**0.198**	0.074	0.044
Period 3	9.31	10.17	13.00	**15.97**	0.121	0.120	0.200	**0.227**

Table 6.3

MVP Performance under Short Sale Allow

	CAGR Ranking				**Sharpe Ranking**			
	NASDAQ	EQW	Raw	Filter	NASDAQ	EQW	Raw	Filter
Full Period	2.00	**4.00**	2.40	1.60	2.40	**3.60**	2.60	1.60
Period 1	2.20	**2.80**	2.20	**2.80**	1.80	2.80	2.20	**3.20**
Period 2	3.00	**4.00**	2.00	1.00	3.00	**4.00**	2.00	1.00
Period 3	1.20	1.80	3.00	**4.00**	1.20	2.00	3.00	**3.80**
Total	8.40	12.60	9.60	9.40	8.40	12.40	9.80	9.60
	S&P 500	EQW	Raw	Filter	S&P 500	EQW	Raw	Filter
Full Period	1.80	**4.00**	2.00	2.20	1.80	**3.60**	2.40	2.20
Period 1	1.20	2.20	3.20	**3.40**	1.00	2.40	3.20	**3.60**
Period 2	3.00	**4.00**	1.80	1.20	3.20	**3.80**	1.80	1.20
Period 3	1.40	2.00	2.80	**3.80**	1.80	1.80	3.00	**3.80**
Total	7.40	12.20	9.80	10.60	7.80	11.60	10.40	10.80

	Average CAGR%				**Average Sharpe**			
	NASDAQ	EQW	Raw	Filter	NASDAQ	EQW	Raw	Filter
Full Period	9.88	**15.81**	7.92	6.47	0.107	**0.152**	0.101	0.082
Period 1	(9.30)	(6.71)	(5.87)	**(3.33)**	(0.053)	(0.017)	(0.021)	**0.015**
Period 2	27.21	**37.51**	7.22	(4.32)	0.238	**0.298**	0.087	(0.019)
Period 3	11.85	16.82	22.51	**27.20**	0.136	0.181	0.265	**0.288**
	S&P 500	EQW	Raw	Filter	S&P 500	EQW	Raw	Filter
Full Period	4.04	**9.13**	4.32	4.77	0.059	**0.097**	0.066	0.069
Period 1	(13.10)	(9.38)	**(0.03)**	(0.21)	(0.125)	(0.059)	0.018	**0.021**
Period 2	16.00	**26.74**	(3.39)	(4.32)	0.150	**0.198**	(0.017)	(0.099)
Period 3	9.31	10.17	16.43	**18.92**	0.121	0.120	0.249	**0.278**

Chapter 7

Portfolio Performance of MTP

In the last chapter, the performance of the minimum variance portfolio (MVP) generated by both raw and filtered covariance matrices were compared to the benchmark index and the naïve portfolio. The naïve portfolio largely outperforms both MVPs. The filtered MVP is only marginally better than the raw MVP. All three portfolios generally outperform the benchmark indexes.

In the same spirit, this chapter examines the return performance of the minimum tracking portfolios (MTP) generated by both raw and filtered covariance matrices. Their performances are then compared to the naïve portfolio and the benchmark index. The same full sample period (2006 to 2013) and three sub periods are analyzed as well as the 5 random portfolios plus the full sample size for both NASDAQ and S&P samples.

All detailed Tables and graphs are contained in Appendix D. Only the summary of the performance results are shown here. These include the ranking score tables and the Benchmark Excess Ratio (BER). Table 7.1 gives the performance scores for the case where short sale is restricted. Table 7.2 shows the results when short sale is allowed.

Table 7.1

MTP Performance under No Short Sale								
	CAGR Ranking				**Sharpe Ranking**			
	NASDAQ	**EQW**	**Raw**	**Filter**	**NASDAQ**	**EQW**	**Raw**	**Filter**
Full Period	1.40	**3.60**	2.80	2.20	1.40	**4.00**	2.60	2.00
Period 1	1.60	2.60	**3.20**	2.60	1.60	2.60	**3.20**	2.60
Period 2	1.80	**4.00**	2.00	2.20	1.80	**3.60**	2.60	2.40
Period 3	1.60	**3.40**	2.40	2.60	1.60	**3.40**	2.40	2.60
Total	6.40	13.60	10.40	9.60	6.40	13.60	10.80	9.60
	S&P 500	**EQW**	**Raw**	**Filter**	**S&P 500**	**EQW**	**Raw**	**Filter**
Full Period	1.80	**4.00**	2.40	1.80	1.80	**4.00**	2.40	2.00
Period 1	2.20	**3.40**	2.60	1.80	1.00	**3.60**	3.00	2.60
Period 2	2.20	**3.20**	2.40	2.20	2.00	**3.40**	2.80	2.00
Period 3	1.60	**3.20**	2.40	2.80	2.40	2.80	2.00	**3.00**
Total	7.80	13.80	9.80	8.60	7.20	13.80	10.20	9.60
	Average CAGR%				**Average Sharpe**			
	NASDAQ	EQW	Raw	Filter	NASDAQ	EQW	Raw	Filter
Full Period	9.88	**15.81**	13.14	12.91	0.11	**0.15**	0.13	0.13
Period 1	(9.30)	**(6.71)**	(7.08)	(7.43)	(0.05)	**(0.03)**	(0.04)	(0.04)
Period 2	27.21	**37.51**	30.70	30.08	0.24	**0.30**	0.26	0.26
Period 3	11.85	**16.82**	15.95	16.22	0.14	**0.18**	0.17	0.17
	S&P 500	EQW	Raw	Filter	S&P 500	EQW	Raw	Filter
Full Period	4.04	**8.71**	5.13	4.81	0.06	**0.10**	0.07	0.07
Period 1	(13.10)	**(8.69)**	(11.29)	(11.79)	(0.13)	**(0.06)**	(0.08)	(0.08)
Period 2	16.00	**25.33**	17.75	17.44	0.15	**0.20**	0.17	0.16
Period 3	9.31	**9.60**	9.03	8.87	**0.12**	0.12	0.12	0.11

Using both CAGR and Sharpe scores, Table 7.1 shows that the naïve portfolio outperforms the Benchmark, the MPT $_{\text{Filter}}$ and MPT $_{\text{Raw}}$. The total score shows that MPT

Raw beats the MPT Filter. ..Surprisingly, the worst performing portfolio are the two benchmark indexes. This means that using any random portfolios of 20 stocks, with or without optimization, one can easily beat the market benchmark index. Figures 7.1A and 7.1B shows the performances of these portfolios (full sample size) under the short sale constraint. After the 2008 – 2009 crash, the naïve portfolio clearly dominates all the other portfolios. Both MTPs are tracking their respective benchmarks closely during the first period but deviate further after the second period.

Figure 7.1A NASDAQ Full Sample MTP Performance – No Short Sale

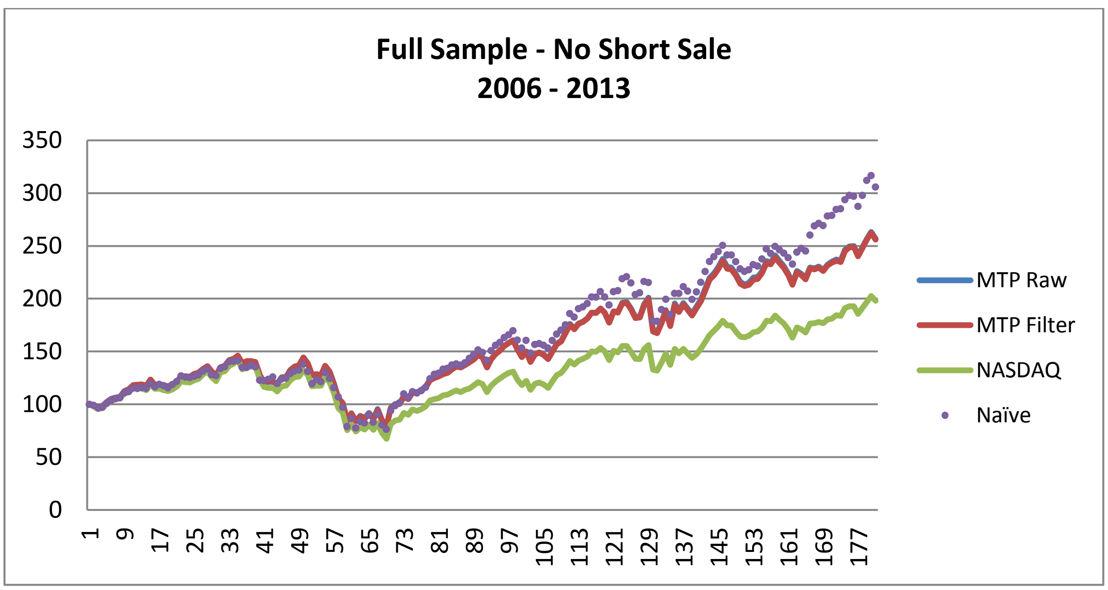


Figure 7.1B S&P Full Sample MTP Performance – No Short Sale

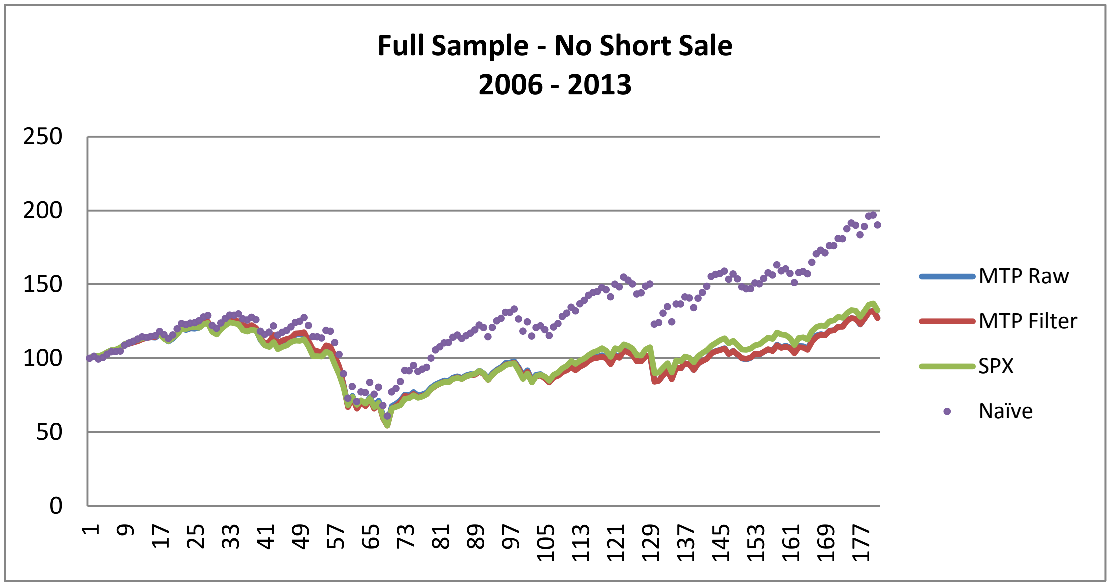


Table 7.2

MTP Performance under Short Sale Allowed

	CAGR Ranking				**Sharpe Ranking**			
	NASDAQ	EQW	Raw	Filter	NASDAQ	EQW	Raw	Filter
Full Period	1.40	**4.00**	2.20	2.40	1.40	**3.80**	2.20	2.60
Period 1	2.00	**2.80**	2.40	**2.80**	2.00	**3.00**	2.40	2.80
Period 2	1.80	**4.00**	1.80	2.40	1.40	**3.60**	2.20	2.80
Period 3	1.60	**3.80**	2.20	2.40	1.60	**3.60**	2.40	2.60
Total	6.80	14.60	8.60	10.00	6.40	14.00	9.20	10.80
	S&P 500	EQW	Raw	Filter	S&P 500	EQW	Raw	Filter
Full Period	2.20	**4.00**	2.20	1.60	2.20	**4.00**	2.40	1.40
Period 1	2.40	**3.40**	2.40	1.80	1.20	**3.60**	3.00	2.20
Period 2	2.40	**3.60**	2.20	1.80	2.40	**3.80**	2.20	1.80
Period 3	1.60	**3.20**	2.20	3.00	2.20	2.80	2.40	**3.00**
Total	8.60	14.20	9.00	8.20	8.00	14.20	10.00	8.40
	Average CAGR%				**Average Sharpe**			
	NASDAQ	EQW	Raw	Filter	NASDAQ	EQW	Raw	Filter
Full Period	9.88	**15.81**	12.23	12.29	0.107	**0.152**	0.128	0.128
Period 1	(9.30)	**(6.71)**	(7.27)	(7.39)	(0.053)	**(0.033)**	(0.041)	(0.040)
Period 2	27.21	**37.51**	30.29	30.46	0.238	**0.298**	0.265	0.266
Period 3	11.85	**16.82**	13.81	13.96	0.136	**0.181**	0.154	0.155
	S&P 500	EQW	Raw	Filter	S&P 500	EQW	Raw	Filter

Full Period	4.04	**8.71**	3.89	3.76	0.059	**0.097**	0.058	0.057
Period 1	(13.10)	**(8.69)**	(11.71)	(12.31)	(0.125)	**(0.059)**	(0.084)	(0.090)
Period 2	16.00	**25.33**	14.38	14.38	0.150	**0.198**	0.137	0.136
Period 3	9.31	**9.60**	9.09	9.29	**0.121**	0.120	0.115	0.117

The results in Table 7.2 confirm the superiority of the naïve strategy. In all four cases, it outscores the other three portfolios. While the MTP $_{\text{Filter}}$ beats the MTP $_{\text{Raw}}$ in the NASDAQ sample, the opposite result occurs in the S&P sample. As before, the worst performers are the two benchmark indexes. Therefore, there is no clear advantage using the filtered covariance over the raw covariance. Their results are almost identical. In fact, both MTPs underperformed the S&P Index. The graphs of the performances are shown in Figures 7.2A and 7.2B.

Figure 7.2A Performance of MTP (NASDAQ sample) – Short Sale Allow

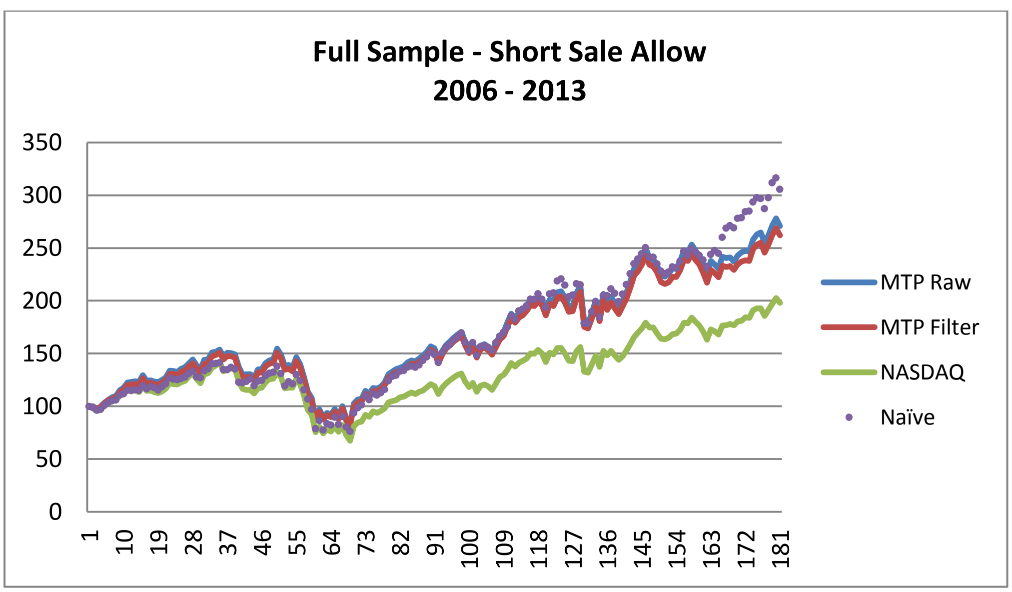
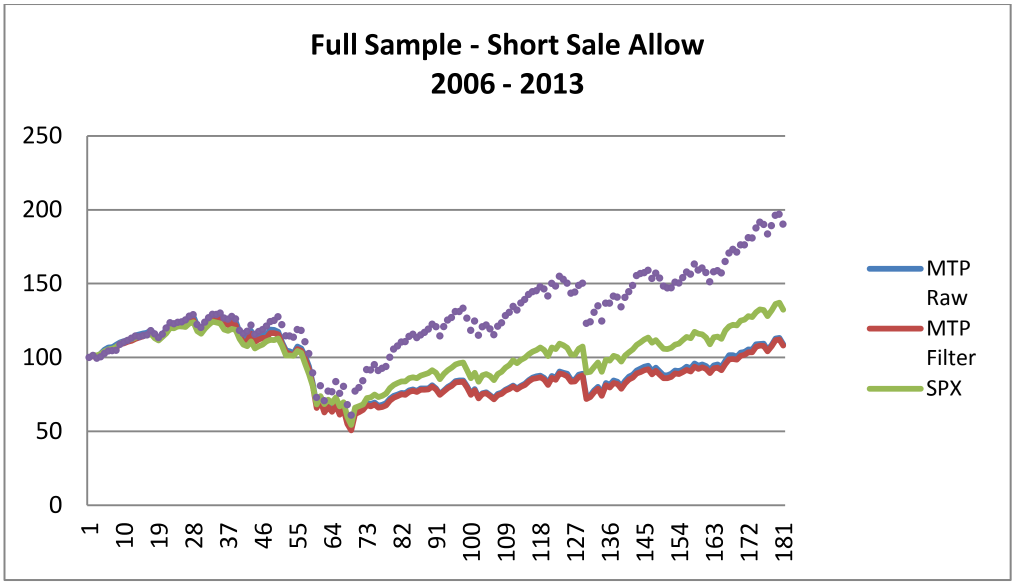


Figure 7.2B Performance of MTP (S&P sample) – Short Sale Allow



The results of the Benchmark Excess Ratio (BER) are reported in Table 7.3

Table 7.3

Benchmark Excess Ratio of MTPs

2006 to 2013

	No Short Sale			Short Sale Allow		
NASDAQ	Raw	Filter	Naïve	Raw	Filter	Naïve
All stocks	27.22%	26.55%	19.38%	23.25%	21.82%	19.38%
Port A	-5.01%	-5.55%	6.51%	-4.87%	-5.08%	6.51%
Port B	14.22%	14.90%	15.05%	14.81%	14.99%	15.05%
Port C	9.42%	7.66%	10.78%	3.35%	4.92%	10.78%
Port D	10.89%	10.47%	22.31%	7.28%	7.65%	22.31%
S&P						
All stocks	-1.36%	-1.29%	19.88%	-10.33%	-11.44%	19.88%
Port A	-2.72%	-2.66%	12.77%	-3.30%	-2.91%	12.77%
Port B	9.65%	8.54%	13.51%	9.52%	8.33%	13.51%
Port C	5.45%	3.97%	13.86%	-0.82%	-1.22%	13.86%
Port D	8.64%	7.13%	17.76%	8.07%	7.08%	17.76%

For the NASDAQ 87 stocks sample, both MTPs perform better than the naïve portfolio, with MTP $_{Raw}$ showing the highest Benchmark Excess Ratio. But the naïve portfolio shows the best BER for all five random portfolios (A to D). For the S&P 80 stocks sample, the naïve strategy clearly beats both raw and filtered MTPs and also for portfolios A to D.

