



EKONOMI
HÖGSKOLAN
Lunds universitet

Master's Thesis in Finance FEK 591
September 23rd, 2003

Noise Reducing Methods for Correlation Matrices Improved Return and Risk Possible in Portfolio Management?



Tutors:

Göran Anderson, Department of Business Administration
Thomas Guhr, Department of Mathematical Physics

Authors:

Per-Johan Andersson
Andreas Öberg

Abstract

- Title Noise Reducing Methods for Correlation Matrices
Improved Return and Risk Possible in Portfolio Management?
- Date of Presentation September the 23rd, 2003
- Authors Per-Johan Andersson and Andreas Öberg
- Tutors Göran Andersson, Department of Business Administration
Thomas Guhr, Department of Mathematical Physics
- Course Master's Thesis in Finance
- Keywords Markowitz Mean Variance Theory, Random Matrix Theory, Power Mapping, Filtering techniques, Correlations, Risk and Diversification
- Resume The importance of the alluring goal of having the “perfect” or “true” information is more than visible in the financial world. In portfolio management financial correlation matrices measure the unsystematic correlations between assets e.g. stocks. Finding the true correlations between stocks would enable the creation of portfolios with less risk. In reality this information is more than hard to come by, partly since the correlations are buried under noise that effectively hides the precious, true correlations. If this noise could be reduced or even eliminated one would be able to unveil the true correlations, a powerful “weapon of knowledge” in portfolio management. Furthermore if patterns could be found in how the correlations change over time even further possibilities would emerge enabling faster adaptation to critical market trends. In the thesis we focus on evaluations of two different methods for noise reduction; Power Mapping, and RMT-filtering (a technique bases upon Random Matrix Theory). Both noise reduction methods are the offspring of Econophysics, a new scientific field combining Economics, Business Administration, and Physics. We also studied a new “hybrid method” previously created by the authors that merges Power Mapping and RMT-filtering. We came to the conclusion that all methods reduce risk for most of the portfolios, sometimes very strongly. Several aspects and behaviors of the methods were examined, with the aim to increase the understanding of these new, complex methods and the correlation structure in general. With a better understanding of the correlation matrices, we might indeed come one step closer to the shining promise of investments with less risk and even perhaps higher return.

Contents

CHAPTER 1 INTRODUCTION	6
1.1 BACKGROUND	6
1.2 PROBLEM STATEMENT	7
1.3 PURPOSE	7
1.4 DELIMITATIONS	7
1.5 TARGET GROUP	8
1.6 BRIEF OUTLINE	8
CHAPTER 2 METHOD	10
2.1 CHOICE OF SUBJECT	10
2.2 INFORMATION SOURCES AND EARLIER STUDIES	10
2.3 DATA PROPERTIES	11
2.4 SOURCE CRITICISM	11
2.5 CRITICISM OF CHOSEN METHOD	11
2.6 MODEL SENSITIVITY TO DATA PROPERTIES	12
2.7 QUANTITATIVE STUDY	12
2.8 METHOD OVERVIEW	12
CHAPTER 3 THEORY	13
3.1 THE MEAN-VARIANCE THEORY	13
3.2 DIVERSIFICATION	18
3.3 RMT-FILTERING - A NOISE REDUCING METHOD FOR CORRELATION MATRICES	20
3.4 OVERVIEW OF THE CORRELATION MATRICES	24
3.4 NOH'S MODEL – A MODEL TO SIMULATE STOCK MARKETS	25
3.4.1 COLLECTED VERSUS SIMULATED STOCK DATA	25
3.4.2 AN INTRODUCTION TO NOH'S MODEL	26
3.4.3 TIME SERIES IN NOH'S MODEL	26
3.4.5 NOISE INTRODUCED BY NOH'S MODEL	28
3.4.6 SPECTRAL DENSITIES OF THE CORRELATION MATRICES	28
3.5 POWER MAPPING – A NOISE REDUCING METHOD FOR CORRELATION MATRICES	30
3.5 EARLIER EMPIRICAL FINDINGS AND RESULTS	33
3.5.1 REDUCING RISK WITH BETTER CORRELATION MATRICES	33
3.5.2 CHANGING CORRELATIONS OVER TIME	35
3.5.3 GRAPHICAL PRESENTATION OF IMPROVED CORRELATION MATRICES	36
3.5.4 THE SPECTRAL DENSITIES FOR DIFFERENT AMOUNTS OF TIME STEPS – CLEARER PEAKS FOR MORE TIME STEPS	38
CHAPTER 4 EMPIRICAL METHODOLOGY	39
4.1 STOCK DATA	39
4.1.1 FIRST DATA SET – SWEDISH STOCK RETURNS	39
4.1.2 SECOND DATA SET – U.S. INDUSTRY PORTFOLIO RETURNS	39
4.1.3 RETURNS	40
4.2 REPLACING MISSING DATA	41
4.3 CREATION OF THE STOCK PORTFOLIOS	44

4.4 MEAN VALUES OF THE COLUMNS OF THE CORRELATION MATRICES	45
4.5 PORTFOLIOS WITH DIFFERENT AMOUNT OF STANDARD DEVIATION	45
4.6 IMAGE FILTERING	46
4.7 OPTIMIZATION PROCESS USED IN MATLAB 6.5	46

CHAPTER 5 RESULTS AND ANALYSIS **47**

5.1 A STUDY ON 197 SWEDISH STOCKS	47
5.1.1 INTRODUCTION	47
5.1.2 VARIABLE DESCRIPTION	47
5.1.3 OVERVIEW	48
5.1.4 MAIN RESULTS - UNCONSTRAINED WEIGHTS	48
5.1.5 MAIN RESULTS - CONSTRAINED WEIGHTS	63
5.1.5 CORRELATION STRUCTURE	74
5.1.6 MEAN VALUES OF THE CORRELATION MATRICES	75
5.1.7 THE SWEDISH STOCK DATA – 15 YEARS FOR 48 COMPANIES	76
5.2 A STUDY ON 48 U.S. INDUSTRY PORTFOLIOS STOCKS	76
5.2.1 INTRODUCTION	76
5.2.2 DATA OVERVIEW	76
5.2.3 MAIN RESULTS – NO CONSTRAINTS IMPOSED	77
5.2.4 MAIN RESULTS – CONSTRAINTS IMPOSED	87
5.2.5 CORRELATION STRUCTURE	88
5.2.5 MEAN VALUES OF THE CORRELATION MATRICES	91
5.3 STANDARD DEVIATION TESTS	94
5.4 IMAGE FILTER TEST	95
5.5 “PM-FILTERING” – A NEW TOOL FOR IMPROVING CORRELATION MATRICES?	96
5.5.1 INTRODUCTION	96
5.5.2 RESULTS FOR 197 SWEDISH COMPANIES (NO CONSTRAINTS)	96
5.5.3 RESULTS FOR 48 U.S. INDUSTRIES (NO CONSTRAINTS)	97
5.5.4 CONCLUSIONS FOR PM-FILTERING	98
5.6 THE MEAN VALUES OF THE COLUMNS OF THE CORRELATION MATRIX	99
5.6.1 INTRODUCTION	99
5.6.2 DATA DESCRIPTION	99
5.6.3 COLUMN MEAN VALUES	99
5.5.3 CORRELATION BETWEEN COLUMN MEAN VALUES AND THE RETURN	103
5.6.4 CONCLUSIONS	105

CHAPTER 6 SUMMARY AND CONCLUSIONS **106**

THE 197 SWEDISH COMPANIES	106
U.S INDUSTRY PORTFOLIO	107
PORTFOLIOS WITH DIFFERENT STANDARD DEVIATIONS	108
IMAGE FILTER TESTS	108
PM-FILTERING RESULTS	108
THE MEAN VALUES OF THE COLUMNS OF THE MATRICES	109
SMALLER WEIGHTS FOR POWER MAPPING AND RMT-FILTERING	109
PROBLEM STATEMENT SUMMARY	109

CHAPTER 7 SUGGESTIONS FOR FUTURE STUDIES **110**

REFERENCES **112**

LITERATURE	112
ARTICLES	112
INTERNET	114
PERSONAL INTERVIEWS	115
DATABASES	115
FRONTPAGE	115
A.1 PORTFOLIOS CREATED USING SWEDISH STOCK DATA	116
A.1.1 PORTFOLIOS NUMBER 2 AND 3 NO CONSTRAINTS IMPOSED	116
A.1.2 PORTFOLIOS 2 AND 3 CONSTRAINTS IMPOSED	119
A.2 PORTFOLIOS CREATED USING U.S. INDUSTRY PORTFOLIOS	122
A.2.1 PORTFOLIOS NUMBER 2 AND 3 NO CONSTRAINTS IMPOSED	122
A.2.2 WEIGHTS EXCLUDED (NO CONSTRAINTS)	127
A.3 COMPANIES OF THE PORTFOLIOS	128
A.3.1 THE 197 SWEDISH COMPANIES	128
A.3.2 U.S. INDUSTRY PORTFOLIOS	136
A.4 NOH'S MODEL	143
A.5 PRINCIPAL COMPONENT ANALYSIS - PCA	147
A BRIEF INTRODUCTION TO PRINCIPAL COMPONENT ANALYSIS	FEL! BOKMÄRKET ÄR
INTE DEFINIERAT.	
A.6 PROGRAM CODE	150
A.6.1 PROGRAM CODE OVERVIEW	150
A.6.2 MAIN_ANALYSIS.M	151
A.6.3 MEANCOLUMN.M	164
A.6.4 DATA.M	169

Chapter 1

Introduction

This first introductory chapter contains a concise background, the problem statement, purpose, and the delimitations made. The chapter ends with a brief outline of the thesis.

1.1 Background

In 1952 Harry M. Markowitz¹ developed theories that took into account both the risk and return of investments, earlier economical theories focused mainly on the properties of return. Harry M. Markowitz later received the Bank of Sweden Prize in Economic Sciences (in Memory of Alfred Nobel) as an appreciation for his work concerning the theory of portfolio choice.

In today's financial world a thorough examination of involved risks in investments is a day to day business and a must for every serious financial institution or investor. Applying the concept of diversification in investments reduces risks without necessary sacrificing return; the best solution would be to use financial assets that are less than perfectly positive correlated². This obviously means that knowledge about the dependence and co-movement i.e. the correlation, between different assets in the portfolio is of crucial interest to the investor. In Markowitz's Mean Variance Portfolio Optimizations portfolios meeting the investor's criteria are created based on these correlation matrices.

A main issue concerning the correlation matrices of the assets e.g. stock data is that the true correlation matrix is hidden under a cloud of random noise. The reason is that when finite length of time series are used to estimate cross-correlations noise is introduced³, and in reality the time series are always finite. The, by the investor, created portfolios are based upon the correlation matrix, which naturally means that the correlation matrix used in the process of making portfolios has to be as close to the true correlation matrix as possible. If these two correlation matrices are greatly dissimilar the created portfolios will not work well in reality. But the calculation of accurate correlation matrices needs enormous amounts of data; often several tens of thousands of data points for every company is needed when often far less data points are available. With new technologies borrowed from as distant areas as Random Matrix Theory, originally developed for complex quantum systems, where the interactions of subunits is not known⁴, the possibility to improve and better understand the nature of the correlation matrices has emerged. With better understanding of the correlation matrices, we might indeed be one step closer to the shining promise of investments with less risk and even perhaps higher return.

¹ Harry M. Markowitz, 1952. Portfolio selection: Journal of Finance 7, pp 77-91.

² Rosenow, B., Plerou, V., Gopikrishnan, P., Stanley, E., 2002, Portfolio Optimization and the Random Magnet Problem.

³ Plerou, V, Gopikrishnan, P et al, 2001, A Random Matrix Approach to Cross-correlations in financial Data.

⁴ Åström, S, 2002, Quantum Chaos

1.2 Problem Statement

Main problems and questions:

- Do Power Mapping and RMT-filtering produce portfolios with less risk than the original Sample matrix?
- Do the methods work differently on different segments of the market, for example on portfolios associated with different levels of risk?
- What effects does imposing constraints such as forbidding negative weights have on the results of the different methods?
- Is there a pattern to be found in the mean values of the correlation matrices? If so are they possible to connect to the return in any way?
- The new “PM-Filtering” has produced good results in earlier limited studies, does this new hybrid method produce good results with more detailed data as well?
- Is it possible to create portfolios with a better return to risk ratio with the new methods?

1.3 Purpose

The main purpose of this thesis is to evaluate the ability to minimize the risk in portfolio management by using two noise reducing methods on the correlation matrix of the assets; the first based on Power Mapping, the other based on filtering techniques derived from Random Matrix Theory (RMT).

Secondary purposes are:

- to evaluate the hybrid method “PM-filtering”
- to furthermore evaluate both noise reduction methods in a more general aspect, examining their properties and behaviours from as many angles as possible
- to study the mean values of the columns of the correlation matrices and try to find possible existing patterns

1.4 Delimitations

The main focus of this thesis has been to evaluate the potential to minimize risk in portfolio management by creating correlation matrices that are improved with two methods to reduce the contained noise (we also study PM-filtering but not as

extensively). We have chosen not to study more methods for noise reduction in matrices to prevent overly complex calculations. Many aspects of the different methods have been studied, but naturally one has to draw the line somewhere. Both the in-sample and the out-of-sample periods could for example be varied even more and be more systematically studied. The image filter studied could be expanded much more, since it now is merely a brief glimpse in the area. The mean values of the columns of the matrices also have several more promising areas but due to the limited amount of time available these areas have not been fully examined. For example, only negative/positive values etc are calculated; the absolute values of the columns have also been examined but are not presented in the thesis due to time limitations. The data is fairly extensive in this test but naturally it could be expanded even more; with e.g. data from different countries and markets and with even higher frequencies. The actual computations and calculations themselves have also created some limitations. For example initially the portfolios were planned to be created every day for the evaluation period, but during the tests it became clear that it would take far too long to execute the calculations for each situation, because of the long calculation times the period length was changed to five days. That is also one reason why the thesis focus more on the minimum variance portfolio (see Chapter 3.1), since it is orders of magnitudes faster to compute thus opening up the possibilities to experiment more.

1.5 Target Group

The target group of this thesis is mainly students and professionals that possess moderate knowledge of finance and risk management. It's somewhat recommended that students have completed their penultimate year of a master's degree in finance or economics.

1.6 Brief Outline

- Chapter 2 In Chapter 2 the properties of the used data are discussed.
- Chapter 3 Chapter 3 contains a description of the relevant theory and a brief overview of earlier findings in scientific articles with similar subjects. Markowitz Mean Variance Theory, Random Matrix Theory (RMT) and Power Mapping, RMT-filtering are among others explained. Previous results from earlier studies are also presented.
- Chapter 4 Chapter 4 contains the empirical discussion
- Chapter 5 Chapter 5 contains the results and the analysis
- Chapter 6 Chapter 6 contains the conclusions

Chapter 2

Method

The overall method used in this thesis is briefly discussed in this chapter. A more detailed and mathematical description of the method is given in Chapter 3 – Theory and Chapter 4 – Empirical methodology.

2.1 Choice of Subject

After an earlier brief study, performed by the authors of this thesis, the idea to continue where we stopped the last time was intriguing. The limited data in the previous study were the largest previous problem; thus the focus of this thesis has been to study (and expand) the methods on more high-frequent data for a larger number of companies. In the earlier thesis a hybrid method was developed that produced superior results. Were these results just a random occurrence, or could this method actually prove to work for more realistic data? With questions like these begging for answers it was natural to continue and expand in every direction possible, looking for “true” correlations and patterns in the sometimes random and chaotic tide of stock information.

2.2 Information Sources and earlier studies

The overall general theory for portfolio optimization i.e. Mean Variance Theory is well known and well documented, and has been studied in both textbooks, scientific articles and on the internet as well. Books like *Econophysics* by Thomas Guhr⁵ and *Quantum chaos* by Sven Åberg⁶ have been used as literature to get an understanding of the subject *Econophysics* as well as relating areas. The filtering theories presented by Gopikrishnan et al⁷ based upon Random Matrix Theory and the quite newly developed Power Mapping by Guhr et al⁸ have been studied in articles.

Personal interviews have also been held with Thomas Guhr and Anders Johansson (representing Vinga Systems AB, the creators of Vinga Vedrò and EcoWin Asset Allocation system) on several occasions during the last year.

⁵ *Econophysics* T. Guhr, Department of Mathematical Physics Lund University, 2001)

⁶ Sven Åberg, 2002, *Quantum Chaos* Lund University

⁷ Gopikrishnan et al, 2001 and Gopikrishnan et al 2002

⁸ Thomas Guhr & Bernd Kälber, 28 Jun 2002, arXiv:cond-mat/0206577 v1

2.3 Data Properties

The main data used for the calculations in the thesis consists of the companies currently listed on the Swedish Stockholm Stock Exchange (Stockholmsbörsen). This means that companies no longer listed on the stock exchange have been excluded from the calculations. There is thus a survivorship bias in the data. In addition, data on merged companies is available only since after the time of the merger. Furthermore in cases when a substantial amount of data points are missing the company has been excluded, this is also the case when data is available only for a short period of time i.e. companies listed within the latest years. The data used in the calculations are adjusted for splits, issues and dividends. Further empirical and methodology discussions regarding the data, selection and processing/handling of the data is referred to Chapter 4.1.1

More data containing monthly returns for 48 U.S. industry portfolios have also been used; the data ranges between 1926 and 2002. For more information see Chapter 4.1.2.

2.4 Source Criticism

This thesis uses the data presented in Chapter 2.3 to calculate the correlations matrices. The Swedish Stockholm Stock Exchange can be regarded as a reliable source for stock data.

The U.S. research return data presented in Chapter 4.2 is collected/obtained and prepared by Professor Kenneth French at University of Dartmouth in association with Professor Eugene Fama at University of Chicago. The reliability of this data should also be regarded as very high.