The final test is to see if the mean excess return from the raw MTP is statistically different from the mean excess return of the filtered MTP. First, the average and standard deviations of the excess returns are contained in Table 7.4.

Table 7.4

Benchmark Excess Return of MTPs

2006 to 2013

	No Short Sale				**Short Sale Allow**			
	Raw		Filtered		Raw		Filtered	
NASDAQ	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev
All stocks	0.00144	0.00528	0.00142	0.00535	0.00179	0.00770	0.00159	0.0073
Port A	-0.00076	0.01514	-0.00087	0.01567	-0.00074	0.01513	-0.00079	0.0155
Port B	0.00156	0.01098	0.00169	0.01135	0.00165	0.01114	0.00175	0.0116
Port C	0.00268	0.02846	0.00239	0.03126	0.00056	0.01664	0.00082	0.0167
Port D	0.00167	0.01537	0.00162	0.01550	0.00117	0.01607	0.00124	0.0161
S&P	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev
All stocks	-0.00010	0.00739	-0.00010	0.00746	-0.00094	0.00915	-0.00099	0.0086
Port A	-0.00045	0.01650	-0.00043	0.01618	-0.00055	0.01659	-0.00048	0.0163
Port B	0.00106	0.01104	0.00094	0.01102	0.00107	0.01120	0.00093	0.0111
Port C	0.00120	0.02211	0.00087	0.02199	-0.00009	0.01099	-0.00014	0.0111
Port D	0.00099	0.01150	0.00086	0.01203	0.00095	0.01173	0.00087	0.0123

The null hypothesis is that

$$H_0\colon\ \mu_{Raw} = \mu_{Filter}$$

where μ_{Raw} is the population mean excess return over the benchmark for raw covariance, and μ_{Filter} is the population mean excess return for filtered covariance. The t-statistic or t-score is used to test whether the two sample means conform to the null hypothesis. This is defined as

$$t - score\ = (\bar{x}_{Raw} - \bar{x}_{Filter})/SE$$

where $$SE = \sqrt{\frac{s_{Raw}{}^2}{N} + \frac{S_{Filter}{}^2}{N}}$$

$\bar{x}_{Raw}, \bar{x}_{Filter}$ are the two sample means, and

$S_{Raw}{}^2,\ S_{Filter}{}^2$ are the two sample variances.

The degree of freedom can be computed as

$$DF = \left(S^2_{Raw}/_N + \ S^2_{Filter}/_N\right)^2 / \{\frac{\left[\frac{S^2_{Raw}}{N}\right]^2}{N-1} + \frac{\left[\frac{S^2_{Filter}}{N}\right]^2}{N-1}\}$$

The P-value is the probability that a t-score having DF degrees of freedom is more extreme than the computed t-value. The test is actually a two-tail test of the null hypothesis. The results are shown in Table 7.5

Table 7.5

Testing Mean Excess Return of MTPs

NASDAQ	**No Short Sale**				**Short Sale Allow**			
				1-tail				1-tail
	t-score	SE	DF	p-value	t-score	SE	DF	p-value
All stocks	0.028	0.00056	358	0.489	0.249	0.00079	357	0.402
Port A	0.069	0.00162	358	0.473	0.031	0.00161	358	0.488
Port B	-0.110	0.00118	358	0.456	-0.082	0.00120	357	0.467
Port C	0.091	0.00315	355	0.464	-0.150	0.00176	358	0.440
Port D	0.031	0.00163	358	0.488	-0.038	0.00170	358	0.485

S&P				1-tail				1-tail
	t-score	SE	DF	p-value	t-score	SE	DF	p-value
All stocks	-0.005	0.00078	358	0.498	0.046	0.000938	357	0.482
Port A	-0.012	0.00172	358	0.495	-0.042	0.001735	358	0.483
Port B	0.106	0.00116	358	0.458	0.118	0.001178	357	0.453
Port C	0.143	0.00232	358	0.443	0.039	0.001164	358	0.485
Port D	0.110	0.00124	357	0.456	0.058	0.001269	358	0.477

Note: 1-tail p-value = P(T > t-score) or P(T ≤ - t-score)

In order to interpret the results, take the example of t-score = 0.249 for the NASDAQ (all stocks) when short sale is allowed. The probability P(T ≤ 0.249) = 0.598. For a one-tail test, then P(T > 0.249) = 1 – 0.598 = 0.402. The p-value that P(T < -0.249) = 0.402. Since it is a 2-tail test, then one is interested in the probability that P(-0.249 >T> 0.249) = 0.402 + 0.402 = **0.804.** Since this probability is greater than the significance level (0.05), one cannot reject the null hypothesis.

The p-value is Table 7.5 is showing only the 1-tail P-value. One can easily approximate the two-tail test by multiplying the 1-tail P-value by two. Since the lowest P-value in the table is 0.402 which far exceeds the significance level of 0.05, one cannot reject the null hypothesis that the mean excess return from the MTP $_{Raw}$ is the same as the MTP $_{Filter}$. This gives concrete proof that the performances between the two MTPs are identical.

I now repeat the same test on the mean excess return between MVP $_{Raw}$ and MVP$_{Filter}$ that was analyzed in Chapter 6. The means and standard deviation of the excess benchmark returns are given in Table 7.6.

Table 7.6

Benchmark Excess Return of MVPs

2006 to 2013

	No Short Sale				**Short Sale Allow**			
	Raw		Filtered		Raw		Filtered	
NASDAQ	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev
All stocks	0.00073	0.02945	0.00164	0.03824	-0.00188	0.05526	-0.00127	0.05126
Port A	0.00060	0.03209	0.00153	0.03995	-0.00002	0.04125	0.00011	0.05266
Port B	-0.00106	0.02653	-0.00136	0.03811	-0.00305	0.03387	-0.00558	0.05003
Port C	0.00106	0.02686	0.00205	0.03706	0.00011	0.03947	0.00029	0.05877
Port D	0.00219	0.02401	0.00204	0.03221	0.00030	0.03875	0.00020	0.05294
S&P	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev
All stocks	-0.00061	0.02478	-0.00038	0.02799	0.00030	0.04861	0.00073	0.04260
Port A	-0.00041	0.02405	-0.00043	0.03180	-0.00085	0.03632	-0.00067	0.04777
Port B	-0.00026	0.02536	0.00005	0.02934	0.00095	0.03421	0.00132	0.04160
Port C	-0.00090	0.02639	-0.00021	0.02953	-0.00106	0.03508	-0.00038	0.04490
Port D	0.00032	0.02666	0.00021	0.02980	0.00031	0.03921	0.00030	0.04983

Table 7.7

Testing Mean Excess Return of MVPs

	No Short Sale				Short Sale Allow			
NASDAQ	t-score	SE	DF	1-tail p-value	t-score	SE	DF	1-tail p-value
All stocks	-0.253	0.00360	336	0.400	-0.108	0.00562	356	0.457
Port A	-0.244	0.00382	342	0.404	-0.026	0.00499	339	0.490
Port B	0.086	0.00346	319	0.466	0.563	0.00450	315	0.287
Port C	-0.288	0.00341	326	0.387	-0.035	0.00528	313	0.486
Port D	0.050	0.00299	331	0.480	0.021	0.00489	328	0.492
S&P	t-score	SE	DF	1-tail p-value	t-score	SE	DF	1-tail p-value
All stocks	-0.083	0.00279	336	0.467	-0.090	0.004817	352	0.464
Port A	0.006	0.00297	342	0.498	-0.038	0.004473	334	0.485
Port B	-0.107	0.00289	319	0.457	-0.093	0.004014	345	0.463
Port C	-0.234	0.00295	326	0.408	-0.160	0.004247	338	0.437
Port D	0.038	0.00298	331	0.485	0.003	0.004726	339	0.499

Note: 1-tail p-value = P(T > t-score) or P(T ≤ - t-score)

The highest t-score in Table 7.7 is 0.563 for Portfolio B, under Short Sale Allow scenario. The 1-tail P-value is 0.287 which far exceeds the 0.05 significance level. The P-values in the table clearly shows that one cannot reject the null hypothesis. Therefore, the MVP's excess return over the benchmark index is insignificantly different between the performance generated by the raw covariance and the performance generated by the filtered covariance.

Chapter 8

Portfolio Performance for Riskier Portfolios

In the last two chapters, the performance of the minimum variance portfolios (MVP) and the minimum tracking portfolios (MTP) were compared to the benchmark index and the naïve portfolio. The naïve portfolio largely outperforms the benchmark index, the MVPs and the MTPs. Moreover, there is no significant difference between the performance, as measured by the excess return over the benchmark, of the raw MVP and the filtered MVP. The same conclusion holds for the performance of the MTPs.

Some may argue that the reason both MVP and MTP are poor performers is because the expected returns are not used in the portfolio optimization. After all, these are "minimal" risk portfolios and as such should expect lower realized returns to reflect for the lower ex ante risk. That may explain why they cannot outperform the naïve portfolio. Therefore in this chapter the objective function is modified such that expected return of the optimal portfolio is set to be equal to or greater than the expected return of the benchmark. The optimal weights $\{w_1^*, \ldots., w_N^*\}$ are computed via

$$\min_{w_1 \ldots w_N} \sum_{i=1}^{N} \sum_{j=1}^{N} w_i \, w_j \sigma_{ij}$$

subject to

$$\sum_{i=1}^{N} w_i \bar{R}_i \geq \bar{R}_{Benchmark} \tag{40}$$

$$\sum_{i=1}^{N} w_i = 1$$

Expected return is normally estimated from historical returns and in general not reliable forecast of future return or it may be subject to certain bias or estimation error. In light of this, the constraint in (40) takes the difference between the portfolio expected return and the benchmark into consideration rather than setting a fixed level for the expected return.

One disadvantage for using the mean return as an estimate for expected return is that both older data and newer data are assigned equal weight. My empirical return series uses 120 daily prices, roughly 6 months of trading data. Using the mean return implies that the return for the last six months is as relevant as the most current return. In order to incorporate the fact that more recent data carries more weight than older data, I adopted a declining weight approach assigned to observations as they go further back in time (see Litterman and Winkelmann (1998). The weight is given as:

$$\gamma_t = \lambda^{T-t+1} \Big/ \sum_{t=1}^{T} \lambda^t$$

and the weighted average return is computed as:

$$\bar{R} = \sum_{t=1}^{T} \gamma_t R_t \qquad (41)$$

The value for the parameter λ is set to 0.75 and T = 119 for all stocks and for the benchmark returns. Larger weight is given to the latest return data than older returns. For λ = 1, (41) is simply the mean return.

In this study only the unconstraint short sale case is examined.[1] The CAGR and Sharpe ratio results are first reported followed by the results for the Benchmark Excess Ratios (BER) and the mean difference tests are reported here. The first goal is to find whether incorporating excess return over the benchmark in the optimization process would improve the performance of the optimal portfolios. The second goal is to analyze whether the optimal portfolio from filtered covariance can produce far superior results than the optimal portfolio from the raw covariance. Table 8.1 shows the average CAGR and average Sharpe ratio across all 5 random portfolios and the full sample portfolio.

Table 8.1

					Minimum Risk + Excess Return with Short Sale Allow			
		Average CAGR%				**Average Sharpe**		
	NASDAQ	EQW	Raw	Filter	NASDAQ	EQW	Raw	Filter
Full Period	9.88	**15.81**	6.55	4.68	0.107	**0.152**	0.086	0.065
Period 1	(9.30)	(6.71)	(6.08)	**(4.47)**	(0.053)	(0.017)	(0.024)	**0.006**
Period 2	27.21	**37.51**	3.99	(7.51)	0.238	**0.298**	0.057	(0.047)
Period 3	11.85	16.82	21.83	**26.17**	0.136	0.181	0.257	**0.275**
	S&P 500	EQW	Raw	Filter	S&P 500	EQW	Raw	Filter
Full Period	4.04	**9.13**	2.09	2.38	0.059	**0.097**	0.042	0.047
Period 1	(13.10)	(9.38)	(0.89)	**(0.88)**	(0.125)	(0.059)	0.011	**0.017**
Period 2	16.00	**26.74**	(8.14)	(10.18)	0.150	**0.198**	(0.059)	(0.060)
Period 3	9.31	10.17	15.39	**18.29**	0.121	0.120	0.233	**0.266**

[1] Under no short sale restrictions, some results are degenerate when excess return constraint is added.

Table 6.3

	MVP with Short Sale Allow							
		Average CAGR%				**Average Sharpe**		
	NASDAQ	EQW	Raw	Filter	NASDAQ	EQW	Raw	Filter
Full Period	9.88	**15.81**	7.92	6.47	0.107	**0.152**	0.101	0.082
Period 1	(9.30)	(6.71)	(5.87)	**(3.33)**	(0.053)	(0.017)	(0.021)	**0.015**
Period 2	27.21	**37.51**	7.22	(4.32)	0.238	**0.298**	0.087	(0.019)
Period 3	11.85	16.82	22.51	**27.20**	0.136	0.181	0.265	**0.288**

	S&P 500	EQW	Raw	Filter	S&P 500	EQW	Raw	Filter
Full Period	4.04	**9.13**	4.32	4.77	0.059	**0.097**	0.066	0.069
Period 1	(13.10)	(9.38)	**(0.03)**	(0.21)	(0.125)	(0.059)	0.018	**0.021**
Period 2	16.00	**26.74**	(3.39)	(4.32)	0.150	**0.198**	(0.017)	(0.099)
Period 3	9.31	10.17	16.43	**18.92**	0.121	0.120	0.249	**0.278**

Table 6.3 is included here as a way to compare whether adding an excess return constraint would improve performance. In fact, the opposite is true. The CAGR values in Table 8.1 are lower than the CAGR values in Table 6.3 for both Raw and Filter portfolios and for all sample periods. In every case, the MVP performs better than the constrained Minimal Risk portfolios. The latter even underperform the benchmarks as displayed in Figures 8.1 and 8.2. Again, the naïve strategy has the best performance.

Figure 8.1 NASDAQ Riskier Portfolio – Short Sale Allow

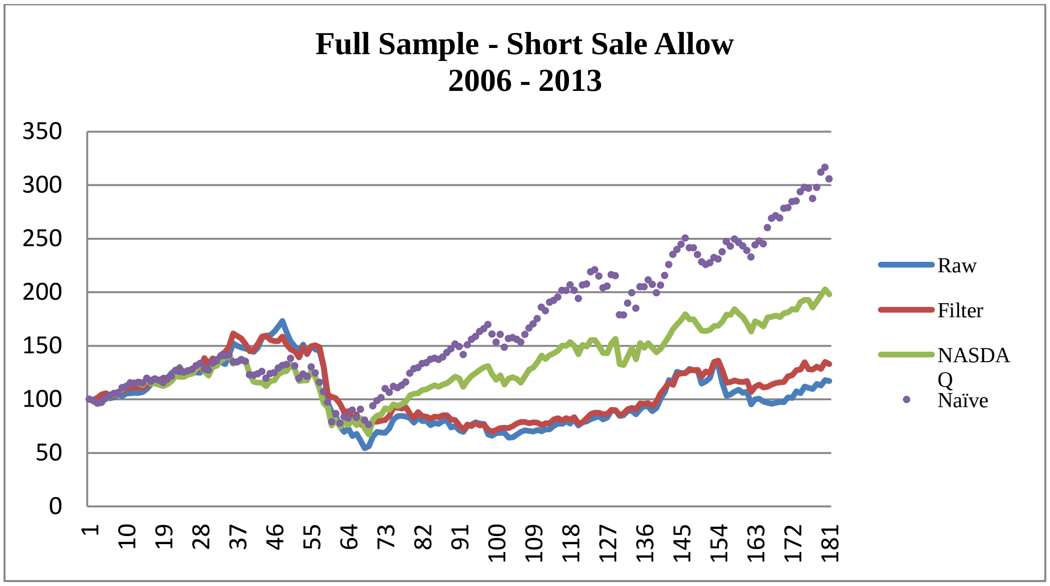


Figure 8.2 S&P Riskier Portfolio – Short Sale Allow

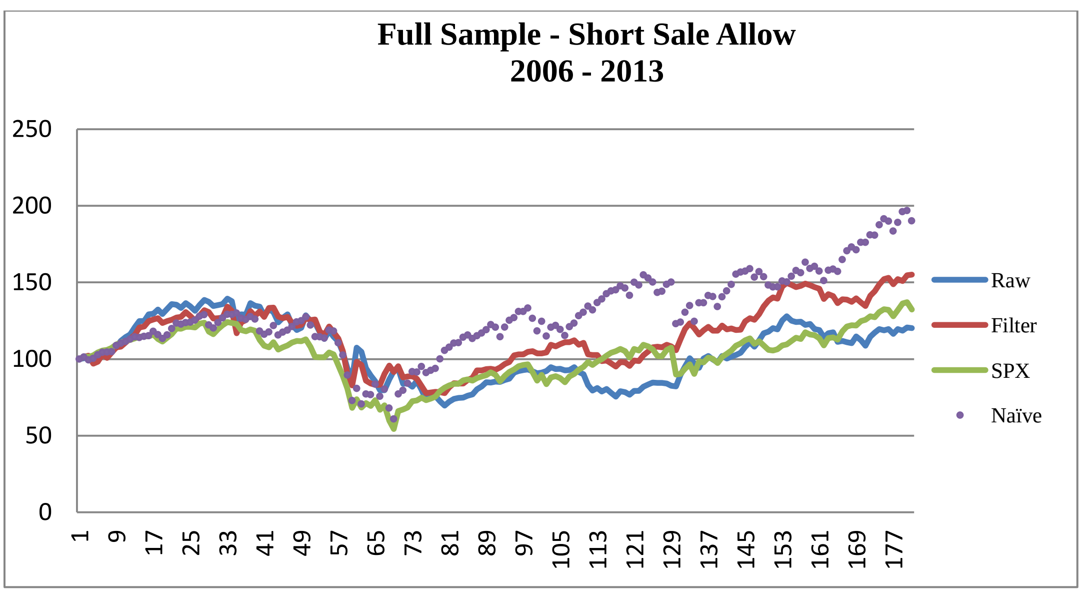


Table 8.2 shows the Benchmark Excess Ratio when excess return constraint is added to the optimization.

Table 8.2 Benchmark Excess Ratio with Excess Return Constraint

Benchmark Excess Ratio

2006 to 2013

	Short Sale Allow		
NASDAQ	Raw	Filter	Naïve
All stocks	-4.74%	-4.59%	19.38%
Port A	-2.52%	-1.93%	6.51%
Port B	-10.23%	-12.07%	15.05%
Port C	0.62%	0.67%	10.78%
Port D	-0.11%	-0.39%	22.31%
S&P			
All stocks	-1.21%	1.60%	19.88%
Port A	-3.01%	-2.37%	12.77%
Port B	1.75%	3.15%	13.51%
Port C	-7.43%	-5.71%	13.86%
Port D	-1.70%	-2.52%	17.76%

One can see that both Raw and Filter portfolios underperform the benchmarks as well as the naïve strategy. In contrast, the Raw and Filter MVPs perform better than the Raw and Filter portfolios here. The mean and standard deviation of excess returns are shown in Table 8.3 and the t-scores in Table 8.4

Table 8.3 Benchmark Excess Return with Excess Return Constraint

Benchmark Excess Return

2006 to 2013

Short Sale Allow

	Raw		Filtered	
NASDAQ	Mean	Std Dev	Mean	Std Dev
All stocks	-0.00264	0.05578	-0.00240	0.05219
Port A	-0.00107	0.04228	-0.00106	0.05523
Port B	-0.00352	0.03440	-0.00618	0.05117
Port C	0.00024	0.03814	0.00039	0.05735
Port D	-0.00004	0.03919	-0.00021	0.05400

S&P	Mean	Std Dev	Mean	Std Dev
All stocks	-0.00060	0.04955	0.00071	0.04413
Port A	-0.00110	0.03639	-0.00114	0.04806
Port B	0.00061	0.03476	0.00134	0.04263
Port C	-0.00274	0.03682	-0.00270	0.04739
Port D	-0.00068	0.04010	-0.00130	0.05167

Table 8.4 Testing Mean Excess Return with Excess Return Constraint.

Testing Mean Excess Return

Short Sale Allow

NASDAQ	t-score	SE	df	1-tail p-value
All stocks	-0.043	0.00569	356	0.483
Port A	-0.001	0.00518	335	0.500
Port B	0.579	0.00460	313	0.282
Port C	-0.029	0.00513	311	0.488
Port D	0.034	0.00497	327	0.486

S&P	t-score	SE	df	1-tail p-value
All stocks	-0.264	0.00495	336	0.396
Port A	0.010	0.00449	342	0.496

Port B	-0.179	0.00410	319	0.429
Port C	-0.007	0.00447	326	0.497
Port D	0.127	0.00487	331	0.450

Note: 1-tail p-value = P(T > t-score) or P(T ≤ -t-score)

From Table 8.4, the lowest 1-tail p-value is 0.282 which far exceeds the 0.05 significance level. Therefore, one can safely conclude that the Filtered covariance has no advantage over the Raw covariance in terms of forming better efficient portfolio.

In this section, the objective function in equation (40) is replaced with the objective function of minimizing the tracking error

$$\mathrm{Min}_{w_1..w_N}\, Var\left(R_p - R_B\right) = \sum_{i=1}^{N}\sum_{j=1}^{N} w_i\, w_j \sigma_{ij} - 2\sum_{i=1}^{N} w_i \sigma_{iB} + \sigma_B^2 \qquad (42)$$

subject to

$$\sum_{i=1}^{N} w_i \bar{R}_i \geq \bar{R}_{Benchmark}$$

$$\sum_{i=1}^{N} w_i = 1$$

The short sale constraint is removed from the optimization problem. The excess return constraint is imposed here. As in the previous section, the goal is to find if this will lead to better performance. Table 8.5 summarizes the CAGR results.

Table 8.5

	Minimum Tracking Error + Excess Return with Short Sale Allow							
	Average CAGR%				Average Sharpe			
	NASDAQ	EQW	Raw	Filter	NASDAQ	EQW	Raw	Filter
Full Period	9.88	**15.81**	12.25	12.26	0.107	**0.152**	0.128	0.128
Period 1	(9.30)	**(6.71)**	(7.24)	(7.53)	(0.053)	**(0.017)**	(0.040)	(0.041)
Period 2	27.21	**37.51**	30.18	30.36	0.238	**0.298**	0.264	0.265
Period 3	11.85	**16.82**	13.93	14.10	0.136	**0.181**	0.155	0.156
	S&P 500	EQW	Raw	Filter	S&P 500	EQW	Raw	Filter
Full Period	4.04	**9.13**	3.82	3.68	0.059	**0.097**	0.058	0.056
Period 1	(13.10)	**(9.38)**	(11.36)	(11.80)	(0.125)	**(0.059)**	(0.080)	(0.085)
Period 2	16.00	**26.74**	14.01	13.83	0.150	**0.198**	0.134	0.132
Period 3	9.31	**10.17**	8.90	9.08	**0.121**	0.120	0.114	0.116

The immediate result from Table 8.5 is that both Raw and Filter portfolios perform far better in the NASDAQ sample than in the S&P sample. But the results are not enough to outperform the naïve strategy. However, the results here are better than those contain in Table 8.1.

Table 7.2

	MTP Performance with Short Sale Allow							
	Average CAGR%				Average Sharpe			

	NASDAQ	EQW	Raw	Filter	NASDAQ	EQW	Raw	Filter
Full Period	9.88	**15.81**	12.23	12.29	0.107	**0.152**	0.128	0.128
Period 1	(9.30)	**(6.71)**	(7.27)	(7.39)	(0.053)	**(0.033)**	(0.041)	(0.040)
Period 2	27.21	**37.51**	30.29	30.46	0.238	**0.298**	0.265	0.266
Period 3	11.85	**16.82**	13.81	13.96	0.136	**0.181**	0.154	0.155

	S&P 500	EQW	Raw	Filter	S&P 500	EQW	Raw	Filter
Full Period	4.04	**8.71**	3.89	3.76	0.059	**0.097**	0.058	0.057
Period 1	(13.10)	**(8.69)**	(11.71)	(12.31)	(0.125)	**(0.059)**	(0.084)	(0.090)
Period 2	16.00	**25.33**	14.38	14.38	0.150	**0.198**	0.137	0.136
Period 3	9.31	**9.60**	9.09	9.29	**0.121**	0.120	0.115	0.117

Comparing results in Table 8.5 to the results in Table 7.2 (MTP without excess return constraint in Chapter 7) both Raw and Filter portfolios show almost identical results for the CAGR and Sharpe ratios. This means the excess return over the benchmark is somehow non-binding when the problem is to minimize the tracking error of a benchmark. Therefore, the conclusions reached for the MTP in Chapter 7 are applicable here as well. One does not expect the Filter portfolio's performance to be any different from the raw portfolio's performance.

Chapter 9

Three-Fund Separation

It becomes quite clear that selecting a single point in the efficient frontier with and without constraints may not even beat the naïve portfolio. It also appears that estimation error in the covariance matrix may be a second-order effect, impacting mostly on the minimum variance portfolio. The question is whether there is an alternative way to enhance the out-of-sample performances of MV portfolios. In this chapter, I explore the notion of three-fund separation as presented by Kan and Zhou (2007).

In portfolio theory, the two-fund separation states that a mean-variance optimizing investor should invest only in the riskless asset and the tangency portfolio. According to Kan and Zhou (2007),

> "If the true parameters are known, as assumed in theory, then two-fund Separation holds and there is no point in analyzing a three-fund portfolio. However, when the parameters are unknown, the tangency portfolio is obtained with estimation error. Intuitively, additional portfolios could be useful if they provide diversification of estimation risk. Indeed, we show that the optimal portfolio weights can be solved analytically in a three-fund universe that consists of the riskless asset, the sample tangency portfolio, and the sample global minimum-variance portfolio. Therefore, a three-fund portfolio rule can dominate all the previous two-fund rules."

The authors consider estimation error in both sample mean and sample covariance and derived analytical solutions for the optimal weights for the two-fund and three-fund separation. Denoting R_t as the vector of excess returns, $R_t = \left(r_t - r_{ft}\mathbf{1}_{\boldsymbol{N}}\right)$, the sample mean of the excess return as $\hat{\mu}$ and the sample covariance matrix as $\sum$. Under normality assumption, it is well-known that $\hat{\mu}$ and $\widehat{\sum}$ are independent of each other and they have the following exact distribution

$$\hat{\mu} = N(\mu, {}^{\sum}/_{T}) \tag{43}$$

$$\widehat{\sum} = W_N(T-1, \sum)/\mathrm{T} \tag{44}$$

where $W_N(T-1, \sum)$ denotes a Wishart distribution with T-1 degrees of freedom and covariance matrix $\sum$. Since the expectation of $E\left[\widehat{\sum}^{-1}\right] = T\sum^{-1}/(T-N-2)$ (see Muirhead (1982)) then the optimal portfolio weight $\widehat{w}$ as given by the solution

$$\widehat{w} = \frac{1}{\gamma}\widehat{\sum}^{-1}\hat{\mu} \tag{45}$$

is actually a biased estimator of the true portfolio weight w^* since

$$E[\widehat{w}] = \frac{T}{T-N-2}w^* \tag{46}$$

when T > N+2. This implies that $|\widehat{w}_i| > |w^*|$, so investors, whose decision in (45) based on the sample estimates, tend to take bigger positions in the risky assets than those who know the true parameters.

In the three-fund separation, Kan and Zhou (2007) chose a linear combination of the global MVP and a sample tangency portfolio. The reasons are twofold. First, the weights of the global MVP depend only on $\widehat{\Sigma}$ but not $\hat{\mu}$. Second, every sample frontier portfolio is a linear combination of two distinct sample frontier portfolios. It is named three-fund because the returns are measured as excess return over the risk-free rate. Therefore, the solution in (45) is really a two-fund separation. Adding the MVP as the third portfolio becomes a three-fund separation.

Suffice it to say, the analytical results of the weights for the three-fund solution is given as:[1]

$$\widehat{w}(c,d) = \frac{1}{\gamma}(c\widehat{\Sigma}^{-1}\hat{\mu} + d\widehat{\Sigma}^{-1}\mathbf{1}_{\boldsymbol{N}}) \tag{47}$$

where c and d are chosen optimally and they are derived as:

$$c^{**} = k\,(\frac{\varphi^2}{\varphi^2 + {}^{N}/_{T}}) \tag{48}$$

$$d^{**} = k\,(\frac{{}^{N}/_{T}}{\varphi^2 + {}^{N}/_{T}}) \tag{49}$$

[1] The optimal solution is derived by maximizing $E[\widetilde{U}(\widehat{w}(c,d))]$ with respect to c and d.

where k is a constant and $c^{**} + d^{**} = 1$. The squared slope of the asymptote to the ex ante minimum-variance frontier is given as:

$$\varphi^2 = \left(\mu - \mu_g \mathbf{1_N}\right)' \Sigma^{-1}(\mu - \mu_g \mathbf{1_N}) \tag{50}$$

and the expected excess return of the ex ante global MVP is given as:

$$\mu_g = (\mathbf{1}_N' \Sigma^{-1} \mu)/(\mathbf{1}_N' \Sigma^{-1} \mathbf{1_N}) \tag{51}$$

Thus, the vector of optimal weights for (47) are expressed as

$$\widehat{w}^{**} = \frac{1}{\gamma}\ (\boldsymbol{c}^{**}\ \widehat{\Sigma}^{-1}\hat{\mu} + \ \boldsymbol{d}^{**}\hat{\mu}_g\ \widehat{\Sigma}^{-1}\mathbf{1_N}) \tag{52}$$

Intuitively, the weight c** applies to the tangency portfolio (or mean-variance portfolio) while d** applies to MVP. Using sample means and covariance to estimate $\hat{\varphi}^2$ will result in the following distributional property:

$$\frac{(T-N+1)\hat{\varphi}^2}{N-1} \sim F_{N-1,\ T-N+1}(T\varphi^2) \tag{53}$$

where $F_{N-1,T-N+1}(T\varphi^2)$ is a F distribution with N-1 and T-N+1 degrees of freedom, and a noncentrality parameter of $T\varphi^2$. Since $\hat{\varphi}^2$ is a heavily biased estimator when T is small, the unbiased estimator is thus given by:

$$\hat{\varphi}_u^2 = \frac{(T-N-1)\hat{\varphi}^2 - (N-1)}{T} \tag{54}$$

As long as $\hat{\varphi}_u^2 > 0$, it is not necessary to include a second term in (54) for adjusting the estimator as suggested by the authors.

While Kan and Zhou (2007) derived these elegant optimal solutions, they only test this on simulated data and, not surprisingly, found that their three-fund separation portfolio improved expected out-of-sample performance over the two-fund portfolio and other portfolios based on Bayesian-Stein estimators. Unfortunately, they did not include performance test on real market data. Therefore, the remaining portion of this chapter is dedicated to provide a partial test of their analytical results.

Although equation (52) calls for solving the optimal weights of the portfolio, one can also interpret the equation as a weighted average between two ex ante portfolios with the weights being defined by c** and d**. Although my studies did not measure returns as excess return over the risk-free asset and, hence, is strictly not a three-fund separation, it would be interesting to implement the results in (52) as a “Modified 2-Fund separation” (M2F). That is, the risk-free asset is not considered here, but is implicitly replaced by a cash position with $r_{ft} = 0$. The next step is to identify the tangency portfolio. One possibility is to simply use the benchmark index such as NASDAQ 100 for the NSADAQ sample and the S&P 500 Index for the S&P sample.

Another possibility is to use the equally weighted portfolio in place of the tangency portfolio. Although the naïve portfolio may not be on the efficient frontier, its performance so far has beaten these other efficient portfolios. In fact, a similar hybrid model that combines the naïve portfolio and the MVP was suggested and studied by DeMiguel et. al (2009). The results of their hybrid portfolio were promising, and the weight is given as:

$$\widehat{w}^{ew-mvp} = \left(a\,\frac{1}{N}\mathbf{1}_{\boldsymbol{N}} + b\,\widehat{\sum}^{-1}\mathbf{1}_{\boldsymbol{N}}\right) \qquad (55)$$

where a and b are solved similarly to c^{**} and d^{**}.