2.5 Criticism of Chosen Method

Many areas have been expanded in this thesis compared to the previous one written on the same subject. The most important expansion is the data. The data is much more high-frequent and there are also many more companies as well as industry data ranging from 1926. Naturally more data would have been even better to study. Also even though many different settings have been tried out it would have been good to try out even more setups, for example the length of the out-of-sample period and the lengths of the in-sample would be interesting to study more, since the results seem to fluctuate fairly strongly when these are changed. Also since we are trying to eliminate the noise in the correlation matrices it would have been meaningful to study a simulated correlation matrix where we know exactly how much noise we have and can calculate how much noise we have reduced. There are also better ways of getting more accurate expected returns than with the model we use. The model is simple but according to Anders Johansson at Vinga System⁹, this method is one of the most common ones. Also to use data that have lacking

⁹ Johansson, Anders. Vinga System. Interviews conducted during 2003-08

information might introduce unexpected errors as well, naturally it would have been better with data with no or less lacking data points. Though the amount of data replaced artificially (12%) is close to the same amount that is being replaced in the study by Swartz¹⁰ (10%).

2.6 Model Sensitivity to Data Properties

The data for the Swedish companies is daily (with 250 values every month). With data with even higher frequency, for example minutes or seconds, the analysis would in some aspects be more correct since we would then be able to study very high-frequent noise, now we are only able to study longer trends. It is unclear how the models would react to extremely high-frequent data (especially for Power Mapping since it has been tried very little on real data). The results presented in Chapter 5.2 produces somewhat extreme results indicating that the choice of the in-sample period can affect the results strongly. Even more tests with different lengths of the in-sample would be preferable. Many tests have been performed to study the sensitivity of the computer algorithms used, generally the loss of information occurring in some calculations are very small (one example is converting to 16bit numbers for the image filter which had a loss of about 10^{-20}) showing that the choice of computer algorithms affect the results extremely little in our cases.

2.7 Quantitative Study

Due to the very nature of the subjects examined, the thesis is quantitatively oriented, though a substantial effort has been made to present the quite mathematically theory and the final results in an easy to understand way. The theory chapter is also written in a not too strict economical or mathematical manner, this to explain the main concepts, ideas and practical applications in text form for the reader who lacks mathematical or economical background. This since the thesis after all is presented at the department of Business Administration at Lund University.

2.8 Method Overview

In the thesis we focus on two collections of data, the first consisting of 197 Swedish stock companies and the other consisting of 48 U.S. industries. The initial Sample correlation matrices are improved by noise reducing methods, mainly Power Mapping and RMT-filtering. Portfolio optimization is then performed with the correlation matrices for Sample, Power Mapping and RMT-Filtering. These portfolios are then tested in the following period for different expected returns. The most important ability we are studying is the ability to create a portfolio with as little risk as possible; our main goal is to reduce the risk with the portfolios created with noise reduced correlation matrices.

¹⁰ Schwartz Tal, March 15, 2000, How to Beat the S&P 500 with Portfolio Optimization

Chapter 3

Theory

This chapter presents the relevant theory used in this thesis

The theory chapter is quite extensive, this serves two important purposes. First of all the theory presented is in many aspects new and therefore needs a sufficient description; secondly the theory is often mathematical and complex and the authors therefore strive to explain the essence in a general easy to understand way.

3.1 The Mean-Variance Theory

What is referred to as modern portfolio theory was first introduced in 1952 by Harry M. Markowitz¹¹, the original theory was soon after developed and refined by among others Mossin, Sharpe and Litner¹². The Mean-Variance Theory is today the simplest model of investments that is sufficiently rich to be directly useful in applied financial problems¹³. In today's financial world the theory is widely used as a efficient and powerful tool in many areas of analytical portfolio and risk management, the reason is mainly because of its simplicity. The equilibrium version of the Mean-Variance Theory is called the Capital Asset Pricing Model (CAPM) and was first developed and presented by William F. Sharpe.

The mean variance method optimizes a portfolio in the aspects of the variance and the expected return. Therefore unlike naive diversification which means selecting a number of assets randomly, diversification according to Markowitz means improving the selection process by finding the portfolios which have certain characteristics e.g:¹⁴

- For a given level of expected return, the portfolio has the lowest standard deviation;
- For a given level of standard deviation, the portfolio has the largest expected return.

There are newer and interesting theories that are more or less linked to the Mean-variance theory but those theories will not be the subject of discussion in this thesis.

¹¹ Harry M. Markowitz, 1952. Portfolio selection: Journal of Finance 7, pp 77-91.

¹² E. Elton, M. Gruber, 1995. Modern Portfolio Theory and Investment analysis

¹³ Philip H. Dybvig, 1997. Mean variance Theory : Lecture notes Washington University in Saint Louis

¹⁴ Kee Hong Bae, 2000, Portfolio Diversification, Korea University Seoul.

The Mean-variance theory and the CAPM is based on some main ideas¹⁵:

- 1) Take on risk in proportion to the risk premium
- 2) Diversification pays.
- 3) The market rewards you for taking a share of economy-wide risk.
- 4) The market does not reward you for taking on security-specific (idiosyncratic) risk.
- 5) All investors hold a mixture of two portfolios, one risk less (if there is a risk less asset involved), and the market portfolio.

Furthermore from a theoretical point of view the assumptions of the mean-variance theory are:¹⁶

- A single-period model
- Preferences depend only on the mean and variance of payoffs, which means:
 - At a given mean, lower variance is preferred
 - At a given variance, a higher mean is preferred
- Price-taking with no taxes or transaction costs (e.g. perfect capital markets)

Further assumptions concerning the preferences of the investors, from the assumptions above one will assume that investors care only about the means and variances of the returns of their portfolios over a particular period. This states the following assumptions¹⁷:

- Investors prefer higher wealth
- Investors are risk-averse, and prefer lower risk

Mean-variance portfolio is, as implied above, a portfolio theory which asserts that investment allocations are made based upon the first two moments of the perceived payoff distribution associated with various asset holdings, or portfolios, namely¹⁸:

- 1) Expected return of a portfolio, $E(r_p)$
- 2) Standard deviation (or variance) “Risk” of the return of a portfolio (σ_p). The expected return of a portfolio is defined as¹⁹:

¹⁵ Philip H. Dybvig , 1997 Mean variance Theory : Lecture notes Washington University in Saint Louis

¹⁶ Philip H. Dybvig, 1997. Investments, Washington University in Saint Louis

¹⁷ Philip H. Dybvig, 1997. Mean variance Theory : Lecture notes Washington University in Saint Louis

¹⁸ <http://www.his.se/ish/kurser/Kursplaner/Ht02/Nek/Overhead01.doc> , Högskolan i Skövde (2002-12-15)

¹⁹ Carlson School of Management University of Minnesota , 2002

$$E(r_p) = \sum_{i=1}^N w_i * E(r_i) \quad \text{With the constrain} \quad \sum w_i = 1 \quad (3.1)$$

$E(r_p)$ = Expected return of the portfolio

w_i = The proportion (weight) of the portfolio invested in asset i

$E(r_i)$ = Expected return of asset i

N = The number of assets in the portfolio

Standard deviation (or variance) is used as a measure of risk²⁰.

The variance of a portfolio is given by:

$$\Omega_p = \left[\sum_{i=1}^N \sum_{j=1}^N w_i w_j \sigma_{ij} \right] \quad (3.2)$$

Where σ_{ij} = the covariance of returns between assets i and j .

If preferred the variance can be expressed by the correlation coefficient instead of the covariance, since:

$$\sigma_{ij} = \sigma_i \rho_{ij} \sigma_j \quad \text{Where} \quad -1 \leq \rho_{ij} \leq 1$$

This gives the following expression:

$$\Omega_p = \left[\sum_{i=1}^N \sum_{j=1}^N w_i w_j \rho_{ij} \sigma_i \sigma_j \right] \quad (3.3)$$

Naturally the standard deviation of the portfolio is then:

$$\sigma_p = \left[\sum_{i=1}^N \sum_{j=1}^N w_i w_j \sigma_{ij} \right]^{\frac{1}{2}} \quad \text{or if preferred} \quad (3.4)$$

$$\sigma_p = \left[\sum_{i=1}^N \sum_{j=1}^N w_i w_j \rho_{ij} \sigma_i \sigma_j \right]^{\frac{1}{2}} \quad (3.5)$$

In vector form we can create a vector of N individual weights w_i :

$$w = [w_1, w_2, \dots, w_N] \quad \text{With the constraint} \quad w^T 1 = 1$$

²⁰ T. Guhr Econophysics, 2001, Lecture notes, 2002

²¹ Ibid.

Each weight w_i correspond to the part invested in the individual asset, i , and 1 is defined as a unity matrix with the dimension $N \times 1$.

We then define the vector of the expected returns as

$$\mu = [E[r_1], E[r_2], \dots, E[r_N]]^T$$

Furthermore we define the covariance matrix Σ for the returns, where

$$\sigma_{ii} = \sigma_i^2 = \text{var}(r_i) \text{ and } \sigma_{kl} = \text{cov}(r_k, r_l)^{22}.$$

The expected return $E[r_p]$ and the variance $\text{Var}[r_p]$ of the portfolio can then be expressed as

$$\begin{aligned} E[r_p] &= w^T \mu \\ \text{Var}[r_p] &= w^T \Sigma w \end{aligned} \tag{3.6}$$

With the constraint $w^T 1 = 1$ (budget constraint)

The objective is to minimize variance at a given expected return under the constraint $w^T 1 = 1$.

By imposing constraints e.g. eliminating short selling or bounds the optimization problem is given by

$$\text{Min}_w \left\{ \frac{1}{2} w^T \Sigma w \mid w^T 1 = 1, w \geq 0 \right\} \tag{3.7}$$

Allowing unlimited short selling and excluding risk free assets, we can write the simplest mean-variance optimization problem as

$$\text{Min}_w \left\{ \frac{1}{2} w^T \Sigma w - \gamma w^T \mu \mid w^T 1 = 1 \right\} \tag{3.8}$$

Where γ is defined as the investors risk tolerance parameter (relates the risk to the return). The solution of the equation above will result in the equation for the optimal portfolio, which describes the portfolios placed on the efficient front.

$$w^{OPT} = \frac{\Sigma^{-1} 1}{1^T \Sigma^{-1} 1} + \gamma \left(\Sigma^{-1} \left(\mu - \frac{1 1^T \Sigma^{-1} \mu}{1^T \Sigma^{-1} 1} \right) \right) \tag{3.9}$$

²² Bengtsson C. & Holst J, 2002, Two Shrinkage Based Covariance Matrix Estimators for Improved Efficiency of Mean Variance Optimization

Or simplified

$$w^{OPT} = \frac{\Sigma^{-1}\mathbf{1}}{\beta} + \gamma \left(\Sigma^{-1} \left(\mu - \frac{\mathbf{1}\alpha}{\beta} \right) \right) \quad (3.10)$$

Where $\alpha = \mathbf{1}^T \Sigma^{-1} \mu$ and $\beta = \mathbf{1}^T \Sigma^{-1} \mathbf{1}$

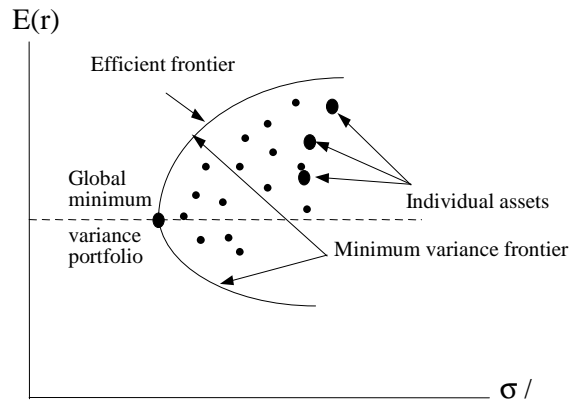


Figure 3.1.1 Source: Finance, College of Business & Information Systems Dakota State University (2003).

By increasing the value of γ one will follow the efficient front upwards. Choosing $\gamma = 0$ will result in the expression for the global minimum variance portfolio:

$$w_{GMVP} = \frac{\Sigma^{-1}\mathbf{1}}{\beta} = \frac{\Sigma^{-1}\mathbf{1}}{\mathbf{1}^T \Sigma^{-1} \mathbf{1}} \quad (3.11)$$

This is the portfolio with the smallest risk. It is also the only portfolio that does not need an estimate of the expected return and the only portfolio that rely solely on the covariance matrix Σ .²³ It is therefore useful as a “test instrument” for evaluation of methods that reduces noise and estimation errors in covariance matrices.

Finally it is worth mentioning that there exists criticism for using variance as a measure of risk, in the literature one can find two major sources of criticism for the use of variance as a measure of risk in the mean variance theory. First of all the variance measures both the positive and negative deviations (volatility) around the expected return. Since the positive deviations in reality are good for the investors, many scientists claim that only negative deviations should be taken into account when calculating the risk e.g. using downside risk²⁴, though on the other hand if investors are using short selling (therefore has an inversed risk preference) it makes sense to use an measure that includes both negative and positive volatility.

²³ Bengtsson C. & Holst J, 2002, Two Shrinkage Based Covariance Matrix Estimators for Improved Efficiency of Mean Variance Optimization.

²⁴ Fabozzi, 1999, Investment management, 2nd ed., Prentice-Hall Inc.

The second source of criticism for the use of variance as a measure of risk deals with the fact that the variance measures how the return varies around the expected return. The assumption that the “return risk” can be summarized by the variance (e.g. the return is normally distributed) is not fully correct, this since the return does not vary around the expected return in a symmetric way (normally distributed) therefore one should add a statistical measure of skewness²⁵.

3.2 Diversification

Diversification (according to Harry M Markowitz) means a strategy that seeks to combine in a portfolio assets with returns that are less than perfectly positively correlated, in an effort to lower portfolio risk (variance) without sacrificing the return²⁶.

Correlation is a familiar measure of association (strength) of the relationship between two variables e.g. assets. The correlation normally varies from 0 (random relationship) to 1 (perfect linear relationship) or -1 (perfect negative linear relationship)²⁷.

For assets (e.g. stocks) one might express the return r_i as a combination of a mean return, a random part of market noise and finally a part of an idiosyncratic risk to the specific firm²⁸:

$$r_i = \text{mean}_i + \beta_i z_m + e_i \quad (3.12)$$

Where mean_i and β_i are constants and the market noise terms z_m and idiosyncratic terms (specific to a certain firm) e_i terms are all uncorrelated (have zero covariance). One should note that the assumption that the idiosyncratic risk is uncorrelated among and across assets is not really true but still makes a fairly good estimate. The variance of a single asset can then be defined as

$$\beta_i^2 \text{var}(z_m) + \text{var}(e_i) \quad (3.13)$$

And for the variance of n companies (if investing the same amount in all companies) one will get the expression

$$\left(\frac{1}{n} \sum_i \beta_i \right)^2 \text{var}(z_m) + \frac{1}{n} \left(\frac{1}{n} \sum_i \text{var}(e_i) \right) \quad (3.14)$$

The risk is divided into two parts, one part of the risk which belongs to the firm, a firm specific risk (idiosyncratic risk) and one part belonging to a “market” risk (systematic). The idiosyncratic risk can be reduced by e.g. not investing in a certain

²⁵ Fabozzi F. J. 1999, Investment management, 2nd ed., Prentice-Hall Inc.

²⁶ www.trading-glossary.com , 2002.

²⁷ <http://www2.chass.ncsu.edu/garson/pa765/correl.htm>, 2002

²⁸ Philip H. Dybvig, 1997. Investments, Washington University in Saint Louis

“risky” company or using diversification strategy; the market risk effects all companies in the market and can not be avoided. For diversified portfolios containing many companies (e.g. containing approximately over 20 firm specific stocks) one can significantly reduce the idiosyncratic risk²⁹ (se graphs below).

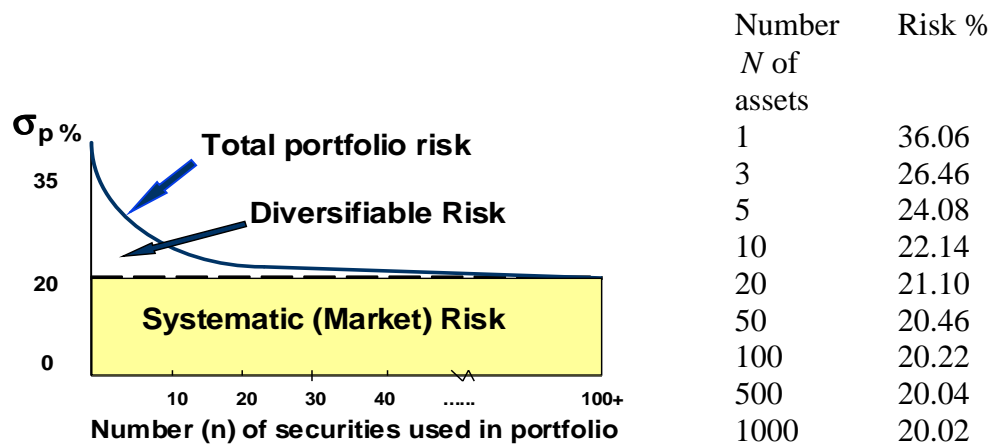


Figure 3.2.1 Risk calculated using $\beta_i=1$, idiosyncratic risk standard deviation = 0.3 and market std deviation = 0.2. In the diagram above one can see that an increased number of assets indeed reduce idiosyncratic risk, especially if one only invests in few assets. (Computation Andersson & Öberg based upon P. Dybvig, Investments, 2001)

Furthermore, by applying diversification on an international basis and not only a national basis, further reduction of risk may sometimes be achieved³⁰ The major advantage of diversification is the possibility to reduce risk without sacrificing the expected return, since applying diversification does not sacrifice the return as might often be the common belief³¹. One problem though is the fact that very few assets have a negative correlation with each other (which is desirable for the investor) which lowers the risk³². Therefore it might also be slightly difficult to find just such assets in a real world situation. Since knowing the true correlations between assets is crucial for the diversification process, one might understand why it is so important to reduce and eliminate the noise that hides the real correlations so that the investors might view the true information. Techniques such as Power Mapping and RMT-filtering reduce noise and help the investor understand the real correlations, which means that his or her risk hopefully can be reduced.