Note that both c^{**} and d^{**} in (48) and (49) are positive numbers implying only positive weights in constructing the M2F. The M2F lies between the MVP and the tangency portfolio. On closer examination of (52), one can see the extra variable μ_g in $(\boldsymbol{d}^{**}\hat{\mu}_g\,\widehat{\sum}^{-1}\mathbf{1}_{\boldsymbol{N}})$. It is possible that the estimate $\hat{\mu}_g$ can be 0, positive or negative. When $\hat{\mu}_g = 0$ then 100% of the fund is invested in the tangency portfolio. When $\hat{\mu}_g < 0$ then one should be shorting the MVP by -d^{**} and c^{**} will be recomputed as (1 + d**). There will be a leverage effect in this case. If the MVP originally consisted of only long positions in stocks, then all these positions are reversed to short positions. If short selling restriction is enforced, then the absolute value of d** is used throughout the test plus employing a long only MVP. The latter case is of great interest here because it is harder to perform short selling, and increasing leverage can present unnecessary margin calls during adverse market conditions.

Using the NASDAQ sample of 87 stocks as a starting point, I first compute $\hat{\mu}_g$ for the local MVP using equation (51). Next, I compute φ^2 using equation (50), the unbiased estimate $\hat{\varphi}_u^2$ in (54) and d** according to (49) using the unbiased estimate. The d** is then applied to the realized returns of the MVP while c** = (1 – d**) is applied to the realized returns of the tangency portfolio, which is either the benchmark index or the naïve portfolio. Two cases are considered here. First, the long-short case when d** can be negative and c** is levered up. Second, the long only case when |d**| > 0 is only allowed. The average CAGR% for the entire sample period 2006 to 2013 are presented in Table 9.1

Table 9.1 Performance of M2F (NASDAQ benchmark)

	Average CAGR%			
	NASDAQ	Naive	MVP(raw)	MVP(filter)
2006 to 2013	9.88	16.19	13.15	15.43
Long-Short				
MVP(raw) + NASDAQ	9.74			
MVP(filter) + NASDAQ	10.04			
MVP(raw) + Naïve		16.13		
MVP(filter) + Naïve		**16.33**		
Long Only				
MVP(raw) + NASDAQ	**10.41**			
MVP(filter) + NASDAQ	10.35			
MVP(raw) + Naïve		15.39		
MVP(filter) + Naïve		16.08		

In the Long-Short case, the M2F strategy underperforms both NASDAQ 100 Index and the naïve portfolio using the MVP $_{\text{Raw}}$. On the other hand, using the MVP $_{\text{Filter}}$ as the anchor portfolio, the M2F strategy actually beats both NASDAQ 100 and the naïve portfolio by around 15 basis points a year. Figure 9.1 and 9.2 illustrate the full sample period performances for the Long-Short case. As for the "Long Only" case, the M2F strategies outperform the NASDAQ 100 Index by 47 basis points per year using MVP $_{Filter}$ and 53 basis points per year using MVP $_{Raw}$. But the combined strategies underperform the standalone results of each MVP.

Figure 9.1

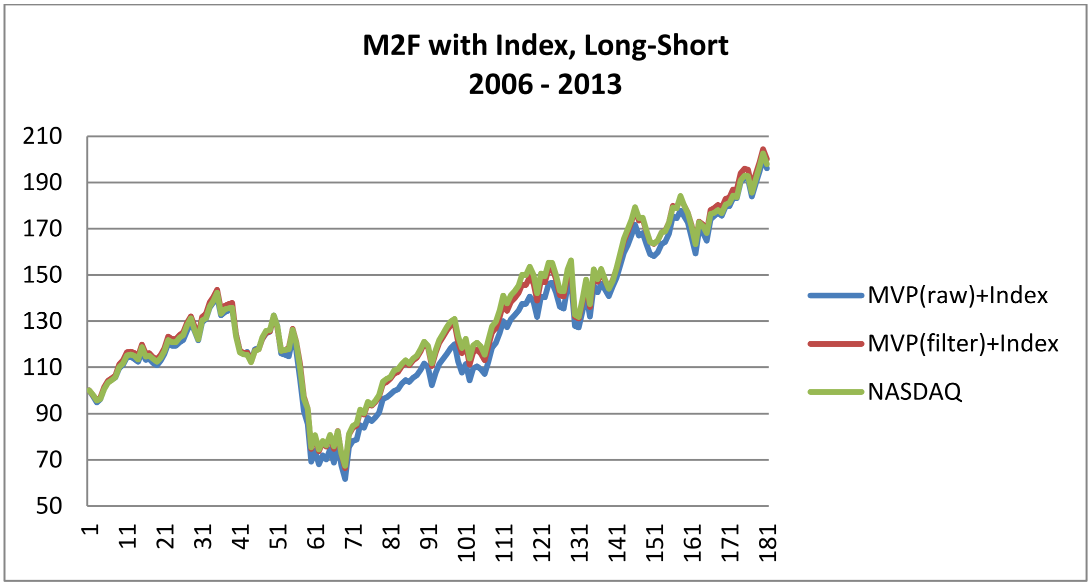
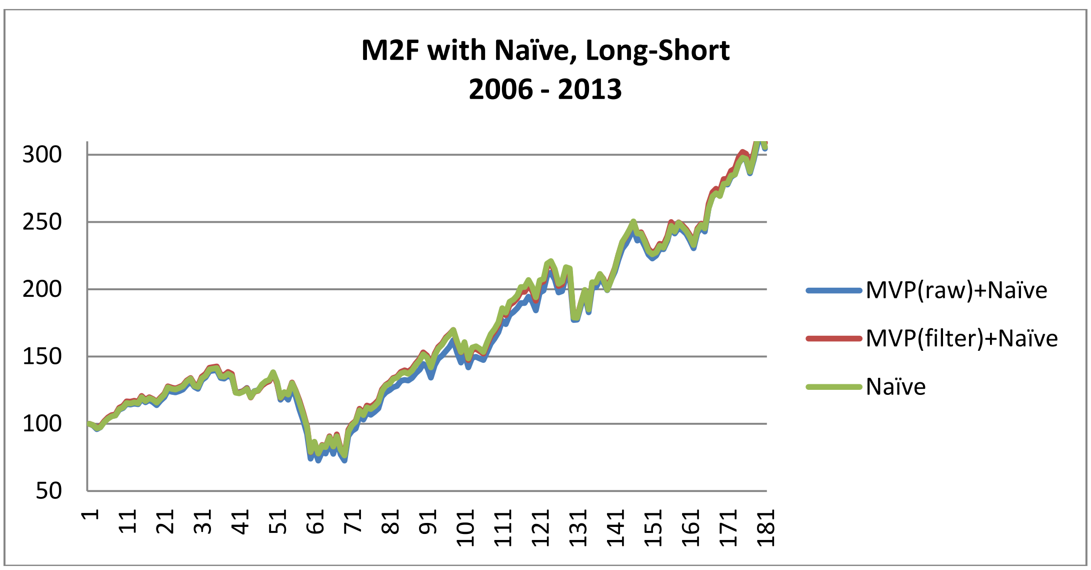


Figure 9.2



Using the naïve portfolio as the tangency portfolio, the M2F strategies actually underperform the naïve portfolio in both "Long-Short" and "Long Only" cases. However, the M2F strategies in this case are better than the standalone MVP performances.

The final stage of this chapter is to repeat the same test for the S&P sample. As before, both "Long-Short" and "Long Only" cases are examined. Table 9.2 presents the results.

Table 9.2 **Performance of M2F (S&P benchmark)**

	Average CAGR%			
	S&P	EQW	MVP(raw)	MVP(filter)
2006 to 2013	**4.04**	**9.30**	3.40	3.85
Long-Short				
MVP(raw) + S&P	3.91			
MVP(filter) + S&P	3.76			
MVP(raw) + Naïve		9.25		
MVP(filter) + Naïve		8.96		
Long Only				
MVP(raw) + S&P	3.90			
MVP(filter) + S&P	3.97			
MVP(raw) + Naïve		8.25		
MVP(filter) + Naïve		8.82		

For the S&P sample, there is no improvement for the M2F strategies over the S&P 500 Index. The main reason is that both MVP $_{Raw}$ and MVP $_{Filter}$ are underperforming the benchmark index. However, by combining the MVP with the naïve portfolio, the M2F strategies produce better results than the standalone result of each MVP. Unfortunately, none of them were able to beat the standalone Naïve portfolio's performance. The closest M2F strategy is the MVP $_{Raw}$ and Naïve combination, under Long-Short scenario, where it underperformed by 5 basis points per year.

Basically, my findings are supporting DeMiguel et.(2009) results that complex models, such as Neutron that employs RMT and eigenvalue modification, cannot beat the naïve (or 1/N) strategy. They look at fourteen models. These include sample based mean-variance models, Bayesian-Stein models, Moment restriction models such as MVP, Portfolio constraint models, three-fund separation and the hybrid model of MVP and Equal Weight strategy. Their studies did not include models using RMT predictions. Therefore, my results here supplement to their conclusion:

> "We find that out-of-sample Sharpe ratio of the sample-based mean-variance strategy is much lower than that of the 1/N strategy, indicating that the errors in estimating means and covariances erode all the gains from optimal, relative to naïve, diversification…
> In summary, we find that of the various optimizing models in the literature, there is no single model that consistently delivers a Sharpe ratio or a certainty equivalent (CEQ) return that is higher than that of the 1/N portfolio, which also has a very low turnover."[2]

The Battle of Equals

Finally, one question still remains. Why does naïve strategy on random portfolios can easily outperform the benchmark indexes such as the NASDAQ 100 and the S&P 500. These benchmarks are value weighted index. Table 9.3 shows these results.

[2] Page 1947 DeMiguel et. al (2009)

Table 9.3

Performance of Naïve Portfolios
2006 to 2013

	Ending Value	CAGR		Ending Value	CAGR
NASDAQ 100 Index	197.915	9.88%	S&P 500 Index	132.267	4.04%
Port A	244.114	12.92%	Port A	196.664	9.79%
Port B	295.712	15.71%	Port B	176.597	8.23%
Port C	277.260	14.77%	Port C	172.258	7.87%
Port D	382.815	19.46%	Port D	205.953	10.46%
Full Sample	305.723	16.19%	Full Sample	190.213	9.30%

The starting value for each portfolio was 100 and the ending value is the cumulative compounded return at the end of the testing period. As mentioned previously, CAGR is the compounded annual growth rate. Notice that each of the portfolios A to D was randomly selected at the outset. Surprisingly, they all beat their respective benchmark index. Portfolio D in both cases has the best performance. This is purely by chance and not by design. One may argue that the market capitalization or size may attribute to differential performance. This is captured in Tables 9.4 and 9.5 below:

Table 9.4 Market Cap at three different dates for NASDAQ sample

NASDAQ 100	1/3/2006	3/9/2009	8/30/2013	
Market Cap	2,105,697	1,253,810	3,571,964	# stocks
Port A	859,488	439,893	964,593	20
Port B	533,836	346,729	1,202,394	20
Port C	424,421	255,731	580,844	20
Port D	244,713	167,104	666,478	27
Full Sample	2,062,458	1,209,457	3,414,309	87

	% of Index		
Port A	40.8%	35.1%	27.0%
Port B	25.4%	27.7%	33.7%
Port C	20.2%	20.4%	16.3%
Port D	11.6%	13.3%	18.7%
Full Sample	97.9%	96.5%	95.6%
	Average Market Cap		
Port A	42,974	21,995	48,230
Port B	26,692	17,336	60,120
Port C	21,221	12,787	29,042
Port D	9,063	6,189	24,684
Full Sample	23,706	13,902	39,245

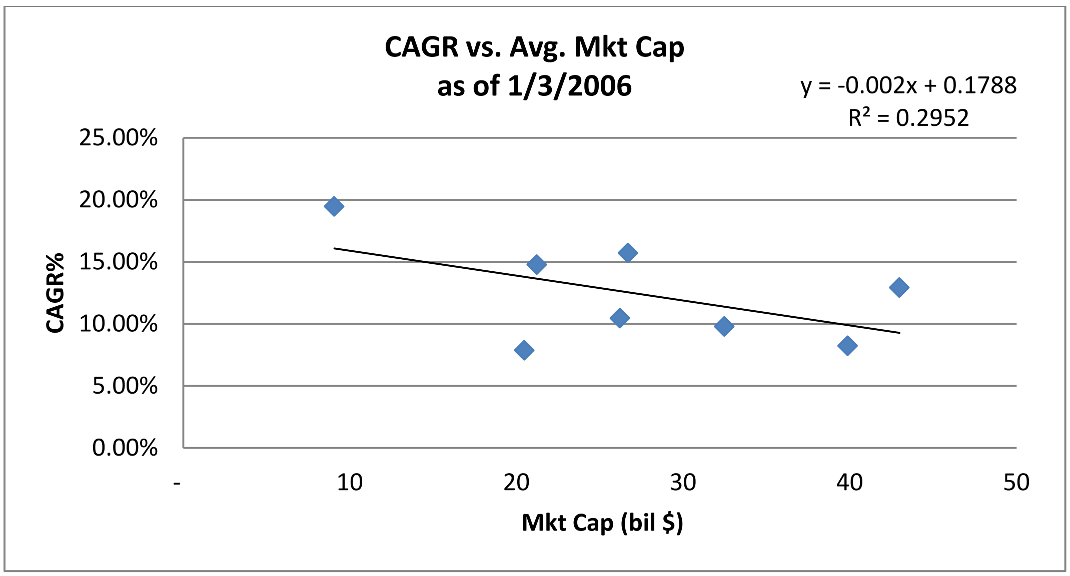
Notice that Portfolio D has the lowest average market cap at the beginning of the sample period. It remains the lowest market cap at the bottom of the crisis.

Table 9.5 Market Cap at three different dates for S&P sample

S&P 500	1/3/2006	3/9/2009	8/30/2013	
Market Cap	11,719,294	6,106,809	14,966,532	# stocks
Port A	649,416	337,639	800,741	20
Port B	797,432	611,965	877,589	20
Port C	409,394	235,128	559,741	20
Port D	524,169	253,022	532,881	20
Full Sample	2,380,411	1,437,755	2,770,952	80
	% of Index			
Port A	5.5%	5.5%	5.4%	
Port B	6.8%	10.0%	5.9%	
Port C	3.5%	3.9%	3.7%	
Port D	4.5%	4.1%	3.6%	
Full Sample	20.3%	23.5%	18.5%	
	Average Market Cap			
Port A	32,471	16,882	40,037	
Port B	39,872	30,598	43,879	
Port C	20,470	11,756	27,987	
Port D	26,208	12,651	26,644	
Full Sample	29,755	17,972	34,637	

For the S&P sample, portfolios C and D have the lowest average market cap. These stocks are not the typical small cap stocks. Next, I plot the CAGR to the Average Market Cap for each of the 8 random portfolios at the start of the testing period. This is illustrated in Figure 9.3

Figure 9.3



The regression results are shown below:

$$\text{CAGR} = a + b^{*}\text{Avg. Market Cap} \qquad R^2 = 0.295$$

$$\begin{array}{cc} 0.178 & -0.002 \\ (t=4.839) & (t = -1.585) \end{array}$$

The slope coefficient is significant at the 90% t-distribution. Therefore, randomly selected portfolios with smaller market cap tend to outperform the portfolios with larger market cap. As the average market cap approaches zero, the implied average CAGR tends toward 17.8%.

Chapter 10

Resolving the Puzzle

The evidence so far indicates two obvious facts. First, the naïve portfolio clearly outperforms the portfolios generated by the more sophisticated portfolio optimization. This is especially true during periods after the market crash of 2008. Second, the high hopes for better performance using “filtered” covariance matrix is completely shattered. It turns out that one cannot differentiate between the excess return over the benchmark using either raw or filtered covariance. This includes looking at various optimal portfolios including the MVP, the MTP and the minimal risk portfolio with excess return constraint. In several cases, some of these portfolios actually underperform the benchmark index.

In Chapter 3, the characteristics of the variance and correlation structures after the filtering process were analyzed in some details. The finding is that the Neutron filtering application, in general, produces higher correlation values across the board. In other words, the filtered correlation structure is skewed towards higher value than the raw correlation structure. On the other hand, the filtered variance actually declined, on average, across stocks. But the decline in variance is not monotonic in nature. Around half of the sample shows decline in variance while the other half shows some increase in variance as compared to pre-filtering variance.