²⁹ Brealy R. & Myers S., 1996, Principles of Corporate Finance, 5th edition.

³⁰ Solnik, Bruno, 1988, International investments

³¹ Seligson & Co, 2002, Risk och Portföljteori. www.seligson.fi

²⁹ Ibid.

3.3 RMT-filtering³³ - a noise reducing method for correlation matrices

Random Matrix Theory based filtering³⁴ is a fairly new technique derived from the field of physics/econophysics and based upon the Random Matrix Theory³⁵. The main idea is to separate estimation errors from the “true and real” correlations in a specific matrix in order to extract the useful and valid information. This can be achieved by comparing a “sample” matrix based upon empirical data with a well known random matrix. This technique has some resemblances with another mathematical technique, namely Principal component Analysis (PCA)³⁶. Although PCA is a rather more refined and delicate mathematical tool.

First the “sample” correlation matrix is created. The process of creating a correlation matrix is fairly simple and straight forward.

Returns for $k = 1, \dots, N$ assets with the individual price S_k over a short period of time dt can be expressed according to³⁷

$$G_k(t) = \ln S_k(t + dt) - \ln S_k(t) \quad (3.15)$$

By computing the logarithms of the price one can somewhat suppress and reduce trends that rely solely on the overall market and seasons. One should notice that this basic technique should be combined with more sophisticated techniques when dealing with certain types of data (e.g. high-frequency data).³⁸

The expression above can easily be rewritten mathematically (if we use a short period of time dt)³⁹

$$G_k(t) \rightarrow \ln S_k(t + dt) - \ln S_k(t) = \ln \frac{S_k(t + dt)}{S_k(t)} \quad (3.16)$$

Since the volatility (standard deviation) concerning the price of the individual assets differs from asset to asset, normalization will be necessary for the comparison between the return of individual assets. This is achieved by the following operation:

$$M_k = \frac{G_k - \langle G_k \rangle}{\sqrt{\langle G_k^2 \rangle - \langle G_k \rangle^2}} \quad (3.17)$$

³³ Note: In previous studies RMT-filtering is often only referred to a “filtering” we chose to use a somewhat more specific name to avoid confusion.

³⁴ B. Rosenow, Vasiliki Plerou, Parameswaran Gopikrishnan, H. Eugene Stanley. Jan 7th, 2002, Portfolio Optimization and the Random Magnet Problem.

³⁵ Åberg S, 2002, Quantum Chaos, Lund University

³⁶ See Appendix for a brief introduction.

³⁷ Thomas Guhr, Econophysics: Lecture notes.

³⁸ Thomas Guhr, Personal interview conducted 2002-11-17

³⁹ Persson, A. & Böiers, L-C, Analys i en variabel, 2000

The method of subtracting the mean $\langle G_k \rangle$ and division by the standard deviation given by $\sigma_k = \sqrt{\langle G_k^2 \rangle - \langle G_k \rangle^2}$ is of course a well known form of a normalization process in the field of Statistics⁴⁰.

An equal time cross correlation matrix, C , with the elements $C_{kl} = \langle M_k(t)M_j(t) \rangle_T$ can then be constructed (The elements range between +1 and -1 by construction).

There are some important main issues⁴¹ associated with the task of analyzing the meaning and importance of the cross correlation coefficients C_{kl} (and the correlation matrix C).

- First of all the conditions on the (stock) market vary with time and correlations between assets may cease to exist at any given time. One could say that cross correlations have a tendency to be non-stationary i.e. they are not constant in time.

- Second, by using short time series (finite length) to estimate cross correlations measurement noise is introduced.

Due to the two issues mentioned above which are obviously not compliant, the empirically measured cross correlations will contain random contributions⁴². The problem is now to find the cross correlations which are not associated with randomness. This is made by comparing the “sample matrix” with a well known random matrix, by doing so one might determine and separate the parts of the “sample” matrix that conforms to the randomness (noise) from the deviating part that contains information.⁴³

Construction of the “sample” matrix C can be done accordingly to⁴⁴:

$$C = \frac{1}{L} MM^T \quad (3.18)$$

Where M is a $N \times L$ matrix, with the elements $\{M_{im} \equiv M_i(m\Delta t); i = 1, \dots, N; m = 0, \dots, L-1\}$

A complete random matrix C_{RM} is then constructed accordingly to:

$$C_{RM} = \frac{1}{L} AA^T \quad (3.19)$$

Where A is a $N \times L$ matrix containing N time series of L mutually uncorrelated (computer generated) random elements with zero mean and unit variance.

⁴⁰ Körner S, Wahlgren, Lund 2000, Statistisk dataanalys, tredje upplagan,

⁴¹ B. Rosenow, Vasiliki Plerou, Parameswaran Gopikrishnan, H. Eugene Stanley. Jan 7th, 2002, Portfolio Optimization and the Random Magnet Problem

⁴² Ibid.

⁴³ Ibid.

⁴⁴ Ibid.

The similarities between the matrices C and C_{RM} above are in some sense obvious, and one might understand the idea of comparing them with each other in order to create an understanding of the properties of the correlation.

The statistical properties of the matrix C_{RM} are well known⁴⁵, these kinds of matrices are commonly referred to as Wishart matrices⁴⁶. It can theoretically be shown that when

$N \rightarrow \infty$, $L \rightarrow \infty$ and when the expression $Q \equiv \frac{L}{N}$ becomes fixed, the distribution of the eigenvalues $P_{RM}(\lambda)$ of the random correlation matrix C_{RM} can be described by the expression:

$$P_{RM}(\lambda) = \frac{Q}{2\pi} \frac{\sqrt{(\lambda_+ - \lambda)(\lambda - \lambda_-)}}{\lambda} \quad (3.20)$$

Where $Q = \frac{L}{N}$, and λ ranging between the smallest and largest eigenvalue(s) ($\lambda_- \leq \lambda \leq \lambda_+$).

The boundaries of the smallest and the largest eigenvalue in the distribution is then given by

$$\lambda_{\pm} = 1 + \frac{1}{Q} \pm 2\sqrt{\frac{1}{Q}} \quad \text{With } Q = \frac{L}{N} \quad (3.21)$$

Recent studies applying RMT methods have shown that approximately 98 % of the eigenvalues of the correlation matrix type C agree and correspond to RMT predictions⁴⁷. The remaining 2 % of the eigenvalues represent deviations from the RMT predictions; these eigenvalues also represent the 2 % of the largest eigenvalues. The fact that only 2 % of the eigenvalues deviate indicates a very high level of randomness in the measured cross correlations⁴⁸. One way to interpret this is that only the largest eigenvalues outside the bulk (given by the boundary equation above) contain non random information.

Plerou et al test the RMT approach on real U.S. stock data by comparing the eigenvalue distribution of the matrix C with the distribution of the random matrix C_{RM} . The conclusion is that approximately 2% of the largest eigenvalues deviates strongly while the remaining 98 % conform to the bulk. This is also a

45 B. Rosenow, Vasiliki Plerou, Parameswaran Gopikrishnan, H. Eugene Stanley. Jan 7th, 2002, Portfolio Optimization and the Random Magnet Problem.

46 Muirhead R, 1982, Aspects of Multivariate Statistical Theory, New York, Wiley, cop.

47 Vasiliki Plerou et al, 22 February 1999. Universal and Non universal Properties of Cross Correlations in Financial Time Series

48 B. Rosenow, Vasiliki Plerou, Parameswaran Gopikrishnan, H. Eugene Stanley. Jan 7th, 2002, Portfolio Optimization and the Random Magnet Problem.

result that is consistent with the results of an earlier study performed by Laloux et al⁴⁹.

Furthermore some conclusions regarding the largest deviating eigenvalues can be drawn and summarized from this and earlier studies⁵⁰.

- The largest eigenvalue of the matrix C represents the influence of the entire market. This eigenvalue is significantly larger than the other eigenvalues
- The remaining deviating eigenvalues show mainly correlations between stocks of similar industries, stocks belonging to companies located in the same geographical area and stocks having large market capitalization.
- The largest deviating eigenvalue(s) are stable for a reasonable long period of time, the smaller the deviating eigenvalues get the more unstable they become.

Plerou et al. continues by providing a thorough examination where they conclude that there is nothing talking against the main idea and statement that the true correlations (information) can only be found in the largest deviating eigenvalues.

This finally provides the conclusion to why the RMT filtering approach (that uses only the largest deviating eigenvalues when creating covariance matrices) might be valuable e.g. for using only the eigenvalues containing the true correlations (information) and excluding those eigenvalues that conform to randomness “noise”.

⁴⁹ Laloux L. et al, 1998, Noise dressing of financial correlation matrices.

⁵⁰ Ibid.

3.4 Overview of the correlation matrices

The figure below presents the general look of an idealized correlation matrix. We will always have a diagonal with ones. There will be a quadratic structure with higher values representing a branch. Outside of the branches there should be smaller values, since these companies are in no branch and should have small values for correlations to other companies. In reality though the values outside a branch might have high values. The complete correlation matrix will also be distorted by noise, especially if the time steps are long (as in monthly data).

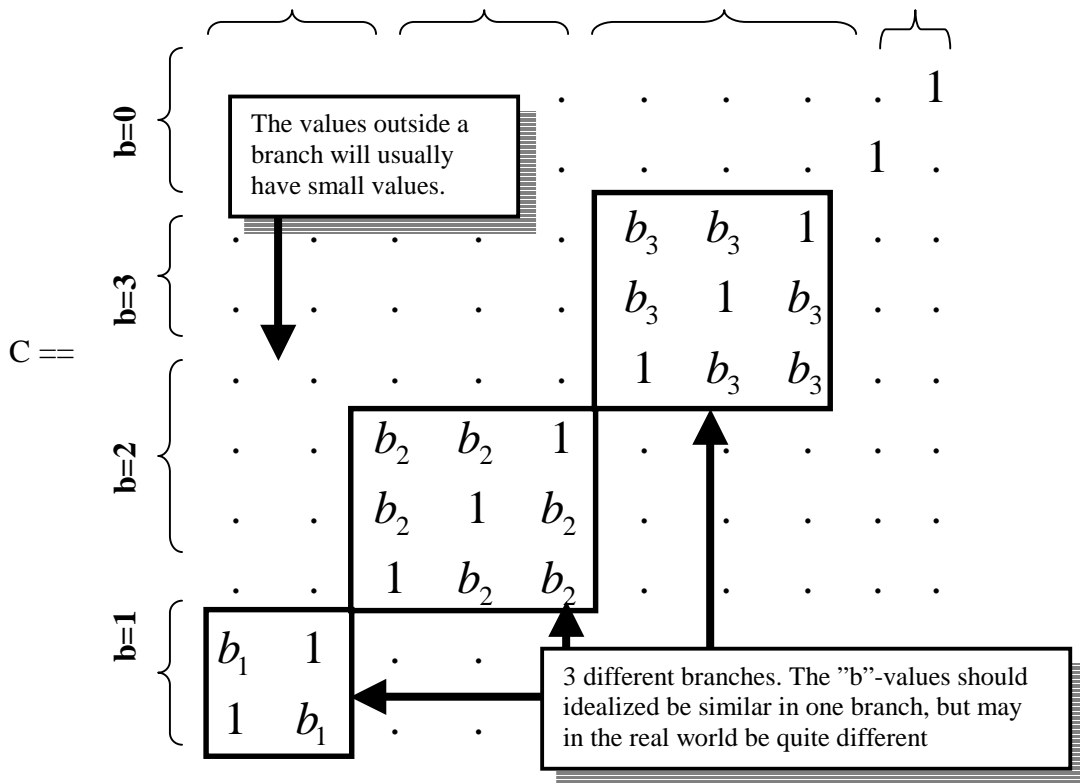


Figure 3.3 An idealized correlation matrix. If the correlation matrix is for real stocks few values will be zero, the p value will also mostly be positive ones and they will always be different values (since the branch structure will never be perfect).

The figure above should always be in our thoughts when we study the pictures of the correlation matrices. Block structures like shown above are always interesting, since they represent industries where the companies react very similar to each other.

An example of a correlation matrix from real companies and descriptions on how to interpret the colours is presented below. As can be seen in the picture the vertical stripe is in another direction (Matlab transposes the matrices for plotting), but a somewhat similar structure is still visible.

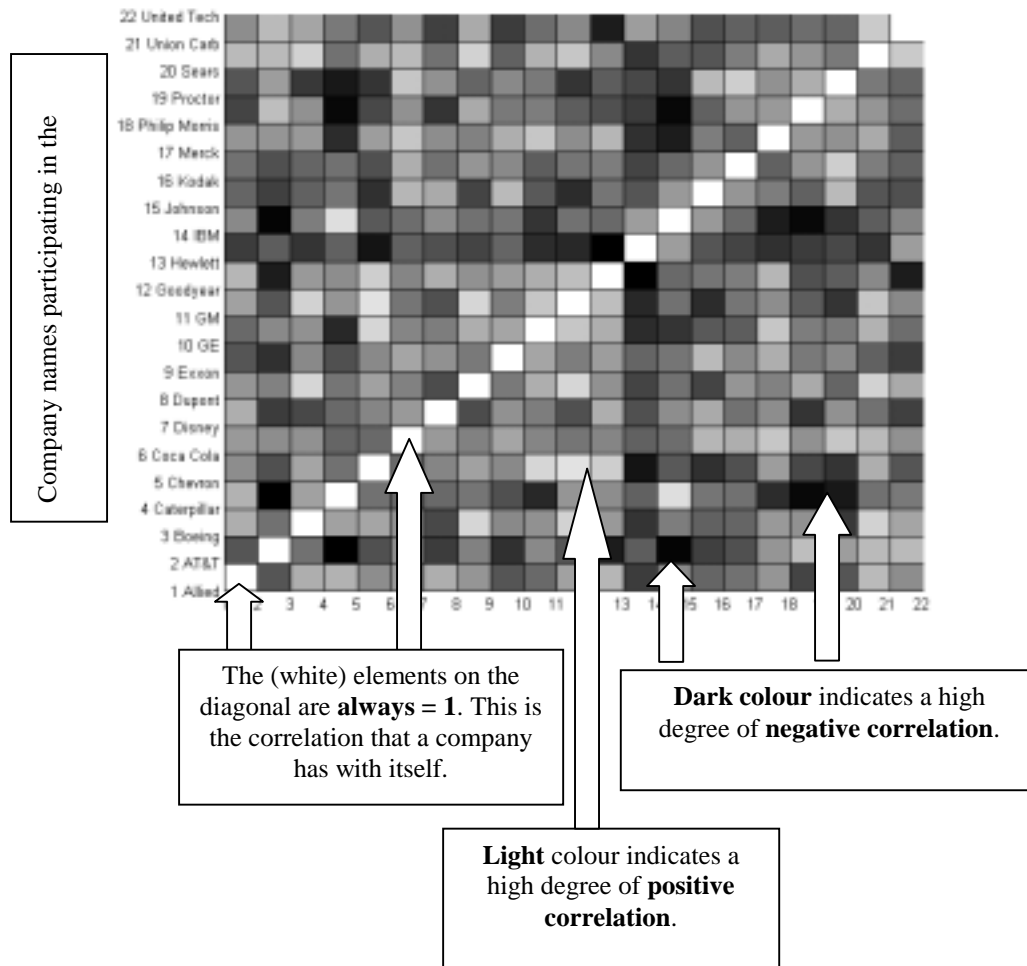


Figure 3.4.2 Explanation of color mapping of correlation matrices

3.4 Noh's model – a model to simulate stock markets

In the following chapter 3.5 we discuss another noise reducing model called Power Mapping. But before we discuss the nature of Power Mapping base theory is established with Noh's model, a market simulating model that we use in the explanation of the concept and idea behind the Power Mapping technique.

3.4.1 Collected versus simulated stock data

To acquiring large amounts of high-frequent stock data for empirical examination can be a time consuming and expensive process that introduces several theoretical and practical problems. The extensive amount of data collected and used in the

study by Gopikrishnan et al⁵¹ took several years to collect and prepare for the analysis. Collected stock data also presents some problems e.g. auto correlation and missing data for companies at certain time periods. To avoid these problems simulating models can be used to create a flexible environment that open up the possibility to do in-depth mathematical analysis of more complex data structures.

3.4.2 An introduction to Noh's model

A short introduction is presented of a model created by Noh⁵²; in chapter 3.4.3-3.4.6 the mathematical aspects of Noh's method are discussed.

Noh's model⁵³ simulates a correlation matrix that is similar to the correlation matrix one would get if a real stock portfolio is studied.

Figure 3.4 illustrates two correlation matrices created by Noh's method. The large squares represent different branches; here the correlation is higher since companies in the same branch are assumed to react fairly similar to each other. The diagonal is the correlation between a company and itself, thus it should always be unity (1). The top-right corner shows an area that doesn't have a square; here the correlation coefficients for the companies in no branch are located.

The two pictures below also show a noise like pattern, this is simply the random noise introduced by Noh's model.

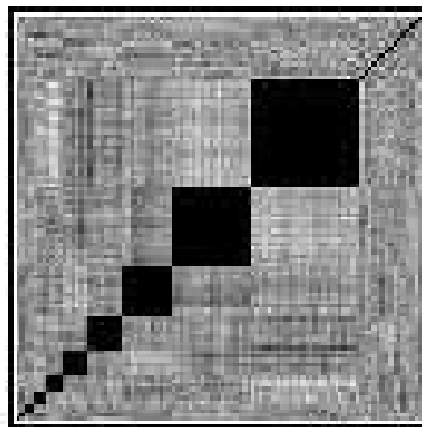


Figure 3.4.1. A correlation matrix created with Noh's method. There are 9 branches in this market. Thomas Guhr & Bernd Kälber, 28 Jun 2002, arXiv: cond-mat/0206577 v1 p. 15

3.4.3 Time series in Noh's model

The model suggested by Noh is from an economic standpoint a one-factor type (and can be interpreted as an application of the arbitrage pricing model⁵⁴), while from a

⁵¹ Vasiliki Plerou, Parameswaran Gopikrishnan et al. July 31, 2001. A Random Matrix Approach to Cross-Correlations in Financial Data.