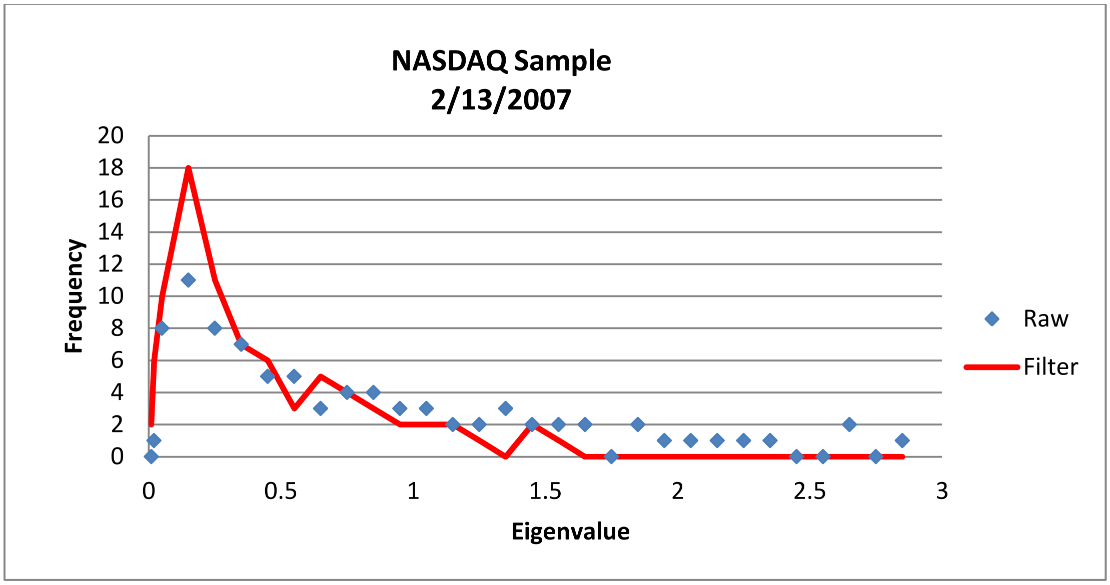
These observations question the usefulness of the Neutron application. If one is only interested in looking at better or enhanced correlation among securities or among variables, then the application appears to provide such a feature – higher correlation structure. Perhaps this may be applicable to pair trading or long-short strategy based on changing correlation structure among securities. If one is interested in getting better risk measures because historical correlations and variances are deemed too low for some reasons, then the Neutron application can be a useful tool.

However, if one is interested in portfolio selection and portfolio performance, then Neutron is not the right tool. So far, there is no evidence to show any advantage of using filtered covariance over the raw sample covariance. There appears to be no new signal or superior information contained in the filtered covariance. The “efficient” portfolios do not automatically translate to better realized performance. The naïve strategy of putting equal weight on each stock (which is always inefficient) proves to be far better than the optimal portfolios based on risk-return tradeoff. Moreover, the naïve strategy beats both S&P 500 and NASDAQ 100 benchmarks.

At this juncture, the focus is to examine the behavior of the final eigenvalues after the filtering process. Recall that the key notion of filtering noise is the method used to modify the eigenvalue distribution within the “noisy” band as predicted by RMT. As mentioned before, some may replace the original eigenvalues by the average eigenvalue within the random band. Others simply flatten them out. It is important to ensure that the trace of the N x N matrix remains intact after the replacement. The trace of an N x N correlation matrix is simply the sum of the diagonal values which is simply N. Although Neutron application is proprietary, one can infer the method used in the application by performing a reverse engineering process. I propose to extract the eigenvalues from the filtered correlation matrix and overlay them with the eigenvalues from the raw correlation matrix. In particular, the frequency distribution between the two sets of eigenvalue is of most interest.

Consistent with the work done in Chapter 3, the correlation matrix for the three dates are analyzed. Each date represents each of the three sub-periods that were examined before. The dates are: 2/13/2007, 3/9/2009 and 1/13/2012 whereby the second date contains the crisis period and exhibits the highest volatility. The algorithm used to extract the eigenvalues is the QR (Orthogonal Upper Triangular decomposition) method (see Press et. al 1988). Figure 10.1 shows the frequency distribution of the eigenvalues within the "random" band for the NASDAQ sample as of 2/13/2007. The larger eigenvalues are not included in the chart. The raw eigenvalues are overlay with the filtered eigenvalues.

Figure 10.1 Frequency Distribution of Eigenvalue for NASDAQ sample, 2/13/2007



The first observation is that the filtered eigenvalues are most heavily concentrated at 0.15. The second observation is that the filtered eigenvalue converges quickly to zero after 1.5 whereas the raw eigenvalues are still positive after that range. Figures 10.2 and 10.3 contain the charts for the other two dates.

Figure 10.2

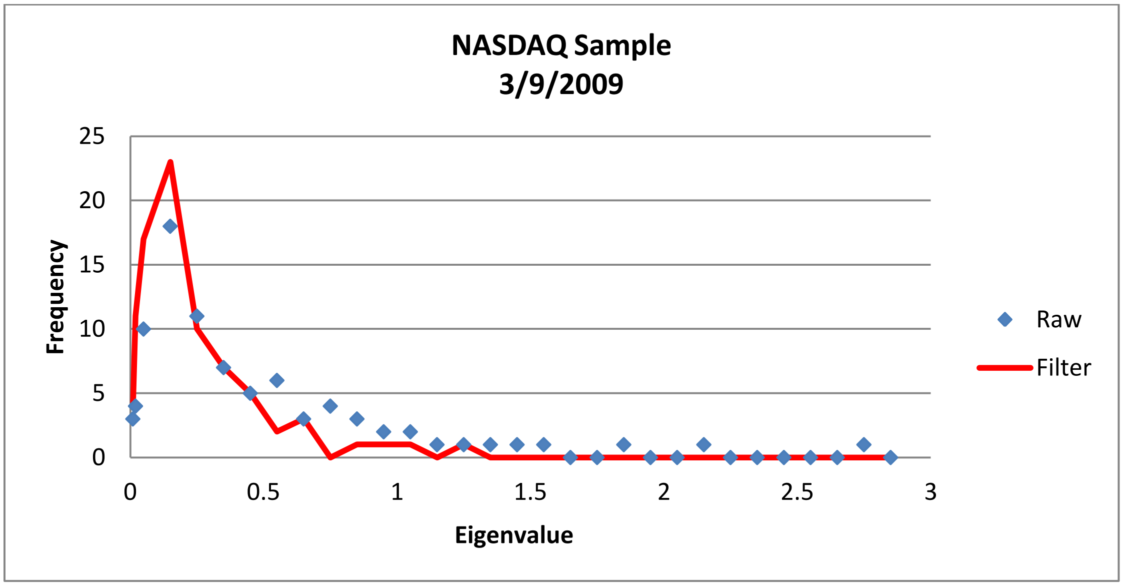
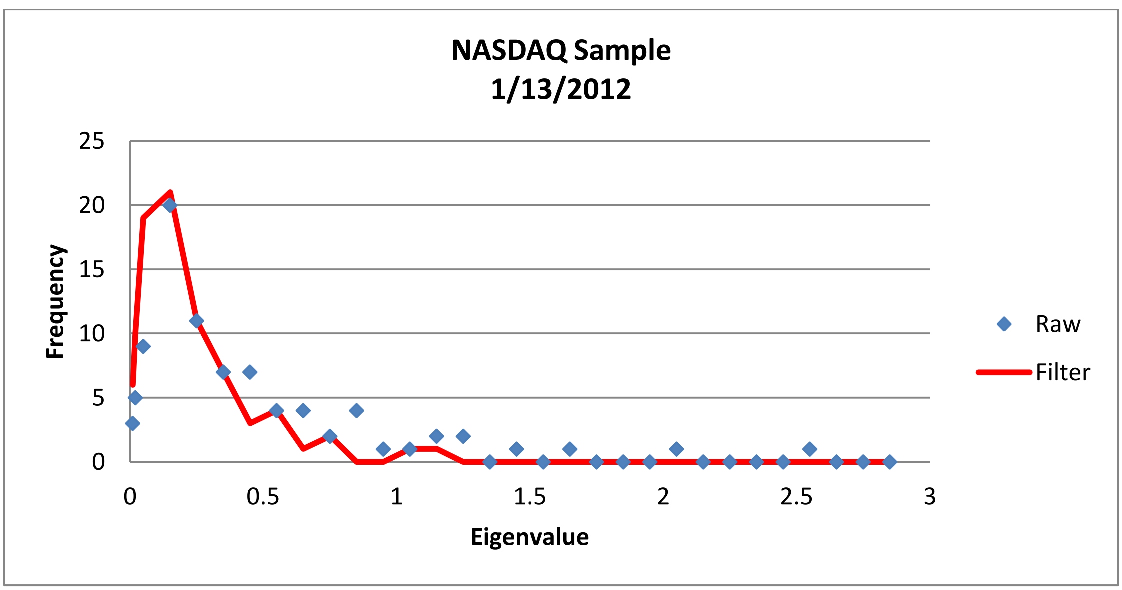


Figure 10.3



The results for all three dates are consistent. There is a high concentration of very low eigenvalues around 0.15 and the filtered eigenvalue converges to zero quicker than the raw eigenvalue. It appears that Neutron "bunches" the smallest eigenvalues at the lower end of the spectrum and narrows the randomness within a tighter band as compared to the raw eigenvalues. Figures 10.4, 10.5 and 10.6 for the S&P sample confirm the same pattern.

Figure 10.4

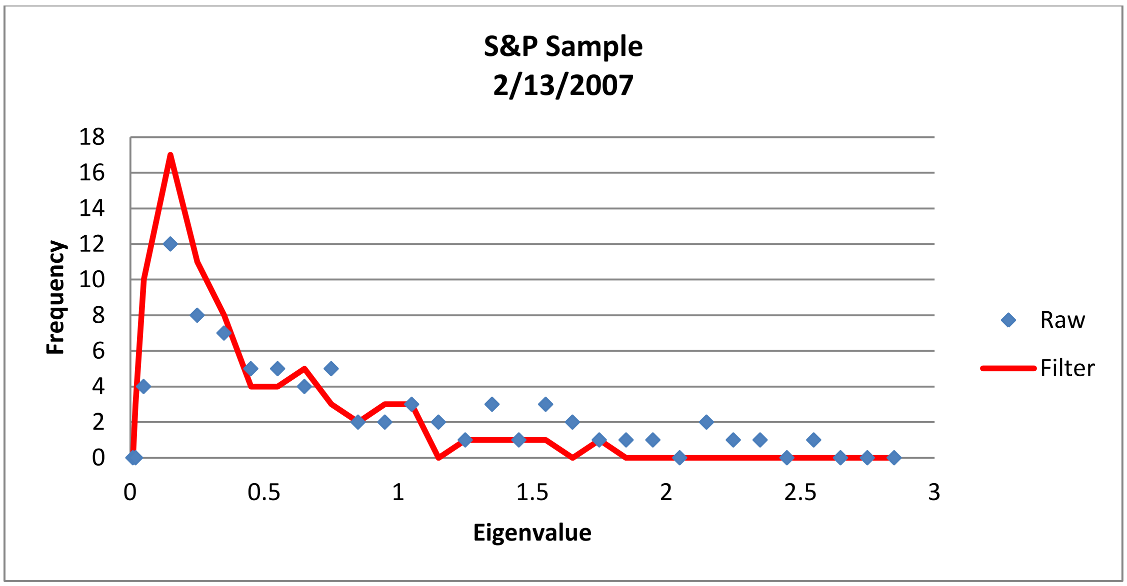


Figure 10.5

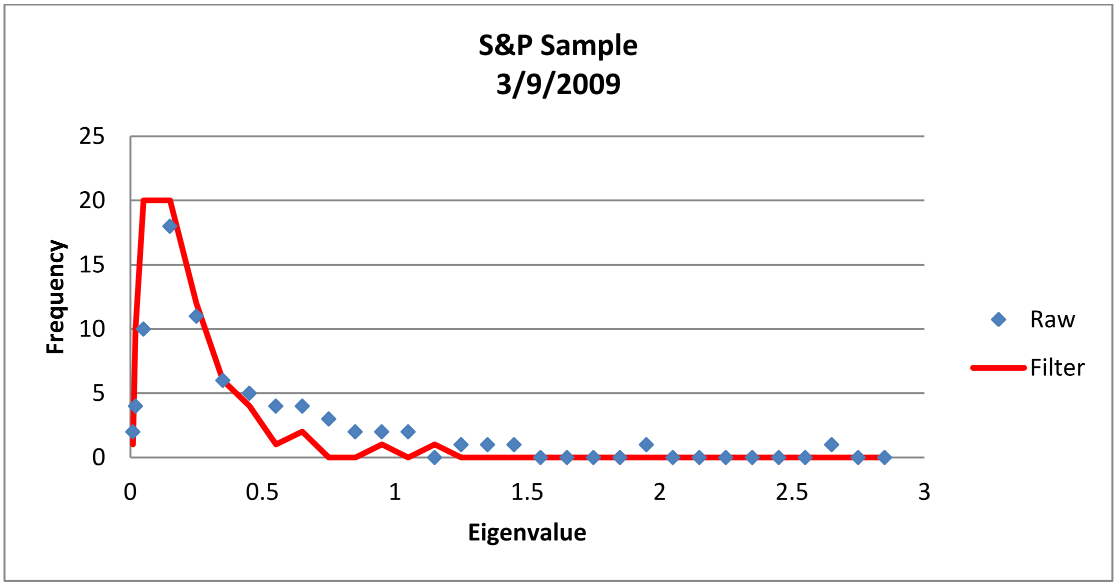
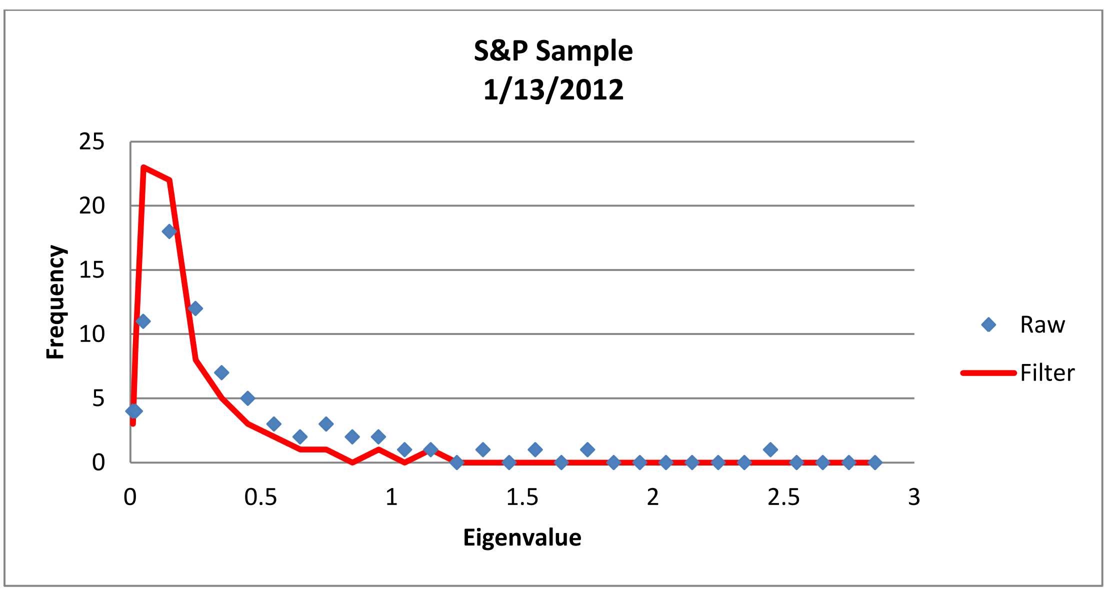


Figure 10.6



The next step is to examine the behavior of the smaller eigenvalues to the larger ones. The lowest 10 eigenvalues are added up and then expressed as a fraction of the trace (N) of the matrix. Then the cumulative sum of the next 10 eigenvalues is recorded and the procedure continues until it reaches 100%. The purpose is to find out how much information the smaller eigenvalue contributes to the total information. The NASDAQ results for the two dates are presented in Table 10.1 and 10.2 and in Figures 10.7 and 10.8

Table 10.1 Cumulative eigenvalues from lowest to highest as of 2/13/2007 (NASDAQ)

	2/13/2007		***NASDAQ***			
		RAW			**FILTER**	
Lowest	**Avg**	**Sum**	**% of Total**	**Avg**	**Sum**	**% of Total**
10	0.0376	0.38	0.43%	0.0154	0.15	0.18%
20	0.0695	1.39	1.60%	0.0283	0.57	0.65%
30	0.1166	3.50	4.02%	0.0469	1.41	1.62%
40	0.1757	7.03	8.08%	0.0714	2.86	3.28%
50	0.2553	12.77	14.67%	0.1056	5.28	6.07%
60	0.3509	21.05	24.20%	0.1499	8.99	10.34%
70	0.4704	32.93	37.85%	0.2110	14.77	16.98%
80	0.6271	50.17	57.66%	0.2930	23.44	26.94%
85	0.7369	62.64	72.00%	0.3553	30.20	34.71%
87	1.0000	87.00	100.00%	1.0000	87.00	100.00%

Table 10.2 Cumulative eigenvalues from lowest to highest as of 3/9/2007 (NASDAQ)

	3/9/2009		***NASDAQ***			
		RAW			**FILTER**	
Lowest	**Avg**	**Sum**	**% of Total**	**Avg**	**Sum**	**% of Total**
10	0.0152	0.15	0.17%	0.0105	0.11	0.12%
20	0.0296	0.59	0.68%	0.0143	0.29	0.33%
30	0.0453	1.36	1.56%	0.0201	0.60	0.69%
40	0.0679	2.72	3.12%	0.0285	1.14	1.31%
50	0.0985	4.93	5.66%	0.0403	2.02	2.32%
60	0.1393	8.36	9.61%	0.0571	3.43	3.94%

70	0.1957	13.70	15.75%	0.0816	5.71	6.57%
80	0.2797	22.38	25.72%	0.1198	9.58	11.02%
85	0.3502	29.77	34.21%	0.1557	13.23	15.21%
87	1.0000	87.00	100.00%	1.0000	87.00	100.00%

Figure 10.7

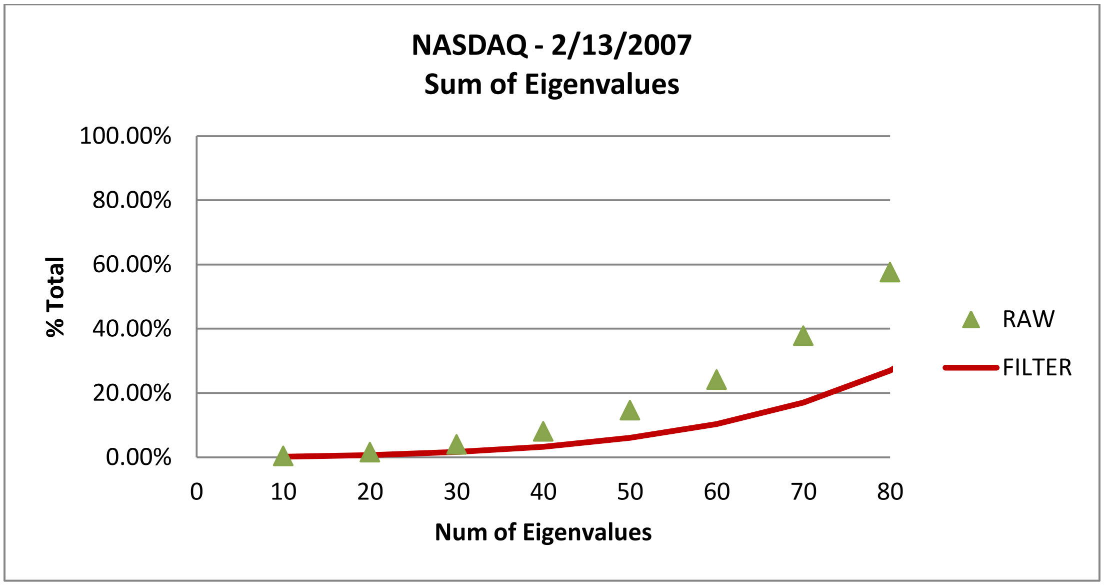
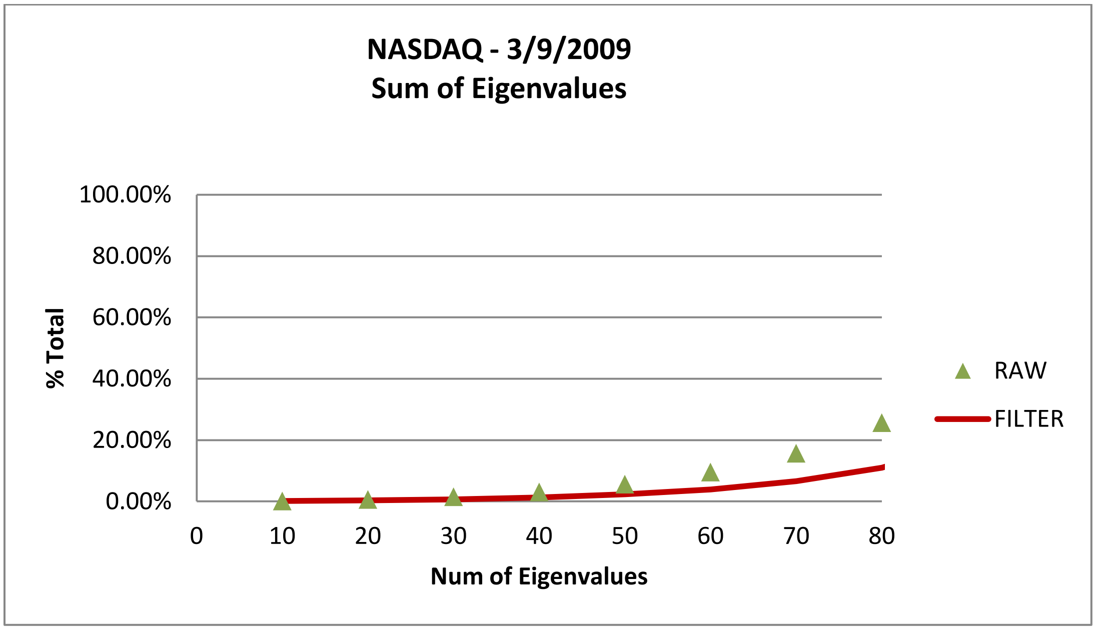


Figure 10.8



There are several interesting facts arising from these results. First, the average filtered eigenvalue is consistently lower than the average raw eigenvalue. On average, the raw eigenvalue is around 2.3 times greater than the filtered eigenvalue. Second, the cumulative sum of the filtered eigenvalues is also smaller than the cumulative sum of the raw eigenvalues. As of 2/13/2007, the raw cumulative sum at the 85th eigenvalue is 62.64 (or 72% of the total) as compared to 30.20 (or 34.71% of the total) for the filtered case. That implies that the filtering process reduces the eigenvalues throughout the entire spectrum except for the largest two eigenvalues. In fact, only 34.7% of the total filtered eigenvalue is explained by the lowest 85 eigenvalues whereas the largest two explain 65.3%. As for the raw case, the largest two eigenvalues account for only 28%.

During the crisis period, the 85 lower eigenvalues explain only 34.21% for the raw case and 15.21% for the filtered case. The largest two eigenvalues explain 65.79% (raw case) and 84.79% (filter case). This means that systemic risk is mainly captured by two eigenvalues. The results for the S&P sample are contained in Tables 10.3 and 10.4 and in Figures 10.9 and 10.10.

Table 10.3 **Cumulative eigenvalues from lowest to highest as of 2/13/2007 (S&P)**

	2/13/2007		***S&P***			
		RAW			**FILTER**	
Lowest	**Avg**	**Sum**	**% of Total**	**Avg**	**Sum**	**% of Total**
10	0.0542	0.38	0.47%	0.0232	0.23	0.29%
20	0.0970	1.94	2.43%	0.0400	0.80	1.00%
30	0.1534	4.60	5.75%	0.0643	1.93	2.41%
40	0.2213	8.85	11.07%	0.0953	3.81	4.77%
50	0.3076	15.38	19.23%	0.1368	6.84	8.55%
60	0.4205	25.23	31.54%	0.1958	11.75	14.69%
70	0.5696	39.87	49.84%	0.2772	19.40	24.26%
75	0.6656	49.92	62.40%	0.3308	24.81	31.01%
80	1.0000	80.00	100.00%	1.0000	80.00	100.00%

Table 10.4 **Cumulative eigenvalues from lowest to highest as of 3/9/2009 (S&P)**

	3/9/2009		***S&P***			
		RAW			**FILTER**	
Lowest	**Avg**	**Sum**	**% of Total**	**Avg**	**Sum**	**% of Total**
10	0.0175	0.18	0.22%	0.0138	0.14	0.17%
20	0.0321	0.64	0.80%	0.0195	0.39	0.49%
30	0.0533	1.60	2.00%	0.0261	0.78	0.98%
40	0.0795	3.18	3.98%	0.0354	1.42	1.77%
50	0.1151	5.76	7.19%	0.0491	2.46	3.07%
60	0.1670	10.02	12.53%	0.0704	4.22	5.28%
70	0.2408	16.86	21.07%	0.1014	7.10	8.87%
75	0.2974	22.31	27.88%	0.1264	9.48	11.85%
80	1.0000	80.00	100.00%	1.0000	80.00	100.00%

Figure 10.9

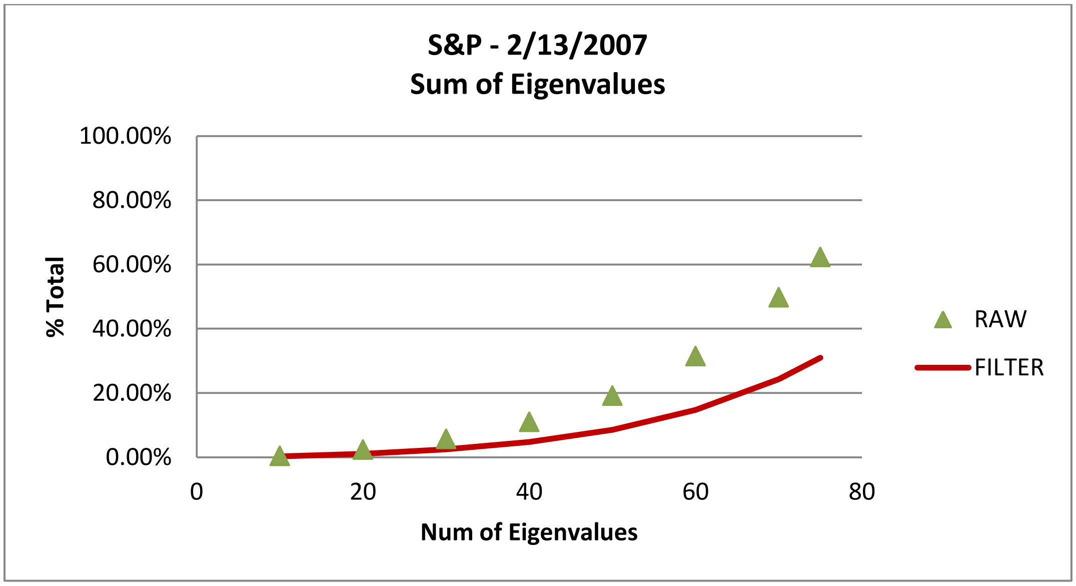
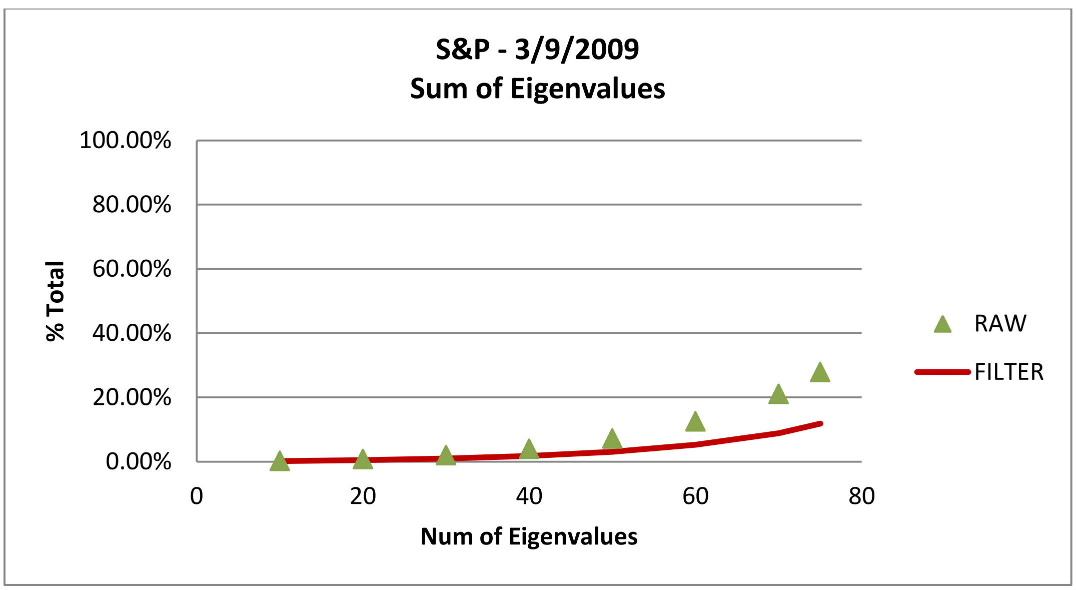


Figure 10.10



The results for the S&P sample are very similar to the results for the NASDAQ sample. It now becomes clear that the Neutron's method is doing three modifications.

First, it is reducing the eigenvalues within the RMT band by half. Second, it limits the RMT band to a narrower range. Finally, it shifts most of the values to the larger eigenvalues. This implies that the filtered correlation matrix is determined highly by a market wide factor and a secondary factor. This could explain why the filtered correlation structure is skewed towards higher values and the filtered variance for most (not all) stocks reduced at the same time.

This analysis may provide an answer as to why the filtered correlation matrix is unable to beat the raw correlation, especially in the unconstrained short selling case. It can also explain why there are fewer stocks in the optimal portfolios using filtered correlation when there is short sale restriction. Consider two stocks called Stock1 and Stock2. Combining the two stocks in the portfolio, one can write the variance of the portfolio as:

$$\sigma_p^2 = (\omega_1^2\sigma_1^2 + \omega_2^2\sigma_2^2) + 2\rho\omega_1\omega_2\sigma_1\sigma_2 \tag{56}$$

where ρ is the correlation coefficient between the returns of Stock1 and Stock2 and the two weights must add up to 100% as:

$$\omega_1 + \omega_2 = 1 \tag{57}$$

and the standard deviation of the portfolio as:

$$\sigma_p = \sqrt{\sigma_p^2}$$

Equation (56) simply states that the portfolio variance consists of two components – the variance component and the correlated component. If both weights ω_1, ω_2 are positive in (56), then portfolio risk increases monotonically with positive correlation. On the other hand, if $\omega_2 < 0$ then $\omega_1 > 1.0$ in order to satisfy the constraint in (57). Then the second term in (56) is negative. Increasing the positive correlation value may result in a reduction of portfolio risk.

Table 10.5 Portfolio Risk for various combinations

			------------------ $\sigma_{Portfolio}$		--------------
		ω_1 =	**0.5**	**1.5**	**1.8**
σ_1	σ_2	ρ	**Long Only**	**Long/Short**	**Long/Short**
0.5	0.5	0.2	0.387	0.742	0.909
		0.5	0.433	0.661	0.781
		0.8	0.474	0.570	0.628
		0.9	0.487	0.536	0.567
0.8	0.5	0.2	0.512	1.176	1.415
		0.5	0.568	1.097	1.287
		0.8	0.618	1.011	1.145
		0.9	0.634	0.981	1.094
0.5	0.8	0.2	0.512	0.776	0.995
		0.5	0.568	0.650	0.802
		0.8	0.618	0.492	0.546
		0.9	0.634	0.427	0.428
0.8	0.8	0.2	0.620	1.187	1.454
		0.5	0.693	1.058	1.250
		0.8	0.759	0.912	1.004
		0.9	0.780	0.858	0.908

Table 10.5 shows the portfolio risk (standard deviation) for different combinations of ω, σ *and* ρ. For the long only portfolio, it is easy to see that when either σ *or* ρ increased, the portfolio risk also increased. However, if the filtered correlation structure

increased with a corresponding reduction in the variance then an overall reduction in the portfolio risk is easily attainable. For $\omega_1 = \omega_2 = 0.5;\ \sigma_1 = \sigma_2 = 0.8;\ \rho = 0.2$ the portfolio risk is 0.62. Suppose post filtering one observes the following $\omega_1 = \omega_2 = 0.5;\ \sigma_1 = 0.5, \sigma_2 = 0.8;\ \rho = 0.5$, then the portfolio risk is reduced to 0.568. If both stock risk are reduced further, say $\sigma_1 = 0.5, \sigma_2 = 0.5$, $\rho = 0.5$, then the portfolio risk is further reduced to 0.433. This explains why the long only optimal portfolio using filtered correlation can have lower risk for the same level of expected return despite the fact that the correlation structure has gone up substantially.

As for the long-short portfolio, an increase in correlation with no change in the individual risk will lead to a reduction in portfolio risk. Take the example of $\omega_1 = 1.5, \omega_2 = -0.5;\ \sigma_1 = \sigma_2 = 0.8;\ \rho = 0.2$ the portfolio risk is 1.187. Post filtering, one observes $\sigma_1 = \sigma_2 = 0.8;\ \rho = 0.8$ and portfolio risk is 0.912. The reduction in portfolio risk comes from the increased correlation and the negative weight of Stock2 in the second term of equation (56). The negative effect from the correlated component may even overcome the variance component in some cases. Suppose the initial case was $\omega_1 = 1.5, \omega_2 = -0.5;\ \sigma_1 = \sigma_2 = 0.5;\ \rho = 0.2$ with portfolio risk at 0.742. The new situation becomes $\omega_1 = 1.5, \omega_2 = -0.5;\ \sigma_1 = 0.5,\ \sigma_2 = 0.8;\ \rho = 0.5$ and the portfolio is reduced to 0.650. The reverse happens if $\omega_1 = 1.5, \omega_2 = -0.5;\ \sigma_1 = 0.8,\ \sigma_2 = 0.5;\ \rho = 0.5$ where the portfolio risk increased to 1.097.

The simple examples illustrate that although correlation structure increased and variance on average has decreased post filtering, the net effect on portfolio risk is not obvious. The overall risk can either increase or decrease depending on the sign and size of the weights, and the increase or decrease in variance relative to correlation. Therefore, it is plausible to have an optimal portfolio, using the filtered correlation matrix, to behave like the optimal portfolio that is based on raw correlation matrix. This can possibly explain the similar performance results of unconstrained optimal portfolios.

Chapter 11

Summary

The main purpose of this dissertation is to evaluate the effectiveness of Neutron QuantumApp in filtering the estimation noise for large covariance matrix. The characteristic of the correlation and variance structure are examined prior to and post filtering. Three observations are established from the analysis. First, the results clearly show that Neutron's filtering process is yielding higher correlation coefficients in almost all cases. Second, there is generally a decline in variance for slightly more than half of the sample securities. The remaining securities show increased volatility after the filtration process although the magnitude is smaller than the magnitude of declining volatility. The average variance of all stocks in the samples declines. Third, the Neutron's filtering process reduces both the eigenvalues and the band width as predicted by RMT. This means the largest eigenvalue (reflecting the market factor) absorbs almost all the differences, thus making every stock in the sample to be highly correlated with the market factor. Not surprisingly, the overall correlation structure has increased dramatically and the larger percentage of stocks show lowered volatility.