⁵² J.D. Noh, Phys. 2000. Rev. E61, 5981, cond-mat/9912076

⁵³ Thomas Guhr & Bernd Kälber, 28 Jun 2002, arXiv:cond-mat/0206577 v1 pp. 1-9

physic's viewpoint Noh's model has several important similarities with models involving Pott's spins^{55, 56}.

The market that Noh's model simulates has K companies, called $k = 1, \dots, K$. Different branches are created labeled $b = 1, \dots, B$. Within one branch the companies are assumed to be correlated, as we shall see later this will create square "blocks" of identical correlations (these values will in a later stage be offset by an introduced noise in the correlation matrices. The indices k of the companies within the same branch are ordered in such a way so that they follow each other. The last "block" is supposed to consist of companies in no branch; the correlations for these companies are close to zero.

The branch index, b , is a function of the company index; $b = b(k)$ The companies which are in no branch, $b = b(k) = 0$. κ_b is the size of the industrial branch b . We have:

$$\sum_{b=1}^B \kappa_b + \kappa = K \quad (3.22)$$

We assume that $\kappa_b > 1$. κ can be any non-negative integer, including zero.

$M_k(t)$, $k = 1, \dots, K$, the normalized time series of the returns for the K companies are modeled as the sum of two purely random contributions⁵⁷. The first contribution models the correlations within a given branch, involving random numbers $\eta_b(t)$, the second contribution is specific for the company and involves random numbers $\varepsilon_k(t)$. $\eta_b(t)$ and $\varepsilon_k(t)$ are assumed to be uncorrelated and standard normal distributed with zero mean value.

$$M_k(t) = \sqrt{\frac{p_{b(k)}}{1 + p_{b(k)}}} \eta_{b(k)}(t) + \frac{1}{\sqrt{1 + p_{b(k)}}} \varepsilon_k(t) \quad (3.23)$$

Contribution 1 and contribution 2 are weighted with the parameter $p_{b(k)}$, a parameter that is common to all companies in the branch b . The weights are also assumed to be positive $p_{b(k)} \geq 0$. When k is in no branch (i.e. for $b = 0$) $p_{b(k)} = 0$. Discrete time steps and normalized time units such that $dt = 1$. $M_k(t)$ consist of T time values at $t = 1, \dots, T$

⁵⁴ S. Ross & J. Econ. 1976. Theory 13, 341

⁵⁵ M. Marsili, cond-mat/0003241 referred to in Thomas Guhr & Bernd Kälber, 28 Jun 2002, arXiv:cond-mat/0206577 v1 p. 3

⁵⁶ L. Kullmann, J. Kertesz and R.N. Mantegna, cond-mat 0002238 referred to in Thomas Guhr & Bernd Kälber, 28 Jun 2002, arXiv:cond-mat/0206577 v1 p. 3

⁵⁷ Thomas Guhr & Bernd Kälber, 28 Jun 2002, arXiv:cond-mat/0206577 v1 p. 4

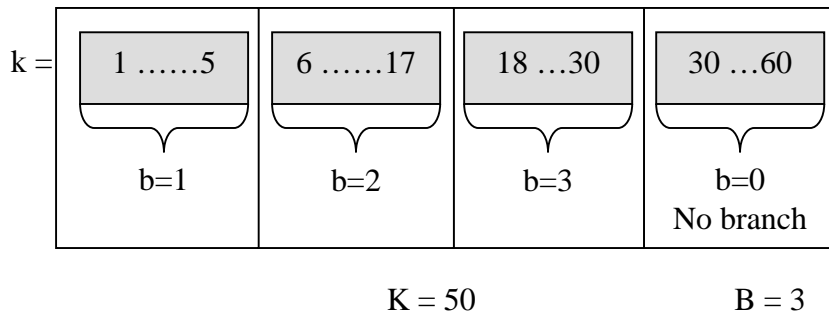


Figure 3.4.2 The branches b and companies k

The example above demonstrates how b is generated from the chosen k values. A market with 50 companies in 3 different branches is studied.

3.4.5 Noise introduced by Noh's model

The finite numbers from the averages above describe the noise dressing⁵⁸ of the financial correlation matrices essential to Noh's model. The always present finiteness of the time series introduces a random offset to every correlation coefficient; the true correlation coefficients are effectively buried under noise. The correlation matrix is then:

$$C = C_{true} + C_{noise} \quad (3.24)$$

3.4.6 Spectral densities of the correlation matrices

As noted by Guhr⁵⁹ spectral density is a “well suited observable to study the noise dressing of empirical correlation matrices”. The spectral density is actually a histogram of eigenvalues of the correlation matrix. The spectral density can present a fairly easy to understand graphical picture of how much noise and useful information the correlation matrix can give us. It also shows “exactly” what the Power Mapping actually does to the information in the correlation matrix. Since the mathematical nature of the spectral densities is fairly complex we do not go into theoretical detail, instead we present a brief overview discussing the aspects of spectral densities and the correlation matrices⁶⁰.

When $C(T)$ is simulated for different length $T = 1650, 5000, 2000$ and 50000 of the time series and one calculates the eigenvalues $\lambda_k, k=1, \dots, K$ and calculate the spectral densities $\rho_T(\lambda)$ we achieve the resulting figure 3.8

⁵⁸ Thomas Guhr & Bernd Kälber, 28 Jun 2002, arXiv:cond-mat/0206577 v1 p. 5

⁵⁹ Ibid.

⁶⁰ See Thomas Guhr & Bernd Kälber, 28 Jun 2002, arXiv:cond-mat/0206577 v1 for further details of the more mathematical aspects.

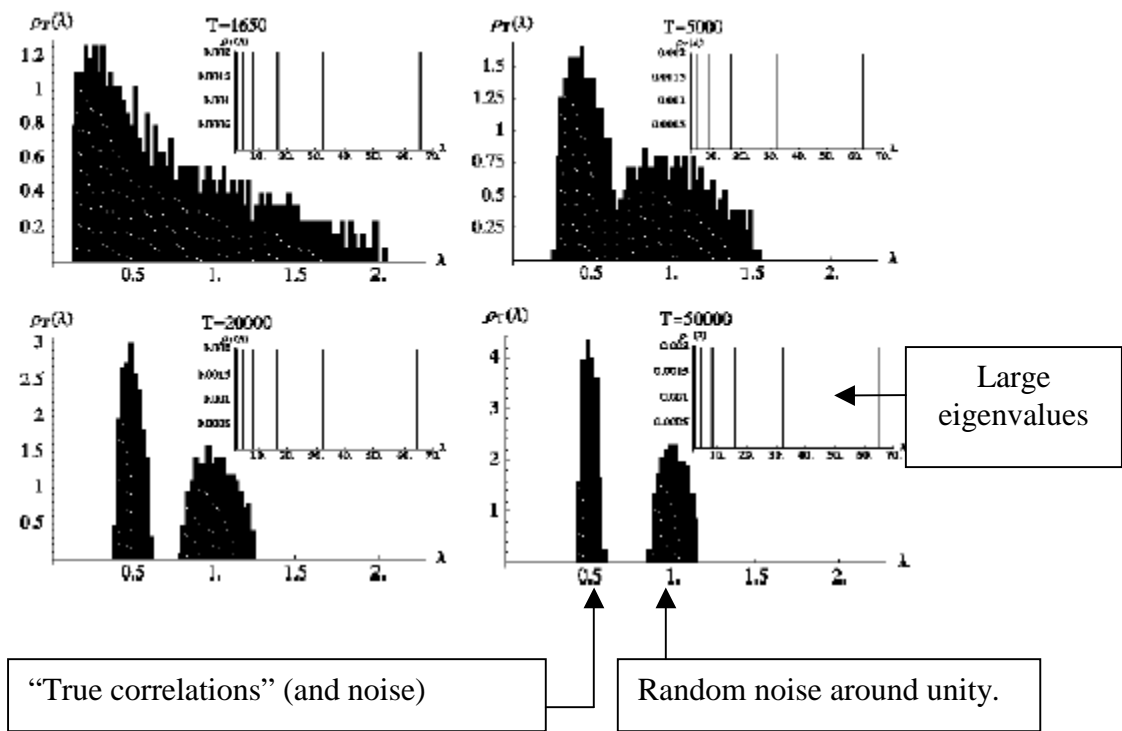


Figure 3.4.3 Spectral densities of the simulated correlation matrices. The presentation is split into regions $0 \leq \lambda \leq 2.2$ and $2.2 \leq \lambda \leq 70$. The densities are given in units of K. Thomas Guhr & Bernd Kälber, 28 Jun 2002, arXiv:cond-mat/0206577 v1 p. 7. Edited version.

Observe the large eigenvalues in the upper-right corner. When T increases the isolated, large peaks are stable; these large peaks are very interesting since it has been shown that they correspond to the individual branches⁶¹. The left peak is centred on $\lambda=1/2$ and the right peak on $\lambda=1$. This left peak can be associated with the true correlations while the right peak only presents noise (the noise is around the trivial self-correlation of the companies for $k = 1$).

The spectral density is given by the formula presented below.

$$\rho_{\infty}(\lambda) = \sum_{b=1}^B (\kappa_b - 1) \delta\left(\lambda - \frac{1}{1 + p_b}\right) + \sum_{b=1}^B \delta\left(\lambda - \frac{1 + \kappa_b p_b}{1 + p_b}\right) + \kappa \delta(\lambda - 1)$$

True correlations

↑

Large eigenvalues

↑

Noise ($\lambda \approx 1$)

↑

Figure 3.4.4 The spectral density.

⁶¹ Thomas Guhr & Bernd Kälber, 28 June 2002, arXiv:cond-mat/0206577 v1 p. 5

As noted by the text boxes below the formula, the first term represents the true correlations, the second term the large eigenvalues and the last term is the noise.

An important aspect of the formula above is that the density is smeared out for finite time series⁶².

3.5 Power Mapping – a noise reducing method for correlation matrices

Power Mapping⁶³ is a new technique available for noise reduction in matrices. The method was first presented by T. Guhr and B. Kälber in June 2002.

The main idea of the method is to “artificially” prolong time series in certain aspects by applying a technique called Power Mapping. This prolongation would help the ability to identify and estimate the noise. Time series that are long enough make the true correlations that otherwise are buried under noise visible in the spectral density⁶⁴. The Power Mapping technique suppresses the matrix elements that can be associated with noise. Prolonging time series is of course especially useful when it is difficult to obtain long time series e.g. during a limited window of time.

The Power Mapping is performed according to the operation

$$C_{kl}^{(q)}(T) = \text{sign}(C_{kl}(T))|C_{kl}(T)|^q \quad (3.25)$$

This means that the matrix $C(T)$ is “mapped” to the matrix $C^{(q)}(T)$ by preserving the sign of the matrix element C_{kl} and raising its modulus to a power of the magnitude q . Note that only the individual elements in the matrix and not the matrix as a whole is powered. Depending on the value of q one has the capability to suppress certain elements from the original matrix $C(T)$ in the matrix $C^{(q)}(T)$ ⁶⁵.

Part of the result in the study performed by Guhr et al. can be found presented in the graphs below. In the study by Guhr et al. correlation matrices for $K = 508$ companies were simulated from time series with the length $T = 1650$ and with specific sizes and weights⁶⁶. For improvement of statistical significance an ensemble of 25 matrices were simulated. These matrices were then powered with different q values. The graphs below show the average spectral densities for the chosen values of $q = 1, 1.5,$ and 2.0 . The densities in the graphs are given in the units K (relates to companies)

⁶² For more in depth analysis of the spectral densities see Thomas Guhr & Bernd Kälber, 28 June 2002, arXiv:cond-mat/0206577 v1

⁶³ Ibid.

⁶⁴ Ibid.

⁶⁵ Ibid.

⁶⁶ Ibid.

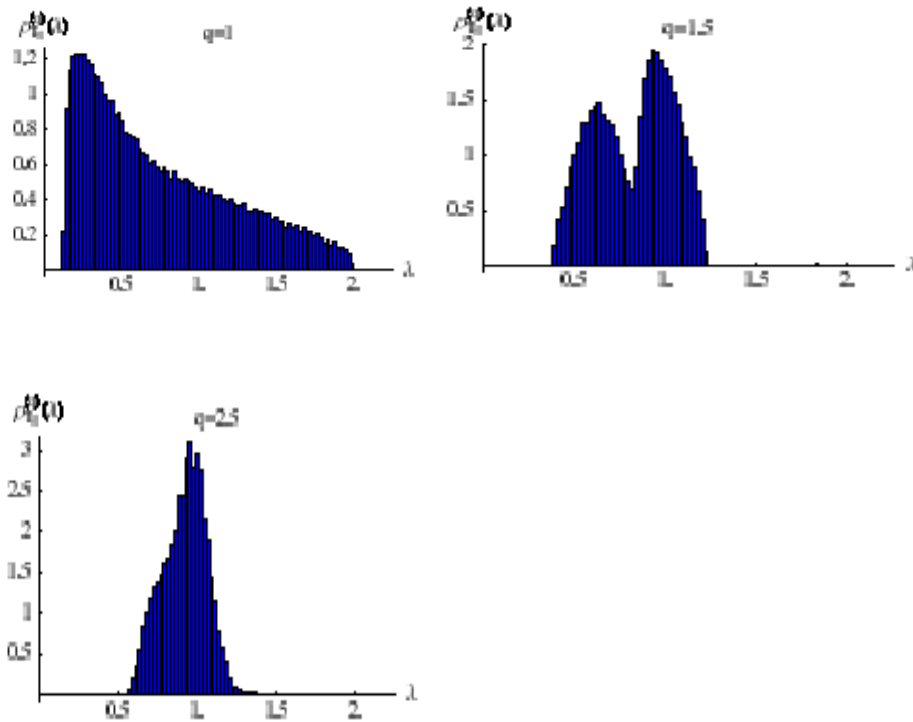


Figure 3.5.1. the average spectral densities for the chosen values of $q = 1, 1.5,$ and 2.5 . The densities in the graphs are given in the units K (relates to companies)

The q value of 1 corresponds to the original matrix (since the powering of 1 does not change the elements in any way). One can see two peaks emerge when $q = 1.5$ and then glue them together again as the value of q increases to $q = 2.5$. It can be shown that the graphs illustrate the separation of the densities into two peaks, one peak corresponding to the noise around the trivial self correlations of the companies (right in the example) and one peak (left in the example) corresponding to the true correlations associated with the branches (see also Chapter 3.4)⁶⁷. The two graphs below show the spectral densities for two other similar simulations where no Power Mapping has been used.

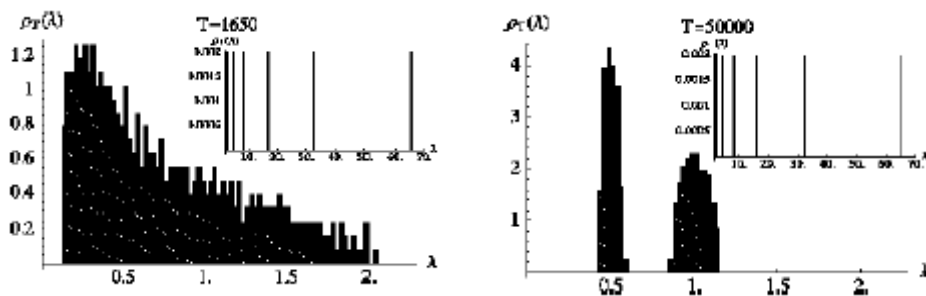


Figure 3.5.2 The spectral densities for $T = 1650$ and $T = 50000$

With the length of $T = 1650$ one cannot separate any peaks at all, at $T = 50000$ one can clearly see the two individual peaks corresponding to the noise around the trivial self correlations and true correlations. If one compares the simulation where $T = 1650$ and Power Mapping where $q = 1.5$ with the two graphs above one can

⁶⁷ Thomas Guhr & Bernd Kälber, 28 Jun 2002, arXiv:cond-mat/0206577 v1

clearly see the advantage of Power Mapping as a prolongation tool and noise reducer. Although the power mapped simulation only uses $T = 1650$ one can already start to see the separated peaks seen in the graph of densities calculated for $T = 50000$. This shows the potential of using Power Mapping in order to prolong time series.

From the graphs above one might suspect that the optimal value of q for maximum separation between the peaks is in the close vicinity of 1.5, this is a correct assumption that has been shown both mathematically and empirically by Guhr et al⁶⁸

Since the technique is fairly complex and mathematical the authors refers to the main article by Guhr et al. (2002).

⁶⁸ Thomas Guhr & Bernd Kälber, 28 Jun 2002, arXiv:cond-mat/0206577 v1

3.5 Earlier empirical findings and results

In a brief earlier study performed by Andersson & Öberg⁶⁹ in the fall of 2002 using only a very limited number of companies and using monthly stock data, the abilities of the power mapping and Random Matrix Theory filtering (denoted RMT-filtering below) techniques to create more “accurate” correlation matrices were briefly examined. A limited part of the results are included in this thesis to make comparison easier. Furthermore several graphical aspects of the Power Mapping and RMT-filtering were studied, since we are only presenting a limited amount of new studies in this area, but are referring to older results through the analysis these results are presented below. The idea has also been to produce a short summary of how the both noise reducing methods affect the Sample matrix visually. Also a new “hybrid method” was developed, in this thesis this method will be further investigated.

3.5.1 Reducing risk with better correlation matrices

The (hopefully) improved matrices from Power Mapping and RMT-filtering were used to create global minimum variance portfolios accordingly to the Markowitz’s Mean Variance Theory, these minimum variance portfolios were created (optimized accordingly to) with the sole purpose of minimizing risk, this is important to remember when interpreting the results presented in this section; this since the aspects of return are not taken into account.

The main results from the examination of global minimum portfolios conducted by Andersson & Öberg are presented in the picture below. The original correlation matrix is denoted as “Sample” and a equally weighted portfolio (i.e. equal weights of all assets) is denoted as “Decided”. “Power Mapping” and “RMT-filtering” both imply the correlation matrices which have undergone respectively noise reduction. RMT-filtering(1) only uses one eigenvalue to calculate the correlation matrix as this was proved to produce better results for the risk. RMT-filtering(2) uses the λ_{\max} to decide the number of eigenvalues that should be included.⁷⁰ PM-Filtering⁷¹ stands for “Power-Mapping-RMT-filtering” and is an experimental technique that uses both RMT-filtering and Power Mapping in a unique way that for at least the simple tests conducted by Andersson & Öberg produced the best results of all filtering techniques. PM-filtering(1) uses the three largest eigenvalues (by default) in the RMT-filtering process and PM-Filtering(2) uses an algorithm to decide how many eigenvalues that should be included in the calculation process. RMT-filtering(1) only uses one eigenvalue to calculate the correlation matrix as this was proved to produce better results for the risk. RMT-filtering(2) uses λ_{\max} to decide the number of eigenvalues that should be used.⁷²

⁶⁹ Andersson, P. Öberg, A. 2003. Noise Reducing Methods for Correlation Matrices New Techniques to Minimize Risk in Portfolio Management?

⁷⁰ See Chapter 2 for further details.

⁷¹ There are two versions of RMT-filtering, 1 and 2. When we later talk about the RMT-filtering we are discussing RMT-filtering1.

⁷² See Chapter 2 for further details.

Table 3.5.1 Showing the different risks for the 24 created portfolios. The standard deviations are annualized through multiplication by $\sqrt{12}$.

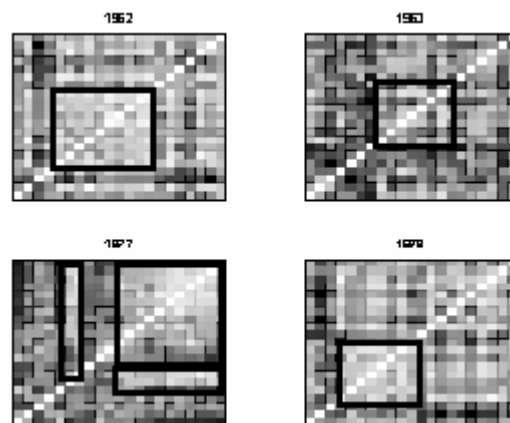
Name	Risk (in percent)	Diff. in Risk compared to Sample
Sample	14.6594	-
Power Mapping	13.8924	-5.2319 %
RMT-filtering(1)	14.4771	-1.2434 %
RMT-filtering(2)	14.6132	-0.31529 %
PM-Filtering(1)	13.7689	-6.0746 %
PM-Filtering(2)	13.8545	-5.4906 %
Decided	16.1969	+10.4884 %

The results above show that Power Mapping and the experimental PM-Filtering1 create the minimum variance portfolios that have the lowest risk. The Power Mapping is approximately 5.2% better than original “Sample” matrix. All RMT-filtering techniques also produce lower risk than the Sample matrix, though not as good results as Power Mapping and PM-filtering1.

3.5.2 Changing correlations over time

To estimate the degree of interdependence and co-movement between financial assets and asset classes in a correct manner is very important in portfolio and risk management, especially in times associated with uncertainty⁷³. This dependence is usually represented by the measure correlation. It is well known from earlier empirical studies that correlations between assets and markets tend to increase with the level of volatility⁷⁴ of the market(s), especially on an international level^{75,76}. This is often referred to as “correlation breakdown. In practical terms this means that the correlation between assets and groups of assets increases in market declines, which increases volatility and reduces opportunities for diversification⁷⁷; this since the possibilities for diversification decreases when the correlations increase⁷⁸. The scenario described above presents a major problem for investors using correlation matrices based upon historical data when correlations suddenly change, this since the usefulness of operations such as diversification and hedging is reduced when they are needed the most i.e. in times of high volatility and extreme negative price movements.⁷⁹ Naturally correlations not only change during extreme market events and trends, but they also change in a lesser extent on a normal basis. Exactly how these correlations and correlations structure vary and how stable in time they have been are currently under debate by scientists.

To briefly illustrate the effects of changing correlations, four correlation matrices have been generated using 22 major American companies for the years 1962, 1963, 1977 and 1978. In the figure below, when observing the correlation matrices for the years 1962 and 1963, one can clearly see that the patterns aren't very similar, but have changed. In the first correlation matrix there is a clear quadratic structure (similar to a structure generated by branches), but in 1963 there is no such structure. The matrices for the years 1977 and 1978 are



also different and have only limited resemblance. Figure 3.5.1. One-year correlation matrices

⁷³ Rubinstein, Mark, 2000, “Comments on the 1987 Stock Market Crash: Eleven Years Later, published in Risks in Accumulation Products, Society of Actuaries,

⁷⁴ Byström Hans NE, January 3, 2002. Cross-Sectional Correlation: New Evidence on Changing Correlations and Correlation Breakdown in Equity Markets

⁷⁵ Longin, F., Solnik, B., 1995. Is the correlation in international equity returns constant: 1960-1990, Journal of International Money and Finance 14, pp 3-26.

⁷⁶ Bookstaber, Richard, Spring 1997. Global risk management: are we missing the point?, Journal of Portfolio Management Vol. 23, No. 3.

⁷⁷ Rubinstein, Mark, 2000, “Comments on the 1987 Stock Market Crash: Eleven Years Later, published in Risks in Accumulation Products, Society of Actuaries.

⁷⁸ Byström Hans NE, January 3, 2002. Cross-Sectional Correlation: New Evidence on Changing Correlations and Correlation Breakdown in Equity Markets

⁷⁹ Rey, David Michael, 2000, Time-varying Stock Market Correlations and Correlation Breakdown. Finanzmarkt und Portfolio Management, Nr. 4, S. 387-412

3.5.3 Graphical presentation of improved correlation matrices

In the picture below one can see the effects of the Power Mapping and RMT-filtering techniques on the very correlation matrices themselves. The matrices are as following: Before and after the noise reduction operations, and the difference between the before and after matrices.

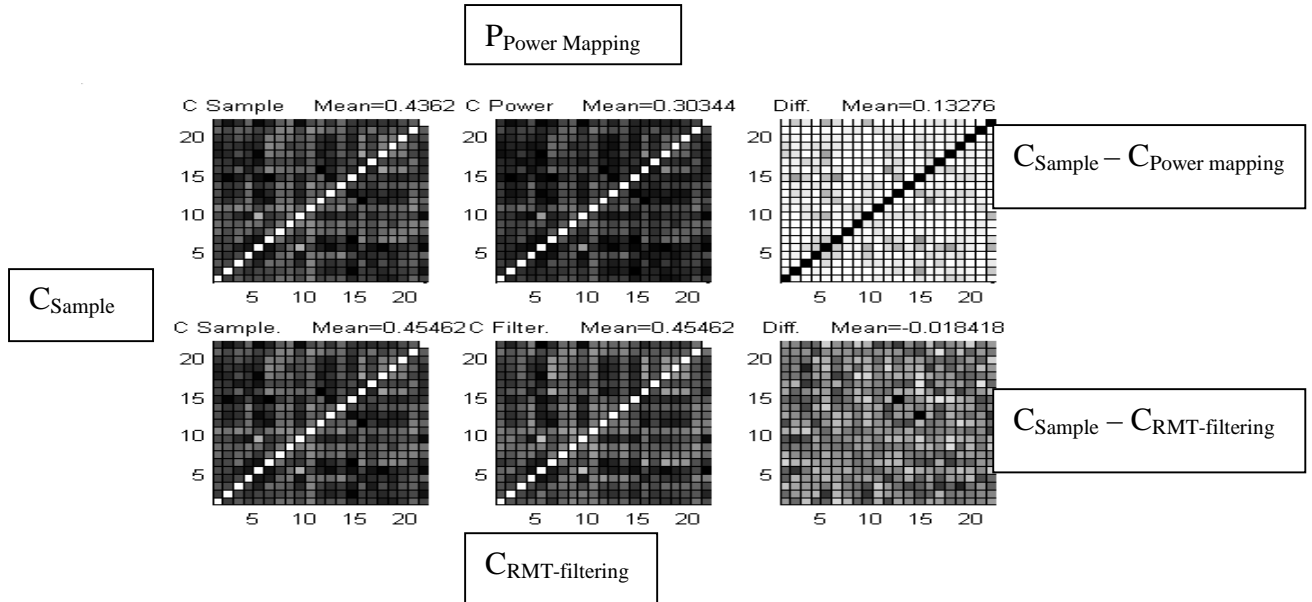


Figure 3.5.2. Calculated 10 year correlation matrices, before and after the noise reduction operations, as well as the difference before and after the operations.

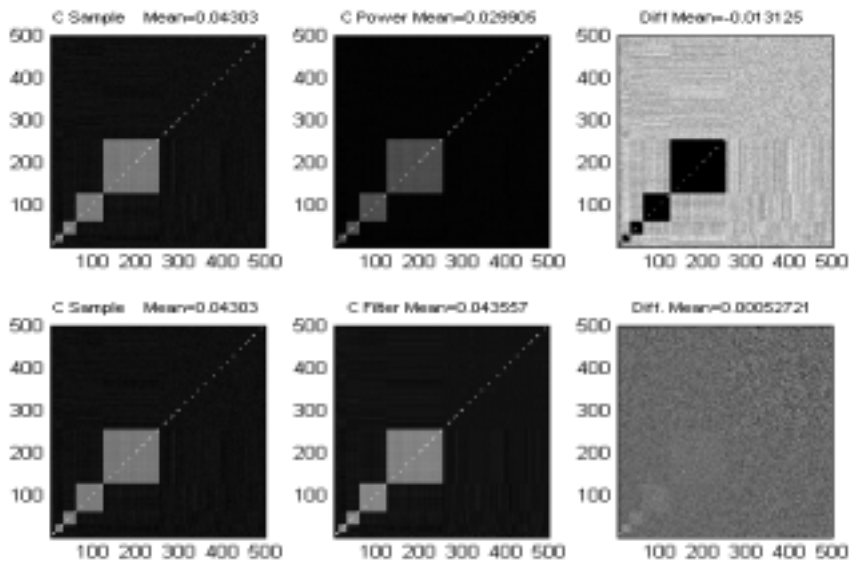


Figure 3.5.3. Correlation simulated by Noh's model. RMT-filtering shows an interesting effect with the lower companies. The smaller branches seem to be suppressed more

The correlation matrix for Sample looks the way we are used to see it. The RMT-filtering has a pattern that looks like faint crossed lines, this might be a problem with the correlation matrix; that these lines are covering the matrix in the RMT-filtering will probably not lead to an improved correlation matrix. The correlation

matrix of the Power Mapping shows no data at all in the areas outside the branches, but a random-like pattern is still visible in the branches. The picture is also noticeably darker than in the other pictures, a normal trait of the Power Mapping. As expected Power Mapping has the lowest mean value since the method suppresses the individual elements in the correlation matrix. One can also note that all the mean values are positive and also very small; this is mostly due to the fact that correlation matrices created by Noh's method have very weak correlations outside the branches.

The RMT-filtering creates perhaps the most interesting picture with new information; information that we cannot see in the previous analysis on the stock data. If one looks at the picture of the $C_{\text{RMT-Filter}}$ (the one in the middle at the bottom) one can see that the picture is a little lighter than the original matrix. An even more interesting pattern can be seen in picture of the difference of the correlation matrix done with RMT-filtering and the Sample correlation matrix (bottom right). Here it can be seen that the smaller the branches are, or perhaps the closer to the bottom-left corner they are the lighter they are. Except for these square areas the picture has a very random structure, the reason for this is simple; RMT-filtering reduces the true noise quite effectively and this is what we see in this picture. One interesting observation is that the mean value of the correlation matrix for RMT-Filtering is actually higher than the original matrix for Sample. Either the RMT-filtering has eliminated more of the negative correlations or it has added positive correlations, this new introduced data is probably partly incorrect.

3.5.4 The spectral densities for different amounts of time steps – clearer peaks for more time steps

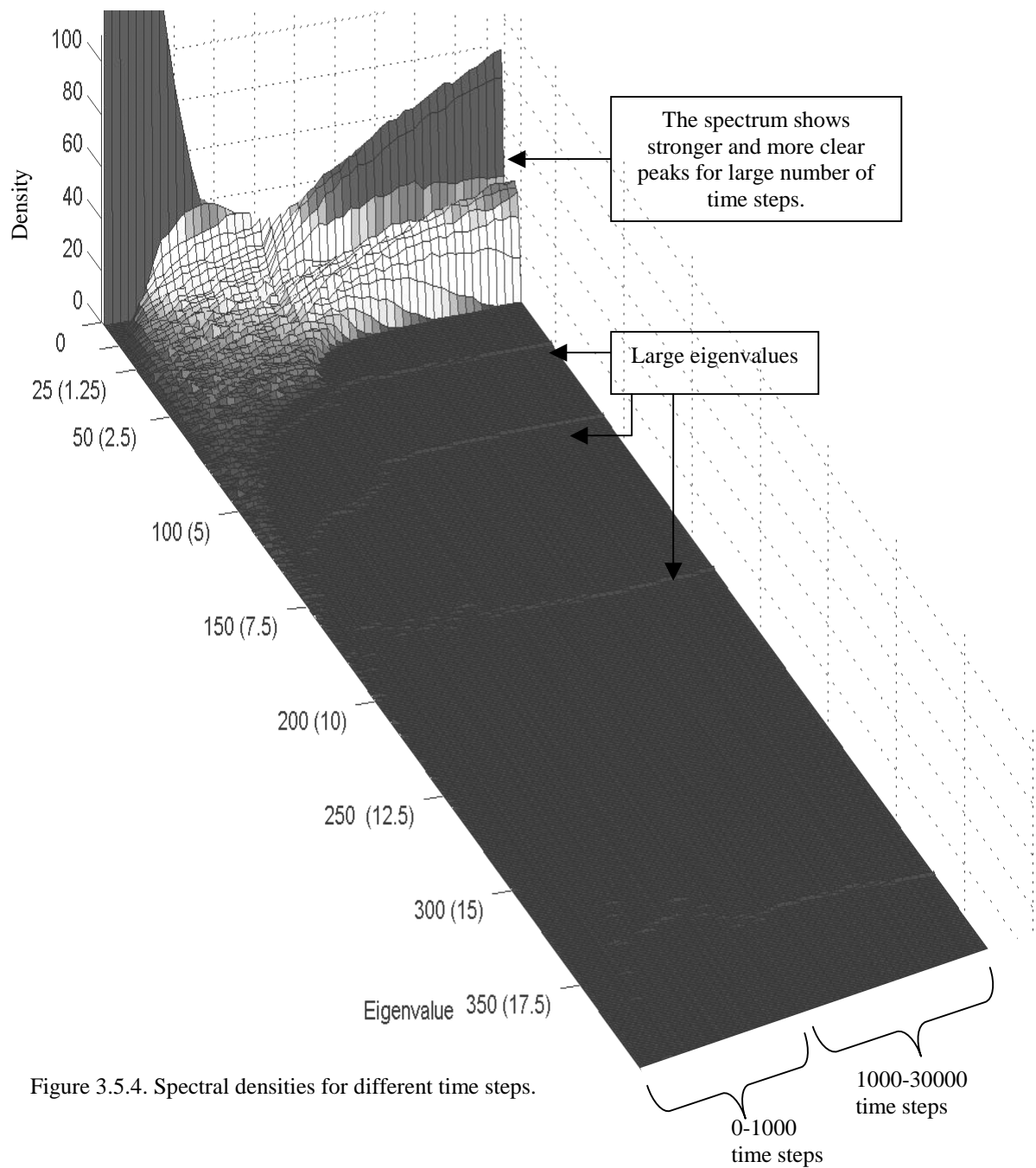


Figure 3.5.4. Spectral densities for different time steps.

The figure above shows several spectral densities for different time steps. A simpler description would be that the picture shows the amount of noise and information that we get from the market. When we have more time steps the peaks are getting smaller and more distinct – there is less and less noise clouding the picture. The large eigenvalues can also be seen (they are hard to spot in print) as pointed out by the arrows; this is the information that the RMT-filtering uses; the rest of the information is ignored.

Chapter 4

Empirical methodology

In this chapter we discuss the data used in the study and the corresponding calculation processes which have been used in the thesis. This chapter is supposed to serve as an overview. The theory has already been discussed and is thus left out of the descriptions. In the analysis experiments have been performed on a large amount of variables, these exact variable setups are not discussed here but in the main analysis in itself to enhance clarity.

4.1 Stock data

4.1.1 First Data Set – Swedish Stock Returns

The first set of data consists of daily stock data including 197 companies currently listed on the Swedish Stockholm Stock Exchange (for a complete compilation and overview of all companies please see Appendix A.3.1). This was the maximum amount of companies that was possible to acquire, many had to be excluded due to constraints presented below. Companies no longer listed on the stock exchange have been excluded; there is thus a survivorship bias in the data. In addition, data on merged companies is available only since after the time of the merger. In cases when a substantial amount of data points are missing the company has been excluded (the extreme limit is set to 35 % missing data points). Companies that only have been available for a short period of time i.e. companies listed within the latest year(s) have also been excluded. The 197 companies used are the companies that were left after the election process presented above. The collected stock data (for all of the 197 companies) range from July 12th 1999 to July 18th 2003. This data for the 197 companies is referred to as Data1a. Further, longer time series containing fewer companies have been created by using only companies having available data prior to 1999, e.g. as long back in time as September 1st 1989. This data is denoted as Data1b (begins at July 1st 1993 and include 67 companies) and Data1c (begins at September 1st 1989 and includes 48 companies).

The data obtained are adjusted for splits, issues and dividends. The data were obtained by the authors of this thesis through the financial databases of the Swedish Stockholm stock exchange, Affärsdata and finally the homepage of the Swedish business magazine Affärsvärlden. The data have been sorted according to the branch structure described in Affärsvärlden.