In terms of portfolio risk, specifically pertaining to the MVP, the minimal risk from MV optimization using raw sample data is lower than the minimal risk from using the filtered data under short sale restriction. The efficient frontier generated from the raw sample dominates the efficient frontier from the filtered sample only at the MVP position.

When risk-return trade-off departs from the minimal position, the filtered efficient frontier quickly dominates the raw optimal frontier.

However, when the short sale restriction is removed the reverse is true; the filtered portfolio risk is lower than the raw portfolio risk. This phenomenon shows up consistently throughout the study. In fact, the filtered efficient frontier dominates the raw efficient frontier at every risk-return point. The results hold well during the depth of the financial crisis of 2008.

The predictive power of risk is showing the same behavior as above under short sale restriction. The MVP_{Raw} is a better predictor of actual risk than the MVP_{Filter}. As soon as the short sale constraint is removed, the opposite is true. Moreover, there is actually improvement in risk prediction when the optimal weights from MVP_{Filter} are applied to the raw data.

Since stocks are more correlated after the filtration process, there is less number of stocks required to generate an optimal portfolio with non-negative weight constraint. There are around 2.6 times more stocks in the MVP_{Raw} than in the MVP_{Filter}. There is definitely an extra degree of efficiency in using the filtered covariance matrix for optimizing risk and return.

In the study of minimal tracking risk portfolio (MTP), the raw covariance produces slightly better result in the minimum tracking error than the filtered covariance, but the latter produces more dominant efficient frontiers than the former. This is true for both constrained and unconstrained optimization. The number of securities with positive weights is very similar for both constrained and unconstrained results.

In terms of return performance, there is no apparent advantage using the filtered covariance over the raw sample covariance. Optimal portfolios such as MVP, MTP and riskier return portfolio have been tested. Although the filtered portfolios may have a slight edge over the raw portfolios at certain times, the differences are insignificant when measuring their return against the benchmarks such as NASDAQ 100 index and S&P 500 index. The naive portfolio or equally weighted portfolio has the best overall performance throughout the study. In fact, the hybrid model, based on the three fund separation between a tangency portfolio and the MVP, has better results than the standalone MVPs, but is still unable to outperform the naïve strategy.

The findings here add further support to other existing studies, that many “fancy” strategies and estimation noise reduction techniques cannot beat the simple naïve portfolio. The latter, surprisingly, performed extremely well during the recovery period of the financial crisis. Furthermore, the naïve portfolio has the least turnover cost of any portfolio in existence.

As for Neutron QuantumApp, there is not much value in using it for portfolio selection. Sample covariance matrix appears to do a decent job when Q=N/T is fixed at some value less than 0.75. Perhaps, Neutron may produce better risk measures for risk management purposes. Alternatively, it may find some use in trading strategies that rely heavily on better correlation estimates.

It is no wonder that Markowitz's elegant optimization will still remain an enigma for some time. The serious attempt by physicists to use Random Matrix Theory in finance is commendable but proved to be futile at this moment.

Bibliography

Baker T.H., P.J. Forrester and P.A. Pearce (1998), "Random Matrix Ensembles with an Effective Extensive External Change," Journal of Physics A, 31, 29, 6087.

Bengtsson, C. and J. Holst, (2002), "On Portfolio Selection: Improved Covariance Matrix Estimation for Swedish Asset Returns", Working paper, Lund University and Lund Institute of technology.

Best, M.J., and R.R. Grauer (1991), "On the Sensitivity of Mean-Variance-Efficient Portfolios to Changes in Asset Means: Some Analytical and Computational Results," Review of Financial Studies 4(2), 315-342.

Bouchaud, J. and M. Potters (2000), "Theory of Financial Risks – From Statistical Physics to Risk Management," Cambridge University Press, U.K..

Britten-Jones, M. (1999), "The Sampling Error in Estimates of Mean-Variance Efficient Portfolio Weights," Journal of Finance, 54(2), 655-671.

Brown, S.J. (1978), "The Portfolio Choice Problem: Comparison of Certainty Equivalence and Optimal Bayes Portfolios," Communication Statistics Simulation and Computing 7, 321-334.

Chan, L.K.C., J. Karceski and J. Lakonishok, 1999, "On Portfolio Optimization: Forecasting Covariances and Choosing the Risk Model", Review of Financial Studies, 12: 937-974.

Canner, N., N.G. Mankiw, and D.N. Weil (1997), "An Asset Allocation Puzzle," American Economic Review, 87(1), 181-191.

Cohen K.J. and J.A. Pogue, (1967), "An Empirical Evaluation of Alternative Portfolio Selection Models," Journal of Business 40, 166-193.

Crisanti A. and H. Sompolinsky (1987), “Dynamics of Spin Systems with Randomly Asymmetric Bonds: Langevin Dynamics and a Spherical Model,” Physics Review A 36, 4922-4939.

DelMiguel V., L. Garlappi and R. Uppal, (2009), “Optimal versus Naïve Diversification: How Inefficient Is the 1/N Portfolio Strategy?,” The Review of Financial Studies, 22:5, 1915-1953.

Disatnik, D.J., and S. Benninga, (2007), “Shrinking the Covariance Matrix – Simpler is Better”, Journal of Portfolio Management, 33-4, 56-63.

Elton E.J., M.J. Gruber, (1973), “Estimating the Dependence Structure of Share Prices – Implications for Portfolio Selection,” Journal of Finance, 28, 1203-1232.

Elton E.J., M.J. Gruber and M. Padberg, (1976), “Simple Criteria for Optimal Portfolio Selection,” Journal of Finance 31-5, 1341-1357.

Elton E.J., M.J. Gruber and M. Padberg, (1978), “Simple Criteria for Optimal Portfolio Selection: Tracing out the Efficient Frontier,” Journal of Finance, 33-1, 296-302.

Eun C.S. and B.G. Resnick, (1992), “Forecasting the Correlation Structure of Share Prices: A Test of New Models,” Journal of Bank Finance, 16, 643-656.

Fama E.F. and K.R. French, (1995), “Size and Book-to-Market Factors in Earnings and Returns,” Journal of Finance 50, 131-155.

Fama E.F. and K.R. French, (1997), “Industry Costs of Equity,” Journal of Financial Economics 43, 153-193.

Frankfurter, G.M., H.E. Phillips, and J.P. Seagle (1971), “Portfolio Selection: The Effects of Uncertain Means, Variances and Covariances,“ Journal of Financial Quantitative Analysis 6, 1251-1262.

Galluccio, G., J.P. Bouchaud and M. Potters (1998),"Rational Decisions, Random Matrices and Spin Glasses," Physica A 259, 449.

Green R.C., and B. Hollifield, (1992), "When will Mean-Variance Efficient Portfolios be well Diversified?" Journal of Finance 47, 1785-1809.

Guhr, T. (2001), "Econophysics", Lecture Notes, Department of Mathematical Physics, Lund Institute of Technology.

Jagannathan, R. and T. Ma, (2000), "Three Methods of Improving the Precision in Covariance Matrix Estimation", Unpublished working paper.

Jagannathan, R. and T. Ma, (2003), "Risk Reduction in Large Portfolios: Why Imposing the Wrong Constraints Helps", Journal of Finance, 54(4): 1651-1683.

Jobson, J.D. and B. Korkie, (1981),"Putting Markowitz Theory to Work," Journal of Portfolio Management, Summer.

Johnson, R. and D. Wichern, (1992), Applied Multivariate Statistical Analysis, Third Edition, Prentice-Hall, Englewood Cliffs, New Jersey.

Jolliffe, I., (1986), Principal Component Analysis. Springer-Verlag, New York.

Jorion, P. (1985), "International Portfolio Diversification with Estimation Risk," Journal of Business 58(3), 259-278.

Jorion, P. (1986), "Bayes-Stein Estimation for Portfolio Analysis." Journal of Financial and Quantitative Analysis, 21, 279-292.

Jorion, P. (1991), “Bayesian and CAPM Estimators of the Means: Implications for Portfolio Selection.” Journal of Banking and Finance, 15, 717-727.

Kan, R., and G. Zhou (2007), “Optimal Portfolio Choice with Parameter Uncertainty,” Journal of Financial and Quantitative Analysis, 42-3, 621-656.

King, B.F., (1966), “Market and Industry Factors in Stock Price Behavior”, Journal of Business, 39-1, 139-190.

Kohn, M. (1987), Practical Numerical Methods: Algorithms and Programs, Macmillan, New York.

Kroll, Y., H. Levy, and H. Markowitz (1984), “Mean-Variance versus Direct Utility Maximization,” Journal of Finance 39, 47-61.

Krzanowski, W., (1984), “Sensitivity of Principal Components”, Journal of the Royal Statistical Society, B 46(3), 558-563.

Laloux, L., P. Cizeau, J.P. Bouchaud, and M. Potters (1999), “Noise Dressing of Financial Correlation Matrices”, Physical Review Letters 83, 1467-1470.

Laloux, L., P. Cizeau, M. Potters and J. Bouchaud, (2000), “Random Matrix Theory and Financial Correlations”, International Journal of Theoretical Applied Finance 3(3), 391-397.

Ledoit, O. and M. Wolf, (2003), “Improved Estimation of the Covariance Matrix of Stock Returns with an Application to Portfolio Selection,” Journal of Empirical Finance, 10, 603-621.

Ledoit, O. and M. Wolf, (2004),“Honey, I Shrunk the Sample Covariance matrix”, Journal of Portfolio Management, 30(4): 110-119.

Lee, Y. (2001), “Noise Detection from Financial Correlation Matrices”, Academic Report, Massachusetts Institute of Technology, US.

Litterman, R., and K. Winkelmann (1998), Estimating Covariance Matrices, Goldman Sachs Risk Management Series.

Liu, L. and H. Lin, (2010), “Covariance Estimation: Do New Methods Outperform Old Ones?” Journal of Economic Finance, 34, 187-195.

Markowitz, H.M. (1952), “Portfolio Selection,” Journal of Finance 7, 77-91.

Markowitz, H.M. (1956), *Portfolio Selection: Efficient Diversification of Investments*, New York, John Wiley and Sons.

Markowitz, H.M. and A.F. Perold (1981), “Portfolio Analysis with Factors and Scenarios,” Journal of Finance 36, 871-877.

Martellini L., P. Priaulet and S. Priaulet (2003), Fixed-Income Securities: Valuation, Risk Management and Portfolio Strategies, Wiley Finance.

Mehta, M.L. (1967), Random Matrices and the Statistical Theory of Energy Levels, New York: Academic Press.

McNamara J.R. (1998), “Portfolio Selection Using Stochastic Dominance Criteria,” Decision Science 29(4), 785-801.

Michaud, R.O. (1989), “The Markowitz Optimization Enigma: Is “Optimized” Optimal?” Financial Analyst Journal, 45, 31-42.

Muirhead, R.J. (1982), Aspects of Multivariate Statistical Theory. New York: Wiley.

Pafka, S. and I. Kondor (2002),“Noisy Covariance Matrices and Portfolio Optimization,” European Physical Journal B, 27, 277-280.

Pafka, S. and I. Kondor (2003), “Noisy Covariance Matrices and Portfolio Optimization II,” Physica A, 319, 487-494.

Pafka, S. and I. Kondor (2004),“Estimated Correlation matrices and Portfolio Optimization,” Physica A, 343, 623-634.

Perold, A.F. (1984), “Large-Scale Portfolio Optimization,” Management Science 30(10), 1143-1160.

Plerou, V., P. Gopikrishnan, B. Rosenow, L.A.N. Amaral, and H.E. Stanley, (1999),”Universal and Nonuniversal Properties of Cross Correlations in Financial Time Series,” Physical Review Letters 83, 1471-1474.

Plerou, V., P. Gopikrishnan, B. Rosenow, L.A.N. Amaral, and H.E. Stanley, (2002), “Random matrix approach to cross correlations in financial data,” Physical Review E 65, 066126.

Press W.H., B.P. Flannery, S.A. Teukolsky and W.T. Vetterling, 1988, “Numerical Recipes in C: The Art of Scientific Computing”, Cambridge University press.

Roll, R., (1992), A Mean/Variance Analysis of Tracking Error, Journal of Portfolio Management 18(4), 341-360.

Ross, S.A., (1976),”The Arbitrage Theory of Capital Asset Pricing,” Journal of Economic Theory 13 341-360.

Sengupta A.M. and P.P. Mitra (1999), “Distribution of Singular Values for Some Random Matrices,” Physical Review E 60, 3.

Shannon, C.E. (1984), “A Mathematical Theory of Communication”, Bell System Technical Journal, 27-3, 379-423.

Sharifi, S., M. Crance, A. Shamaie and H. Ruskin (2004), “Random Matrix Theory for Portfolio Optimization: A Stability Approach”, Physica A, 335, 3-4, April, 629-643.

Sharpe, W.F. (1963), “A Simplified Model for Portfolio Analysis,” Management Science, IX, 277-293.

Sharpe, W.F. (1964),” Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk,” Journal of Finance 19, 425-442.

Sharpe, W.F. (1971), “A Linear Programming Approximation for the General Portfolio Analysis Problem,” Journal of Financial Quantitative Analysis 6 (5), 1263-1275.

Simaan, Y. (1997), “Estimation Risk in Portfolio Selection: The Mean Variance Model versus the Mean Absolute Deviation Model,” Management Sciences, 43(10), 1437-1446.

Stein, C. (1956), “Inadmissibility of the Usual Estimator for the Mean of a Multivariate Normal Distribution.” Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, University of California Press, 1, 197-206.

Stone, B.K. (1973), “A Linear Programming Formulation of the General Portfolio Selection Problem,” Journal of Financila Quantitative Analysis 8(4), 621-636.

Utsugi, A., K. Ino, and M. Oshikawa (2004), “Random matrix theory analysis of cross correlations in financial markets,” Physical Review E 70, 026110.

Zellner, A., and V.K. Chetty (1965), “Prediction and Decision Problems in Regression Models from the Bayesian Point of View,” Journal of American Statistics Association, 60, 608-616.

Appendix A

An Explanation of Random Matrix Theory

Markowitz's theory of optimal portfolio aims at generating efficient portfolios such that the overall risk is minimized for a given reward, or, conversely, reward is maximized for a given risk. Using variance as a risk measure and expected return as a measure for reward, the return on a portfolio is a linear combination of the returns on the assets forming the portfolio with weights given by the proportion of wealth invested in the assets. The portfolio variance can be expressed as a quadratic form of these weights with the volatilities and correlations as coefficients. The volatility and correlation estimates are extracted from historical data and they have to be reliable for any practical use.

Consider N assets, the correlation matrix contains $N(N-1)/2$ entries, which must be determined from N time series of length T. Suppose there are 500 securities, then there are 124,750 unique entries in the correlation matrix. In this example if one takes 5 years of daily returns for each security (assuming 250 trading days a year), then one has 625,000 of total observations. Dividing 625,000 by 124,750 one has around 5 data points. One has to increase T in order to increase the data points (30 years of daily data will yield 30 data points). Therefore, if T is not very large as compared to N, one would expect that the determination of the covariance matrix structure to be dominated by estimation noise. If this is the case, then the smallest eigenvalues of this matrix are the most sensitive to this noise. Any application that uses such correlation matrix can produce inefficient results.

Laloux et al. (1999) argue that in the MV optimization**, the eigenvectors corresponding to the smallest eigenvalues determine the least risky portfolios.** According to them:

> "Indeed, in the case of the S&P 500, 94% of the total number of eigenvalues falls in the region where the theoretical formula applies. Hence less than 6% of the eigenvectors, which are responsible for 26% of the total volatility, appear to carry some information."

Independently, Plerou et al. (1999) also find that empirical covariance matrices deduced from financial return series to contain a high degree of noise.

In this Appendix, a brief introduction to eigenvalues, eigenvectors and Principal Component Analysis (PCA) is in order.[1] Then a more detail discussion of RMT follows. Consider a square matrix $\boldsymbol{M}$. An eigenvector of $\boldsymbol{M}$ is defined as a non-zero vector $\boldsymbol{v}$ such that when the matrix is multiplied by $\boldsymbol{v}$, the result is the vector $\boldsymbol{v}$ scaled by $\boldsymbol{\lambda}$

$$Mv = \lambda v \tag{A.1}$$

[1] Refer to Kohn (1987) for detail explanations of eigenvalues and eigenvectors.

where λ is called the eigenvalue (or characteristic root) of ***M*** corresponding to ***v***. In short, the scalar value is a way to "stretch" out the information contained in M. Rewriting (A.1) as

$$(M - \lambda I)v = 0$$

where ***I*** is the *N x N* identity matrix. Therefore, a non-zero solution ***v*** exists if and only if

$$det(M - \lambda I) = 0 \qquad \text{(A.2)}$$

In other words, the eigenvalues of ***M*** are precisely the real numbers λ that satisfy (A.2).