4.1.2 Second Data Set – U.S. Industry Portfolio Returns

The second set of data consists of monthly returns on U.S. value weighted industry portfolios. The industry portfolios contain stocks from all NYSE (New York Stock Exchange), AMEX (American Stock Exchange), and NASDAQ (National Association of Securities Dealers Automated Quotation) firms. For a complete

compilation and overview of all portfolios please see Appendix A.3.2. More in detail the portfolios were constructed by assigning each NYSE, AMEX, and NASDAQ stock to an industry portfolio based on its four-digit SIC code or equivalent (U.S. Standard Industrial Classification code) at that time, then returns were calculated.

The original data range from July 1926 to December 2002 and contain 43 to 48 industry portfolios (depending on time preference). The authors of this thesis have chosen to work with data ranging from January 1973 to December 2002, this since all portfolios and data are available after December 3 1972. When we later study the mean value of the correlation we look back at a longer time period starting 1931 with 43 companies. These long time series enable us to study the market trends even better. The U.S. research return data is collected/obtained and prepared by Professor Kenneth French at University of Dartmouth and Professor Eugene Fama at the University of Chicago⁸⁰.

4.1.3 Returns

For a number of statistical reasons it is preferable not to work with price series⁸¹, therefore raw series of prices are transformed into series of returns. Additional returns have the benefit of being unit free. Normally there are two ways of calculating returns; simple returns and continuously compounded returns (log returns). Normally one uses continuously compounded returns, due to better statistical properties (certain problematic trends can be eliminated). Though calculating portfolio returns is easily done by using simple returns, while when using continuously compounded returns the value of the portfolio has to be estimated at each time period and then the returns can be determined. In the limit (increasing data sampling frequency) using lesser time steps the two returns will be identical.

Continuously compounded returns:

$$r_t = 100\% \times \ln\left(\frac{p_t}{p_{t-1}}\right)$$

Where r_t denotes continuously compounded return at time t , and p_t denotes asset price at time t .

Simple returns:

$$r_t = \frac{P_t - P_{t-1}}{P_{t-1}} \times 100\%$$

Here r_t denotes simple return at time t , and p_t denotes asset price at time t .

⁸⁰ U.S Stock Return Research Data Base of Professor Kenneth French, 2003

⁸¹ Brooks, C., 2002, Introductory Econometrics for Finance, Cambridge University Press

4.2 Replacing Missing Data

Initially the stock data of 300 listed Swedish companies were compiled, roughly 50 companies were excluded since they had too short time series, a minimum of four (4) years was set as a time limit. Furthermore, approximately 50 companies were excluded since they had too many data points missing. The cut-of-value was set to 35 % of the overall data points (a mean value of 12% missing data points was acquired). This is a rather high value, but the size of the portfolio is important for the analysis since we often are studying the correlation matrices and eigenvalue spectra. The missing data points have been replaced by artificially generated data points. It is worth mentioning that for most of the companies very few or no data points were missing, furthermore this procedure is only carried out on the first data set (containing listed Swedish companies).

The new data points are generated by measuring the closest available data points i.e. next to the missing data point, and drawing a direct line between these points. The standard deviation is then measured for a certain number of data points (30 points) back in time; this obtained standard deviation is then applied to the drawn line between the existing data points creating a simple simulation of the stock behavior. Some special cases are also treated. For example if there is a hole in the very beginning of the data, the first existing data point found is copied to the starting position. Also if holes exist before 30 points have been gathered for the standard deviation then the first found 30 points are used as a basis for the standard deviation. In the picture below one can see the two plotted graphs for the listed Swedish company TV4 AB, one graph representing the original data with missing data points and a second graph (in the same picture, shifted downwards for easy comparison) representing the “new” complete curve. One can see that the two graphs seem to conform quite nicely, after the new artificial data points have been added.

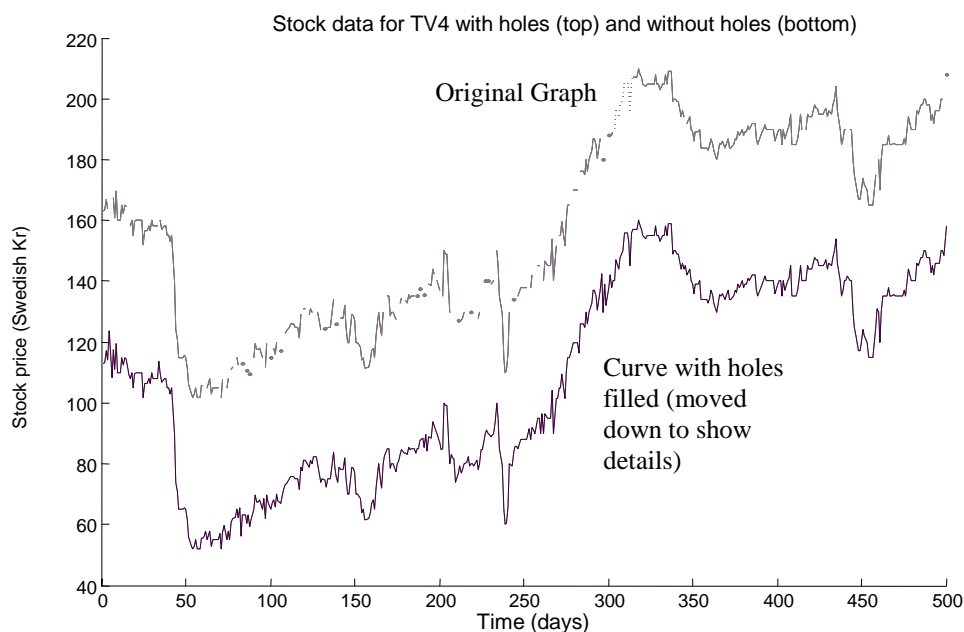


Figure 4.2.1 Two stock charts, one with holes, and one with the holes filled in.

The picture below shows a close up from the picture of the above. It is visible that the simulated data reacts similarly to what one would expect.

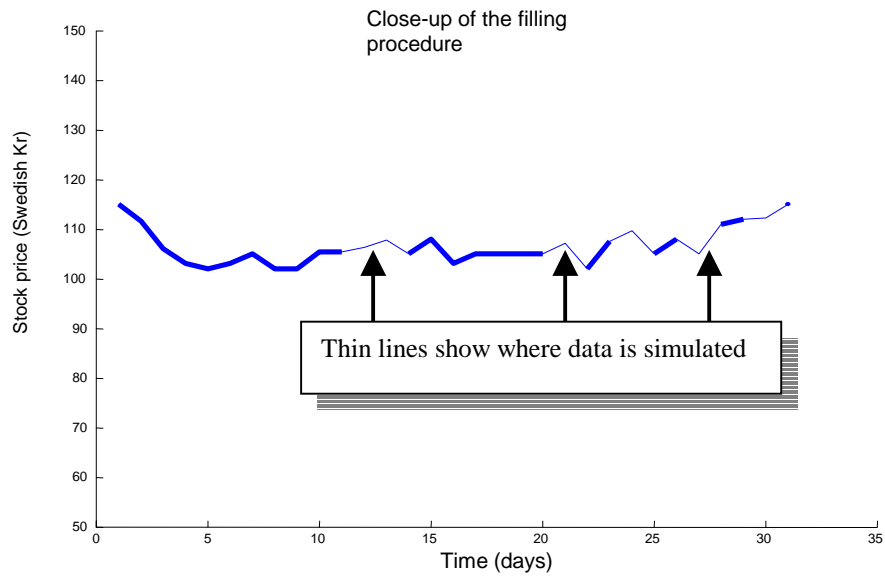


Figure 4.2.2 Close up of the hole-filling procedure

Below are two pictures showing the number/percentages of holes in the initial data with close to 300 companies (note that the width on the x-axis is longer than one unit):

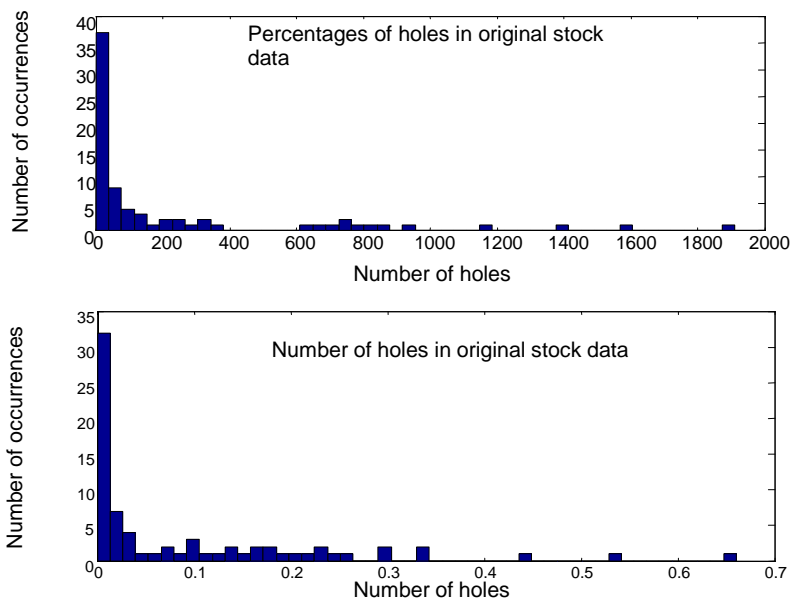


Figure 4.2.3 Information on the existing holes in the data.

4.3 Creation of the Stock Portfolios

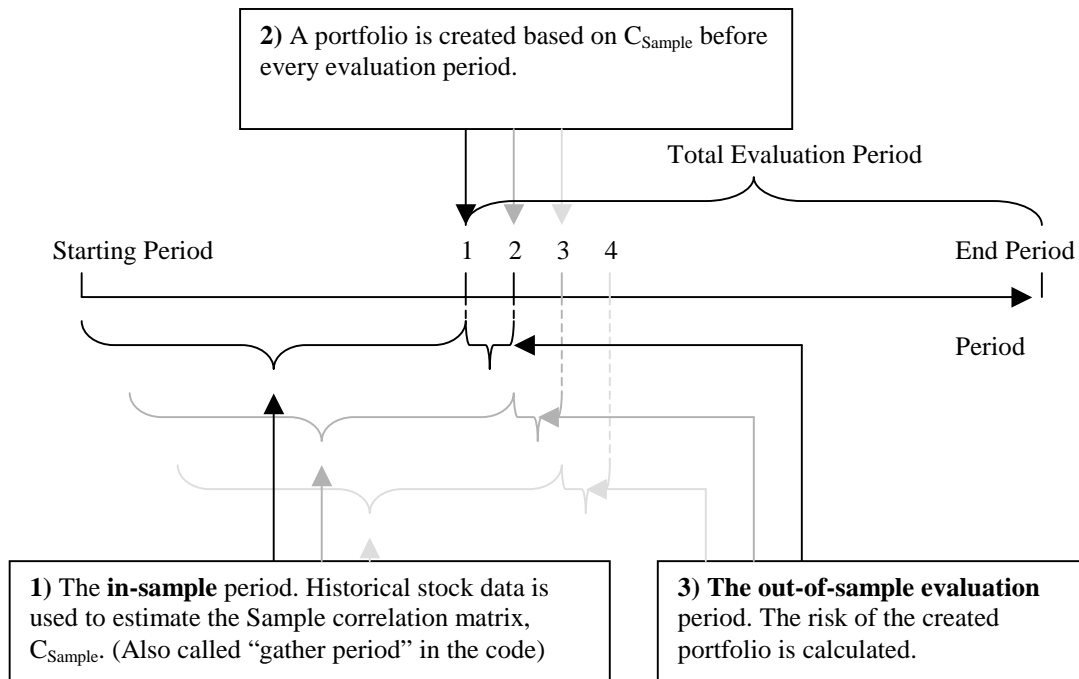


Figure 4.3.1 Portfolios are created and evaluated.

The calculation process can be simplified as described in figure 4.3.1. Historical data is used to create portfolios that we try to improve with Power Mapping and RMT-filtering and later evaluate during a time frame. The length of the in-sample period is different for different parts of the analysis (see Chapter 5 for details). Also the total length of the periods change; the exact numbers are specified in the individual tests in Chapter 5. In Chapter 5.1 where we look at Swedish stock data a period of 5 days is used, for the other analysis of monthly data one year is used. This period is the actual time when we evaluate the created portfolios. This process is repeated for every period in the evaluation period.

Figure 4.3.2 shows a more detailed overview of the structure for all the steps involved. The stock data in the in-sample period are used to calculate the correlation matrix C_{Sample} . C_{Sample} is then improved with Power Mapping, creating the correlation matrix $C_{Power\ Mapping}$, and RMT-filtering that creates the correlation matrix named $C_{RMT-filtering}$. Markowitz Mean-Variance-Theory is then used to calculate the different weights for Sample, Power Mapping and RMT-filtering; the portfolios created with these weights are later evaluated in the following period.

Visual presentation of the calculation process

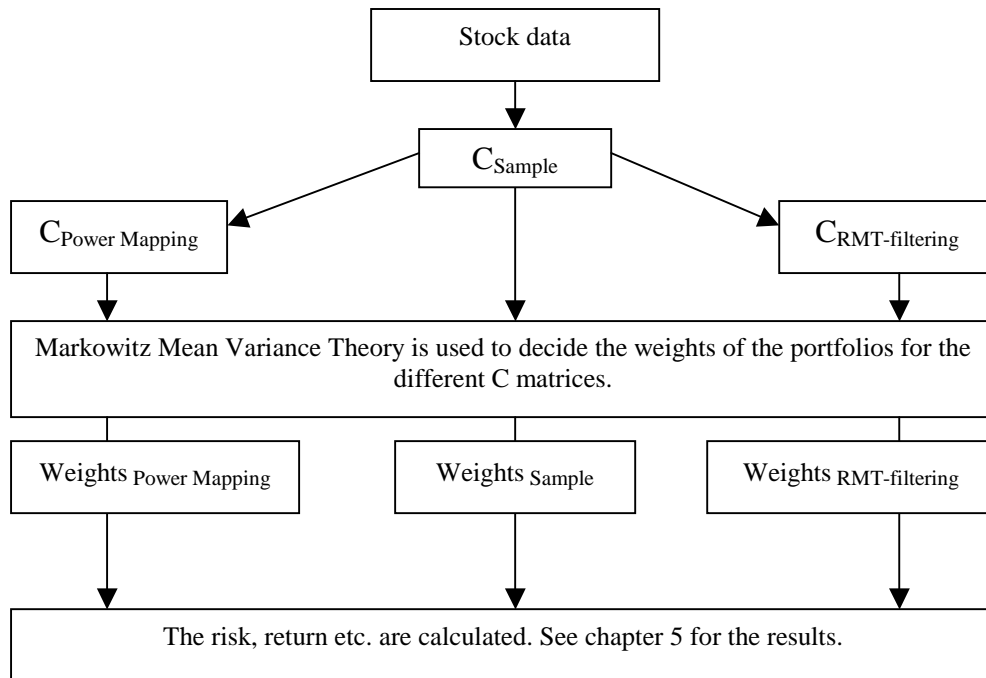


Figure 4.3.2 Portfolios are created for C_{Sample} , $C_{\text{power Mapping}}$ and $C_{\text{RMT-filtering}}$.

4.4 Mean Values of the Columns of the Correlation Matrices

In Chapter 5.6 we study the mean values of the columns of the correlation matrices. The correlation matrices are calculated for Data1b with an in-sample period of 3 years. A new correlation matrix was calculated for every day. Later on in Chapter 5.6 we also study the correlation between the mean values of the correlation matrices and the return of the company. The column-mean-values are split up into monthly values as well as the returns and then the correlation between these two vectors for each month is calculated.

4.5 Portfolios with Different Amount of Standard Deviation

In Chapter 5.3 we divide the 197 Swedish stock data (Data1a) into 7 different portfolios (or groups), they are all sorted according to standard deviation. Below the mean values of the groups standard deviations are presented.

Group 1	Group 2	Group 3	Group 4	Group 5	Group 6	Group 7
0.01845356	0.02268897	0.02813739	0.03540807	0.04294199	0.05214441	0.08778298

4.6 Image Filtering

A short test was also performed using an image filter; the filter used is a contrast filter (see Chapter 5.4). The image filter cuts off all the values lower than a certain threshold and rescales the remaining values so that the smallest values are as low as the lowest one before the cut-off procedure⁸². In this experiment 5% and 10% were the cut-off thresholds used. Since Power Mapping visually is similar to a contrast filter this method was found interesting to try. The data used were the one for the 197 Swedish stock companies used in Chapter 5.1. The main reason for using this specific data is that it's the most high-frequent data in this thesis and generally the data that we are focusing on in our tests.

4.7 Optimization Process used in Matlab 6.5

To solve the quadratic minimization problem associated with finding the efficient frontier (using the standard Markowitz Mean Variance formulas presented in chapter 3) the function showed below, denoted `Fmincon`, was used.

This procedure is very demanding for the processor of the computer, therefore this function is only used for the sole purpose of calculating the efficient frontiers. The calculation time seems to increase exponential with the numbers of companies in the calculations. For example a computer equipped with a 2.0 GHz CPU and 768 MB RAM needs almost half a day to perform calculations (finding efficient portfolios, portfolio weights and estimating the frontier) using 200 companies and 1000 time steps, this even after the Matlab code was seriously optimized and simplified by the authors. The function finds the constrained minimum of a function of several variables, accordingly to the included documentation provided by Matlab⁸³:

The purpose of the function `Fmincon` is to find a minimum of a constrained nonlinear multivariable function subject to

$$\begin{aligned} \min_x f(x) & \text{ subject to} \\ c(x) & \leq 0 \\ ceq(x) & = 0 \\ A \cdot x & \leq b \\ Aeq \cdot x & = beq \\ lb & \leq x \leq ub \end{aligned}$$

where x , b , beq , lb , and ub are vectors, A and Aeq are matrices, $c(x)$ and $ceq(x)$ are functions that return vectors, and $f(x)$ is a function that returns a scalar. Following $f(x)$, $c(x)$ and $ceq(x)$ can be nonlinear functions.

⁸² The mean value of detail that is lost when the data is converted to 16bit data for image filtering is $1.533 \cdot 10^{-5}$

⁸³ Matlab 6.5 Optimization Toolbox

Chapter 5

Results and Analysis

In this chapter we analyze several aspects of the different noise reducing methods; RMT-filtering, Power Mapping and PM-filtering, and their ability to improve certain properties of portfolios. Studies have been performed on different markets, e.g. all (applicable) stocks on the Swedish Stockholm Stock Exchange and industry portfolios from the USA. Different aspects of the correlation matrices are also studied, like the change of correlations over time and possible structures and patterns in the columns of correlation matrices.

5.1 A Study on 197 Swedish Stocks

5.1.1 Introduction

This first analysis (Chapter 5.1) produces the main bulk of the results with heavier commenting and descriptions (than Chapter 5.2 to 5.5). Most of the pictures and graphs are included in the analysis and not placed in Appendix, this to make it easier for the reader since every picture and graph are studied and commented. The figures with little new information have been moved to Appendix. The main goal of this thesis has been to study the abilities of Power Mapping and RMT-filtering to create portfolios with certain preferable characteristics such as lower risk and higher returns; this is what we will begin studying in this first part of the analysis.

5.1.2 Variable Description

Sample is often used as an abbreviation and denotation for the initial unprocessed correlation matrix, in the same way **Power** denotes Power Mapping and **Filter** denotes RMT-filtering. **Total Actual Risk** is the actual risk that the portfolio creates. The **Total Return Actual** is closer to a real word situation than Total Mean Return since it describes how much an investor's capital would grow under the total amount of time; we also call this the "actual return". The **Total Mean Return** is the total actual mean return (expressed in percent) that the portfolio achieves. This is not normally a value that an investor would use, since you seldom discuss the mean changes over a period of time. One advantage of the mean return values is that we are less vulnerable to the sometimes more exponential changes of the "actual return". The **Total Mean Return Div. Total Risk** is simply the Total Mean Return divided by the total Actual Risk. The **Total Return Actual Div. Total Risk** is Total Return Actual divided by Total Actual Risk; this gives a relationship between the return and the risk. A higher ratio would mean that the created portfolio will create higher return with less risk than a portfolio with a smaller ratio.

5.1.3 Overview

In this first part we will study a portfolio containing 197 Swedish stocks. Four portfolios were created with different expected returns (on a daily period basis). Portfolio(1) with 0.002 expected return, Portfolio(2) with 0.0025, Portfolio(3) with 0.003 and the Minimum Variance Portfolio (MVP). The expected returns have been chosen so that portfolios have been possible to create; the expected return cannot be lower than the one for MVP and naturally not higher than the highest possible expected return for the portfolio. As the market fluctuates this can be a noticeable problem since at times the mean return is much higher/lower than what is normal. The fourth portfolio is the Mean Variance Portfolio, thus its return and risk changes for every portfolio. As noted by Bengtsson et al⁸⁴ the MVP is good to use in a study since its connection to the information in the correlation matrix is stronger than in the portfolios created with different expected return. The portfolios with different expected return suffer from the fact that the expected return is non-trivial to create with adequate accuracy. The method used in this thesis is a mean-average method to decide the expected returns, which is one of the most common methods.

5.1.4 Main Results - Unconstrained Weights

The following tables show the main result of the study. The first portfolio is the portfolio closest to the Minimum Variance Point (MVP). As can be seen in the table both Power (Power Mapping) and Filter (RMT-filtering) produce much lower risks than Sample (the unfiltered correlation matrix). The poor performance conducted by Sample is expected, since it has long been recognized that mean-variance efficient portfolios constructed solely by using the original sample covariance matrix perform poorly in-sample (the evaluation period)⁸⁵. The primary reason for this is that the large covariance matrices encountered in practice require estimating too many parameters. One solution to this problem would be to use data with a higher frequency i.e. having a longer studied time frame compared to the number of assets (companies). For an investor this is naturally important, what is also notable though is that the mean return is lower, especially for Power. The total return is the highest for Sample, almost as high for Filter and noticeably lower for Power Mapping. The measure total Return (yearly) divided by total risk (yearly) favors RMT-filtering indicating that RMT-filtering is capable of creating portfolios with a much better return to risk ratio than Sample, and far better than Power Mapping.

⁸⁴ Bengtsson Christoffer. & Holst Jan, Nov 15th 2002, Two Shrinkage Based Covariance Matrix Estimators for Improved Efficiency of Mean-Variance Optimization

⁸⁵ Black, Fisher and Robert Litterman, 1992, Global Portfolio Optimization, Financial Analysts Journal, September-October, 28-43

Table 5.1.1. Results of portfolio number one, no constraints imposed

Portfolio 1 (Exp Return= 0.002)

Total Actual Risk (Sample)= 1.0894% (daily) , Yearly= 17.2249%
Total Actual Risk (Power) = 0.59512% (daily), Yearly= 9.4096%, (Vs Sample= **-45.372%**)
Total Actual Risk (Filter) = 0.59888% (daily), Yearly= 9.4692%, (Vs Sample= **-45.0262%**)

Total Mean Return (Sample)= 0.041397% (daily return)
Total Mean Return (Power)= 0.013145% (daily return)
Total Mean Return (Filter)= 0.031842% (daily return)

Total Mean Return div. Total Risk (Sample)= 0.037999 (daily)
Total Mean Return div. Total Risk (Power) = 0.022089 (daily) (Vs Sample= -41.8714%)
Total Mean Return div. Total Risk (Filter) = 0.053168 (daily) (Vs Sample= 39.9195%)

Total Return Actual (Sample)= 30.47% (whole period) Yearly= 9.270%
Total Return Actual (Power)= 8.9% (whole period) Yearly= 2.883%
Total Return Actual (Filter)= 25.27% (whole period) Yearly= 7.799%

Total Return Actual div. Total Risk (Sample)= 0.538
Total Return Actual div. Total Risk (Power) = 0.306 (Vs Sample= **-43.1%**)
Total Return Actual div. Total Risk (Filter)= 0.8237 (Vs Sample= **55.1%**)

In the second portfolio we see very similar results as the one from portfolio(1) presented earlier. The risk is much lower for both methods but the return is lower here also. This is a general trend for all portfolios. Since we move upwards the efficient front the risk and return is naturally increased for respectively method compared to the previously portfolio. Looking at the measure Total Return divided by total risk we can see that both Filter and Power have improved in comparison with portfolio number one.

Table 5.1.2. Results of portfolio number two, no constraints imposed

Portfolio 2 (Exp Return= 0.0025)

Total Actual Risk (Sample)= 1.192% (daily) ,Yearly= 18.8473%
Total Actual Risk (Power) = 0.64966% (daily) ,Yearly= 10.2721%, (Vs Sample= **-45.49%**)
Total Actual Risk (Filter) = 0.65689% (daily) ,Yearly= 10.3863%, (Vs Sample= **-44.89%**)

Total Mean Return (Sample)= 0.046052% (daily return)
Total Mean Return (Power)= 0.017601% (daily return)
Total Mean Return (Filter)= 0.037191% (daily return)

Total Mean Return div. Total Risk (Sample)= 0.038634 (daily)
Total Mean Return div. Total Risk (Power) = 0.027092 (daily) (Vs Sample= -29.8754%)
Total Mean Return div. Total Risk (Filter) = 0.056617 (daily) (Vs Sample= 46.5468%)

Total Return Actual (Sample)= 33.93% (whole period) Yearly= 10.221%
Total Return Actual (Power)= 12.31% (whole period) Yearly= 3.949%
Total Return Actual (Filter) = 30.04% (whole period) Yearly= 9.150%

Total Return Actual div. Total Risk (Sample)= 0.542
Total Return Actual div. Total Risk (Power) = 0.383 (Vs Sample= **-29.33%**)
Total Return Actual div. Total Risk (Filter) = 0.881 (Vs Sample= **62.5%**)

The third portfolio also shows a very similar picture again. The same remarks as for table 5.1.1 and 5.1.2 are also valid here.

Table 5.1.3. Results of portfolio number three , no constraints imposed

Portfolio 3 (Exp Return= 0.003)

Total Actual Risk (Sample)= 1.3103% (daily) ,Yearly= 20.717%
Total Actual Risk (Power) = 0.71401% (daily) ,Yearly= 11.2894%, (Vs Sample= **-45.506%**)
Total Actual Risk (Filter) = 0.72356% (daily) ,Yearly= 11.4405%, (Vs Sample= **-44.7773%**)

Total Mean Return (Sample)= 0.050734% (daily return)
Total Mean Return (Power)= 0.022079% (daily return)
Total Mean Return (Filter)= 0.042553% (daily return)

Total Mean Return div. Total Risk (Sample)= 0.038721 (daily)
Total Mean Return div. Total Risk (Power) = 0.030922 (daily) (Vs Sample= -20.1407%)
Total Mean Return div. Total Risk (Filter) = 0.05881 (daily) (Vs Sample= 51.8834%)

Total Return Actual (Sample)= 37.19% (whole period) Yearly= 11.111%
Total Return Actual (Power)= 15.77% (whole period) Yearly= 5.000%
Total Return Actual (Filter)= 34.91% (whole period) Yearly= 10.496%

Total Return Actual div. Total Risk (Sample)= 0.536
Total Return Actual div. Total Risk (Power) = 0.443 (Vs Sample= **-17.4%**)
Total Return Actual div. Total Risk (Filter) = 0.917 (Vs Sample= **71.09%**)

We can see that Power Mapping and Filter consistently produces very similar and significantly lower risks than Sample. Surprisingly though the Power Mapping produces a level of return that is far lower than the similar levels of return produced by Filter and Sample. Important to consider in these analysis is that the simulations for the expected returns introduce an error that could and probably will affect the results quite noticeably. Several studies have shown that the overall effects of errors in variances and co-variances are much smaller when compared to the effects of errors in forecasted expected returns.⁸⁶ In practical terms this means that investors would be wise to emphasis acquiring better forecasts of asset's expected returns than in attempting to get better forecasts of variances and co-variances⁸⁷. In reality though this is not easily done, since it is usually far more difficult to estimate expected returns in comparison to estimate the covariance matrix. This indicates that better estimates of the covariance matrix is worth spending time on rather than trying to improve the estimations of the expected returns, even though the improvement of the errors introduced is smaller compared to improvements of the errors connected to the expected returns.⁸⁸

It is also a possibility that the expected returns introduce an error that favors one method. A method to rule this possibility out would of course be to use more methods for creating the expected returns e.g. a Baynesian or factor method/model for the estimations and compare the results. On the other hand the simulations used

⁸⁶ Chopra, V. K., William T. Ziemba, "The Effect of Errors in Means, Variances and Covariances on Optimal Portfolio Choice," Journal of Portfolio Management, July 1993, pp. 6.

⁸⁷ Schwartz Tal, March 15, 2000, How to Beat the S&P 500 with Portfolio Optimization

⁸⁸ Bengtsson Christoffer. & Holst Jan, Nov 15th 2002, Two Shrinkage Based Covariance Matrix Estimators for Improved Efficiency of Mean-Variance Optimization

for the expected returns are similar to the methods used by real analysts⁸⁹, meaning that in a real-world situation this common tool would be used. To avoid possible errors introduced by simulated expected return the Mean Variance Portfolio (the portfolio with the least risk) has also been created. The results for the MVP are presented in the table below. After excluding the expected return vector from the calculations one can see that Filter and Power still are the winners, Sample produces a significantly higher risk.

Table 5.1.4. Results of portfolio number four (Minimum Variance portfolio), no constraints imposed

Portfolio 4 (MVP)

Total Actual Risk (Sample)= 0.9315% (daily), Yearly= 14.7283%
 Total Actual Risk (Power)= 0.52033% (daily), Yearly= 8.2271%, (Vs Sample= **-44.141%**)
 Total Actual Risk (Filter)= 0.50039% (daily), Yearly= 7.9119%, (Vs Sample= **-46.281%**)

Total Mean Return (Sample)= 0.03715% (daily return)
 Total Mean Return (Power)= 0.0079063% (daily return)
 Total Mean Return (Filter)= 0.024293% (daily return)

Total Mean Return div. Total Risk (Sample)= 0.039882 (daily)
 Total Mean Return div. Total Risk (Power)= 0.015195 (daily) (Vs Sample= -61.9009%)
 Total Mean Return div. Total Risk (Filter)= 0.048549 (daily) (Vs Sample= 21.7294%)

Total Return Actual (Sample)= 27.9% (whole period), Yearly =8.54%
 Total Return Actual (Power)= 5.03% (whole period), Yearly = 1.74%
 Total Return Actual (Filter)= 18.86% (whole period), Yearly = 5.93%

Total Return Actual div. Total Risk (Sample) = 0.580
 Total Return Actual div. Total Risk (Power) = 0.212 (Vs Sample= **-63.4%**)
 Total Return Actual div. Total Risk (Filter) = 0.750 (Vs Sample= **29.23%**)

To calculate the weights for the previous 3 tables, the function Fmincon was used (See Appendix). Portfolio(4) is created with the standard formulas presented in the theory chapter, Fmincon has not been used here, this since the same results can for this case be created with different formulas that are much faster⁹⁰ (Fmincon is used in the calculations where constraints are imposed). The MVP shows the same patten as the earlier tables. The MVP is only optimized for risk, meaning that the information about the return is less relevant, and is thus grayed out. It is still interesting to note that the results are similar to the other portfolios. The main reason for the similarities is displayed in the picture below. This is the efficient front for the last years (the efficient fronts generally look fairly similar for all of the studied ones, and this last one was not chosen for a specific reason) created portfolio. Every period will have a different efficient front.

⁸⁹ Anders Johansson Vinga Systems AB., 2003

⁹⁰ Our internal testing showed that the different between our direct calculation and the solution presented by Matlab had a mean difference of $2.602 \cdot 10^{-18}$ indicating that the error most likely only stems from internal CPU rounding errors.

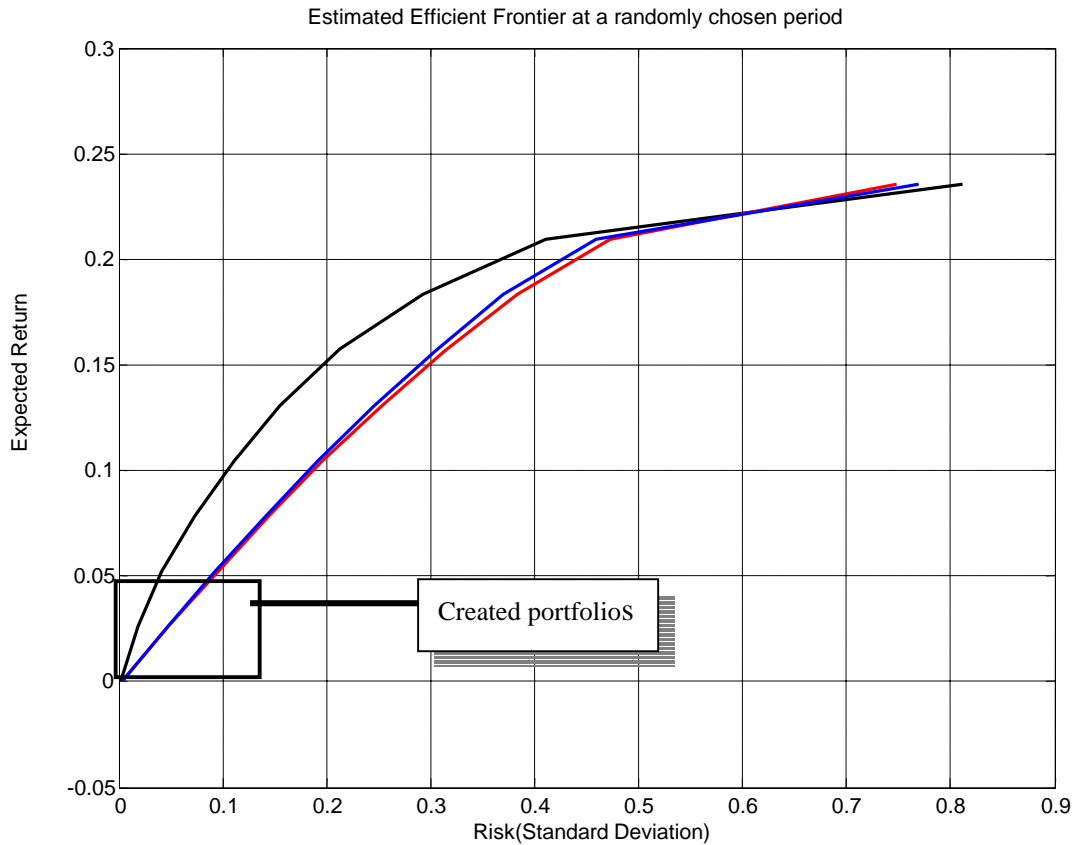


Figure 5.1.1. Efficient frontier and area of examination.

Normally one would like to study portfolios spread out over the efficient front. This has not been possible with our method, since the market changes quite severely over the studied years. This means that at some years it's not possible to create portfolios that are far between, bear markets will make high expected returns impossible. In the later analysis we will mostly study Portfolio(1) and the MVP since they are the farthest apart; Portfolio(2) and Portfolio(3) will be very similar.

The picture below shows the estimated risk (obtained when optimizing the weights) It's important to note that these risk are **estimates** and are not the real ones we will actually get when we create the portfolio. Also note that these risks are only meant to be compared to each other relatively, this risk is not directly comparable to the risk in the rest of the thesis. The picture below shows an interesting pattern, RMT-filtering and Power mapping are very close to each other. They almost look like they are converging which might indicate that they both are getting closer to the "real" information with little noise. The estimated risk for Sample is much lower, which the initial tables of the chapter showed was incorrect. This indicates that the information from C_{SAMPLE} (the correlation matrix from Sample) provides less correct information, resulting in poor estimates for the risk. This is an important finding of the thesis and something that we will discuss more at various points of the thesis.