As an illustration, consider a matrix ***M***

$$M = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & 4 \\ 0 & 4 & 9 \end{bmatrix}$$

$$det(M - \lambda I) = det\left(\begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & 4 \\ 0 & 4 & 9 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\right) = det\begin{bmatrix} 2-\lambda & 0 & 0 \\ 0 & 3-\lambda & 4 \\ 0 & 4 & 9-\lambda \end{bmatrix} = 0$$

which is

$$(2 - \lambda)[(3-\lambda)(9-\lambda) - 16] = -\lambda^3 + 14\lambda^2 - 35\lambda + 22 = 0$$

and the roots (eigenvalues) of the polynomial are 2, 1, and 11. The three eigenvectors corresponding to the three eigenvalues are:

$$\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \quad \begin{bmatrix} 0 \\ 2 \\ -1 \end{bmatrix} \quad \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix}$$

In order to illustrate the relationship between the eigenvalues and their eigenvectors, consider an *n x n* matrix $\boldsymbol{U}$ whose columns v_i are the eigenvectors of a symmetric *n x n* matrix $\boldsymbol{M}$

$$\boldsymbol{U} = \begin{pmatrix} v_1 & v_2 \ldots & v_n \\ | & | & | \\ \downarrow & \downarrow & \downarrow \end{pmatrix}$$

Premultiply both sides of the equation by *M*

$$\boldsymbol{MU} = \begin{pmatrix} Mv_1 & Mv_2 \ldots & Mv_n \\ | & | & | \\ \downarrow & \downarrow & \downarrow \end{pmatrix} = \begin{pmatrix} \lambda_1 v_1 & \lambda_2 v_2 \ldots & \lambda_n v_n \\ | & | & | \\ \downarrow & \downarrow & \downarrow \end{pmatrix}$$

$$= \begin{pmatrix} v_1 & v_2 \ldots & v_n \\ | & | & | \\ \downarrow & \downarrow & \downarrow \end{pmatrix} \text{x} \begin{pmatrix} \lambda_1 & 0 & 0 \\ \vdots & \lambda_2 & \vdots \\ 0 & 0 & \lambda_n \end{pmatrix} = \boldsymbol{U\Lambda}$$

where $\boldsymbol{\Lambda}$ is a diagonal matrix whose nonzero elements are the eigenvalues of $\boldsymbol{M}$**.** Premultiply both sides of the equation by $\boldsymbol{U^{-1}}$ yields

$$\boldsymbol{U^{-1}MU} = \boldsymbol{U^{-1}U\Lambda} = \boldsymbol{\Lambda}$$

The matrix $\boldsymbol{U}$ is an orthogonal matrix which has the property that $\boldsymbol{U^{-1}} = \boldsymbol{U^{T}}$. Therefore, there are *n* eigenvalues corresponding to *n* eigenvectors**.**

A common application of decomposition of a matrix into eigenvalues and eigenvectors is the Principal Component Analysis (PCA). A symmetric *n x n* real matrix ***M*** is said to be ***positive definite*** if $z^T Mz$ is positive for every non-zero column vector *z* of *n* real numbers. A ***positive semi-definite*** matrix (PSD) requires $z^T Mz$ to be always non-negative. Thus, the eigen decomposition of a PSD matrix yields an orthogonal basis of eigenvectors, each of which has a non-negative eigenvalue.

The sample covariance matrices are PSD and the orthogonal decomposition is the PCA. For the covariance or correlation matrix (in which each variable is scaled to have a unit sample variance), the eigenvectors correspond to principal components and the eigenvalues to the variance explained by the principal components. PCA of the correlation matrix provides an **orthonormal eigen-basis** for the space of the observed data. In this eigen-basis, the largest eigenvalues correspond to the principal-components that are associated with most of the covariability among a number of observed data.

Similarly, one can extract the eigenvalues and eigenvectors from a sample covariance matrix and then analyze the spectrum of these values that are affected by noise. Once a better understanding of the spectrum, one can proceed to find ways to mitigate this measurement noise and then “transform” the original covariance matrix into a “less noisy” covariance matrix. In other words, how can one distinguish the “signal” (or true information) from the “noise” from the eigenvectors and eigenvalues of a correlation matrix?

In order to answer the question, one starts off with the simple concept of comparing an N x N empirical correlation matrix **C** to a purely random matrix as one could obtain from a finite time series of N uncorrelated assets. The presence of true information can then be inferred from the deviations from the random matrix.

The element in the empirical correlation matrix ***C*** is written as:

$$\boldsymbol{C}_{\mathrm{ij}} = \left(\frac{\boldsymbol{1}}{\boldsymbol{T}}\right) \sum_{\boldsymbol{t}=\boldsymbol{1}}^{\boldsymbol{T}} \boldsymbol{\delta x}_{\mathrm{i}}(\boldsymbol{t}) \boldsymbol{\delta x}_{\mathrm{j}}(\boldsymbol{t}) \qquad \text{(A.3)}$$

for assets i and j, and time period t.[2] The δx's are time series of price changes, subtracted from their average values, and rescaled to have constant unit volatility. In matrix notation, one can rewrite (A.3) as $C = \left(\frac{1}{T}\right) M\, M^T$ where M is a N X T rectangular matrix and ^{T} is matrix transposition. This is also known as the Wishart matrix (Baker et. al (1998)). Denote $\rho(\lambda)$ as the density of eigenvalues of C . This is given as:

$$\rho(\lambda) = \left(\frac{1}{N}\right) \frac{d\, n(\lambda)}{d\lambda} \qquad \text{(A.4)}$$

where $n(\lambda)$ is the number of eigenvalues of C less than λ.

[2] The analyses follow Laloux et al. (1999)

If M is a $T\ x\ N$ **random matrix**, then the density function in (A.4) is exactly known in the limit $N \ \rightarrow \infty, T \rightarrow \infty\ and\ for\ a\ fixed\ Q = \ {}^{T}/_{N} \ \geq 1$ (Sengupta and Mitra (1999))

$$\rho(\lambda) = \frac{Q}{2\pi\sigma^2} \frac{\sqrt{(\lambda_{max}-\lambda)(\lambda-\lambda_{min})}}{\lambda} \qquad \text{(A.5)}$$

where $$\lambda_{min}^{max} = \sigma^2(1 + \frac{1}{Q} \pm 2\sqrt{1/Q}) \quad \text{with } \lambda \in \lfloor \lambda_{min} - \lambda_{max} \rfloor \qquad \text{(A.6)}$$

and σ^2 is the variance of the elements of M, equal to 1 with normalization. Equations (A.5) and (A.6) are approximately valid at finite N and T when N and T are not small. According to (A.5) and (A.6), the eigenvalues of the Wishart matrix distribute only in the range: $\lfloor \lambda_{min} - \lambda_{max} \rfloor$. In the limit Q =1 the normalized eigenvalue density of M is the **Wigner semi-circle law**[3] which can be written as:

$$\rho(\lambda) = \frac{1}{2\pi\sigma^2} \sqrt{4\sigma^2 - \lambda^2}, \ |\lambda| \leq 2\sigma \qquad \text{(A.7)}$$

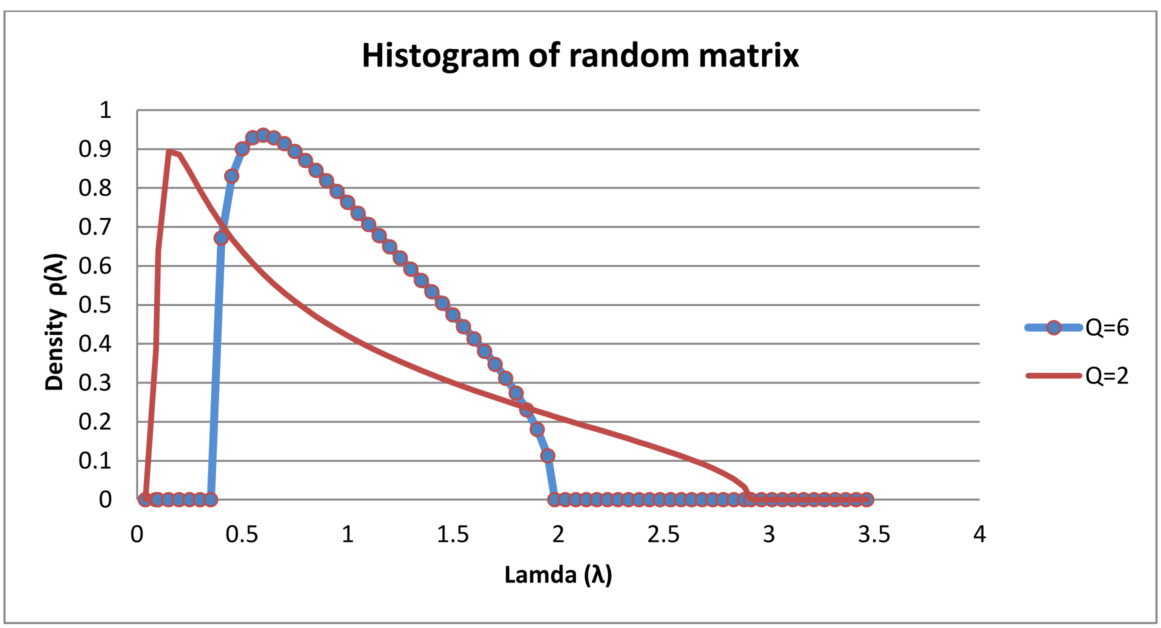
According to RMT predictions as in (A.5), for the lower spectrum from 0 to λ_{min}, there are no eigenvalues. Near this edge, the density $\rho(\lambda)$ has the highest value (except in the limit Q=1; $\lambda_{min} = 0$) and then declines slowly until it reaches 0 at the upper edge λ_{max}. In essence, $\lambda_{min}\ and\ \lambda_{max}$ are the theoretical minimum and maximum eigenvalues that determine the bounds of the theoretical eigenvalue distribution. If the eigenvalues of C are beyond these bounds it is said that they deviate from the random (or

[3] See Mehta (1967)

theoretical) boundary. Note that the results are valid only in the limit, $N \to \infty$. For finite N, the edges at both ends are somewhat smoothed with a small probability of finding eigenvalues above λ_{max} and below λ_{min}.

Consider two cases in order to illustrate the equation (A.5). In the first case, T=3,000 and N=500 such that Q = 6. The eigenvalue associates with the random noise ranges between [0.35 and 1.983]. The density function is represented by the blue line with the circle marker in Figure A2.1. In the second case, T=2,000 and N=500 such that Q=2. The eigenvalues associates with the random noise ranges between [0.085 and 2.914]. This is represented by the red line in Figure A-1. For a smaller Q value, the noisy eigenvalues cover a larger area. The random spectrum is reduced for larger T. Noise is reduced as T increases relatively to N, the number of securities.

Figure A-1



Recall that Laloux et al. (1999) argue that in the MV optimization, the least risky portfolios happen to correspond to the smallest eigenvalues; 94% of the spectrum of these correlation matrices can be fitted with random matrix. In other words, the distribution of the smaller eigenvalues resembles the distribution generated by random noise as in Figure 2.1. Only the larger eigenvalues (a smaller set) are unaffected by the measurement noise (i.e., those $\lambda > \lambda_{max}$).

At this juncture, I present the behavior of the distribution of eigenvalues from a sample correlation matrix with N=87 stocks from the NASDAQ 100 sample and T=119 returns; Q=1.367. Three separate dates were chosen to represent three sub periods. These dates are: 2/13/2007, 3/9/2009 and 1/13/2012. The eigenvalues for the correlation matrix for these three dates were extracted, combined and sorted into frequency bins between $\lambda_{min} = 0.02$ and $\lambda_{max} = 3.44$, which is the RMT prediction interval of random noise.

Figure A-2

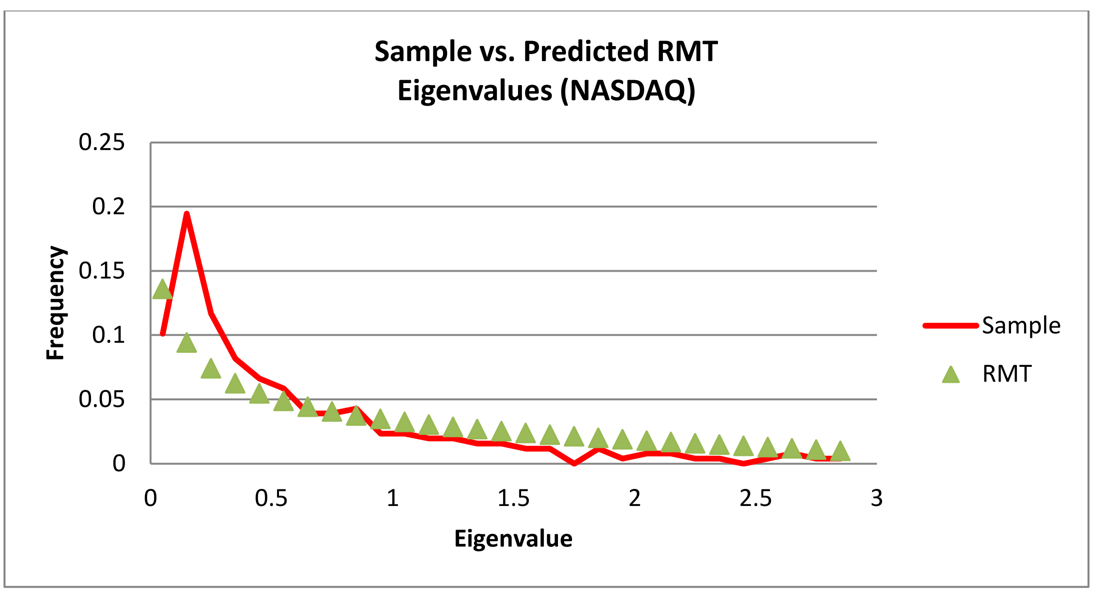
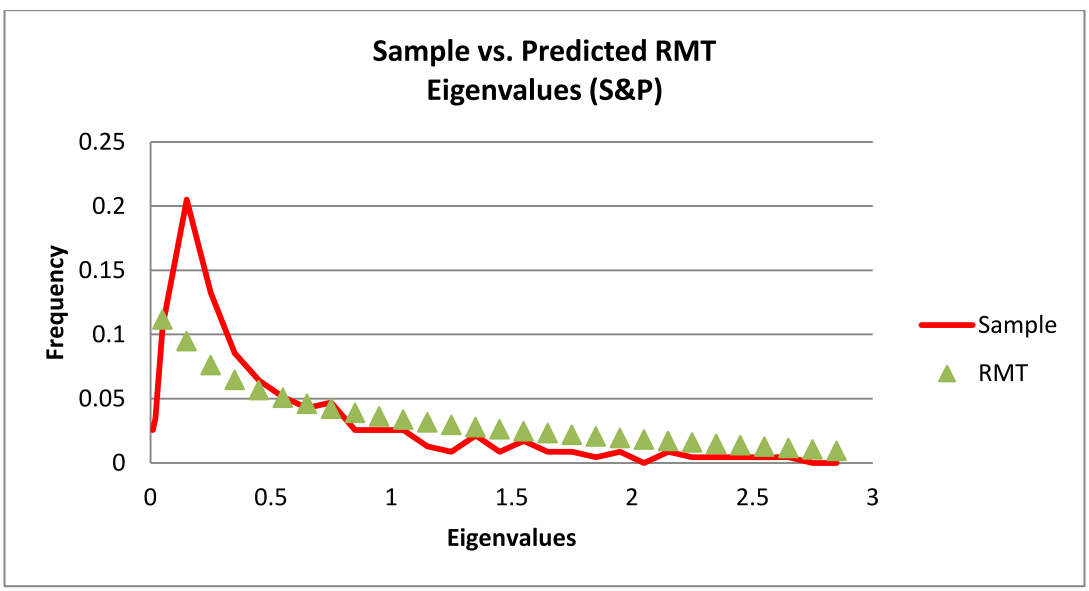


Figure A-2 shows the histogram of the sample eigenvalues from the combined 3 dates. Note that those eigenvalues greater than λ_{max} are not displayed. The largest eigenvalue for the 3 dates has a value of 54.7 which is 15.9 times the value of λ_{max}. Moreover, 97.7% of the total eigenvalues are within the RMT prediction band which is consistent to Laloux et al. (1999) finding.

The experiment is repeated for 80 randomly selected stocks from the S&P 500 sample with T=119; Q=1.487; $\lambda_{min} = 0.032$ and $\lambda_{max} = 3.312$. The histogram of eigenvalues for the same three dates is displayed in Figure A-3

Figure A-3



The largest eigenvalue in the S&P sample for the 3 dates is 54.3; 16.4 times greater than λ_{max}. Again, between 97.5% and 98.75% of the total eigenvalues fall into the RMT prediction band. The implication is clear. The information content contained in the sample correlation matrix is indistinguishable from random noise. Only the first and second largest eigenvalues appear to contain real correlation data.

Shannon Entropy

Entropy is a measure of uncertainty associated with a random variable. Shannon entropy can be interpreted as the average number of bits required in order to specify the probability of being in a system state. In other words, it quantifies the expected value of the information contained in a message, measured in bit (see Shannon (1984)). Mathematically, the Shannon entropy of **A** is written as:

$$H(\boldsymbol{A}) = -\sum_{i=1}^{n} p_i log_2 p_i$$

where

H(A) is the information size of a set A

p_i is the probability of i^{th} element.

The higher the Shannon entropy, the more noisy or uncertainty the elements are.