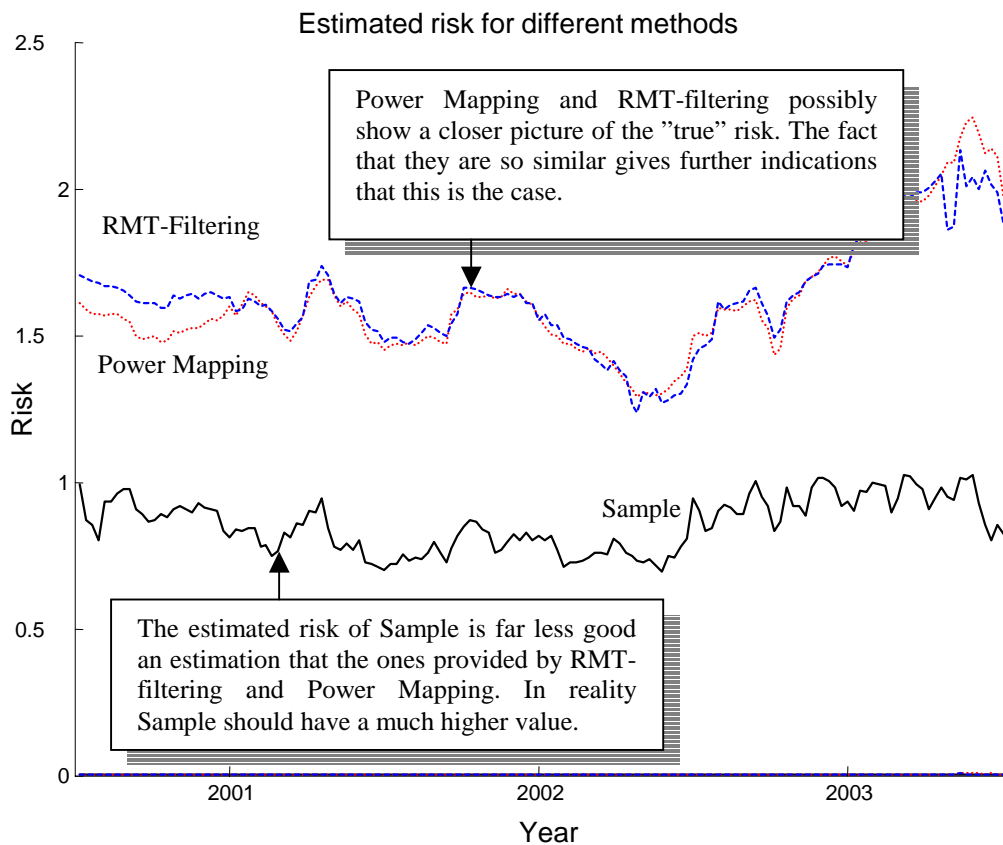


Figure 5.1.2. Estimated risk for portfolio (1) for and Sample, RMT-filtering and Power Mapping

The picture below shows the difference in estimated risk for Power - Sample for Portfolio(1), Portfolio(2) and Portfolio(3). The patterns are very similar, this is most likely a result that the set levels of the expected returns are so close to each other.

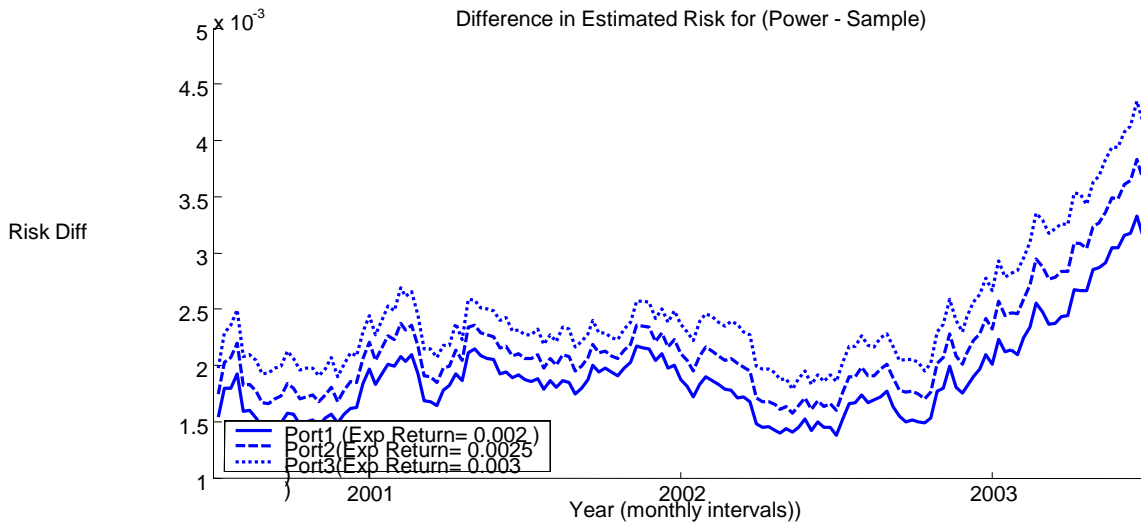


Figure 5.1.3. Difference in estimated return between Power mapping and Sample for portfolios number 1 to 3.

The picture below shows the estimated risk for RMT-filtering - Sample. As can be seen in the picture the estimated risk from the optimization process is not correct for RMT-filtering either, since RMT-filtering actually produces portfolios with lower risk (as was shown in the beginning of the chapter).

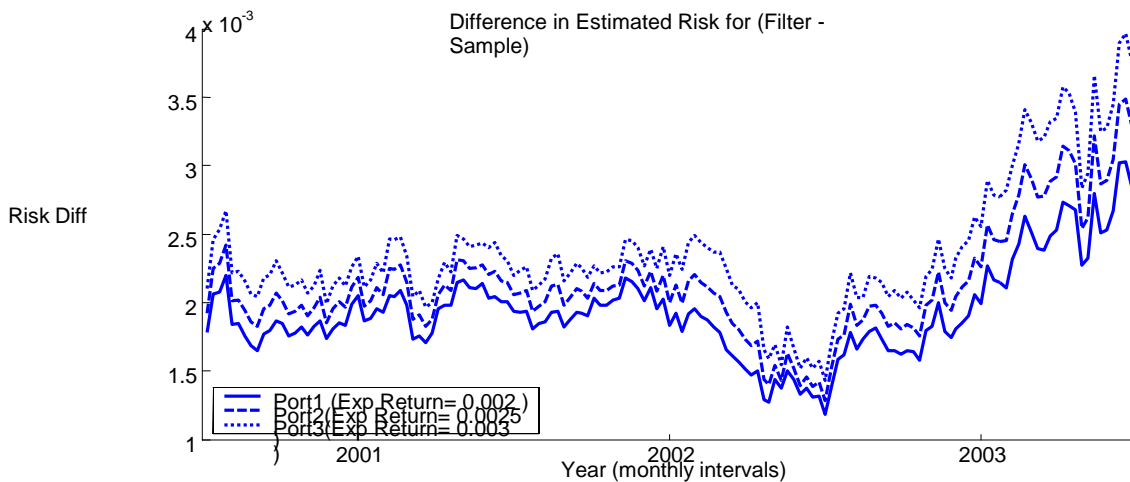


Figure 5.1.4. Difference in estimated return between RMT-filtering and Sample for portfolios number 1 to 3.

The picture below shows the weights for portfolio(1) at the first in-sample period. As can be seen Sample creates portfolios with far larger and smaller weights than the other portfolios. This indicates that what we are actually seeing here is a result from the noise in the matrix C_{Sample}^{91} . Large negative portfolio weights are associated with the extreme high row-sums of the estimated covariance matrix, and these extreme high row-sums are more likely to be caused by estimation error⁹². This might also create portfolios that have severely higher or lower return than the other methods since the portfolio fluctuates so strongly. E.g. it is possible that Sample by pure chance manages to invest heavily in very profitable stocks. In the long run this effect should however even out if enough portfolios are created. It is also visible that Power Mapping and RMT-filtering generate a smaller amount of short selling compared to Sample. This creates portfolios that are more suitable to real world investors.

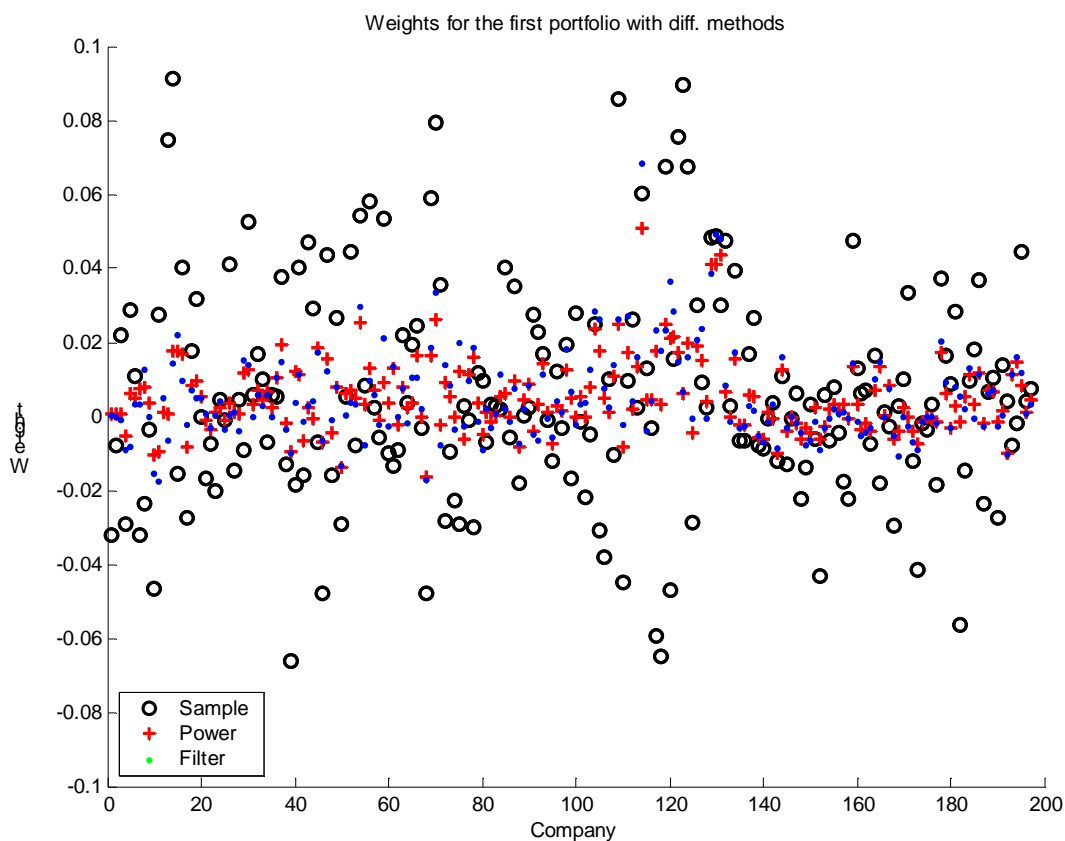


Figure 5.1.5 .Weights corresponding to Power mapping, RMT-filtering and Sample calculated for portfolio number one.

The next picture shows the actual daily risk calculated at monthly intervals for Portfolio(1). It is clear that Power Mapping and RMT-filtering have less risk in the period at practically all occasions. Some areas are interesting though. The left arrow on the picture indicates a situation where the risk is near identical for all methods. The reason for this is unclear; however for the later periods where all methods

⁹¹ Thomas Guhr, personal interview, 2003

⁹² Jagannathan Ravi. Tongshu Ma. Risk Reduction in Large Portfolios: A Role for Portfolio Weight Constraints. September 19, 2001

create portfolios with higher risk there is a plausible explanation. When the market is in turmoil it is much harder to create good portfolios no matter the content of the correlation matrix. Since most stocks are going down it is hard for the Markowitz optimization to spread the risk. This should explain why the methods create fairly similar results.

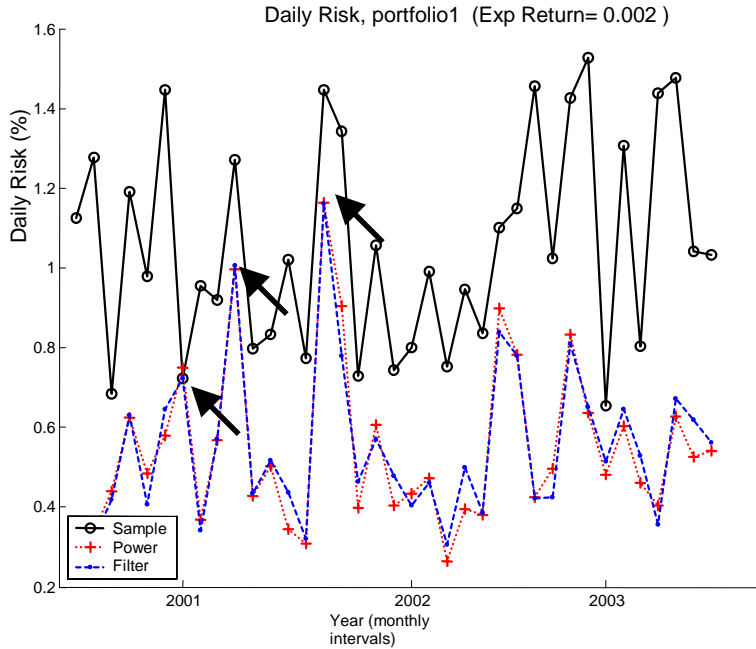


Figure 5.1.6. Observed daily risk for portfolio number one.

The picture below shows the mean daily returns for the portfolios created. Sample fluctuates much more for the return also but it manages to create a portfolio with higher mean return anyway.

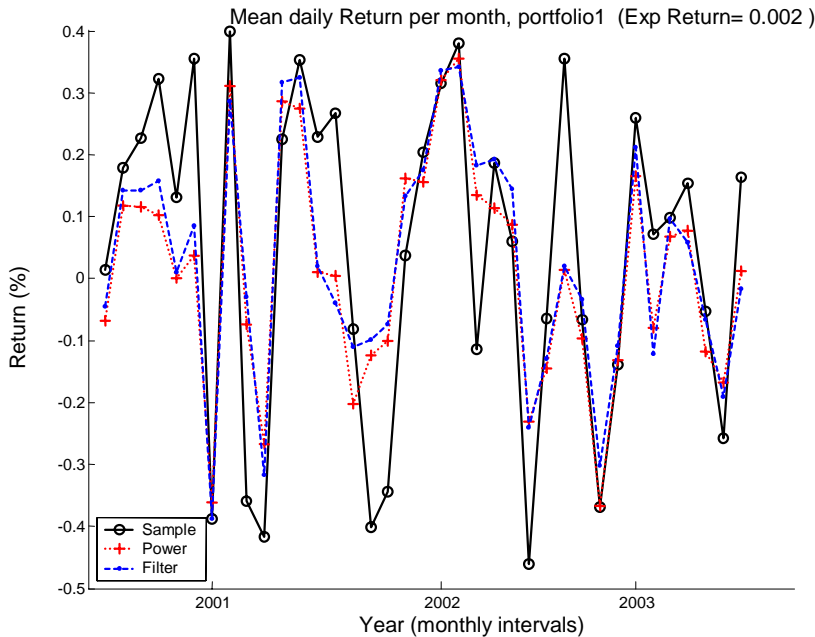


Figure 5.1.7. Observed mean daily return calculated each month for portfolio number one

The picture below shows the actual monthly return (meaning this is the actual return an investor would get over a month if he invested one unit. 1.0 means the investor

will get no return). The picture is practically identical to the previous one. The actual return over the complete return over the complete period is however quite different, since it changes exponentially. It is also worth mentioning that these values look fairly similar to each other when in reality they are not, many small changes lead to the quite different results presented in the initial table; here reviewed for convenience:

Table 5.1.5. Total daily mean return for portfolio number one

Total Mean Return (Sample)= **0.041397%** (daily return)
 Total Mean Return (Power)= **0.013145%** (daily return)
 Total Mean Return (Filter)= **0.031842%** (daily return)

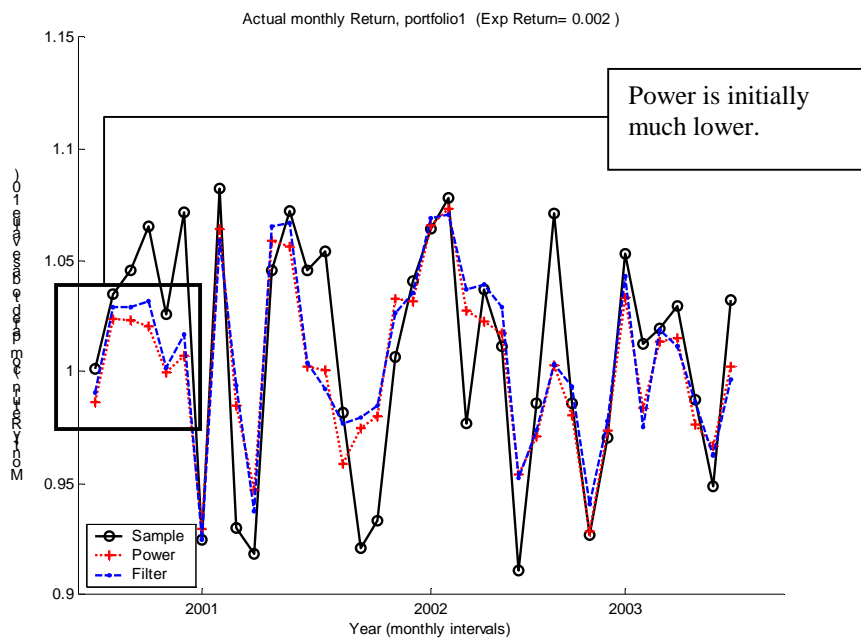


Figure 5.1.8. Actual monthly return for portfolio number one

As previously mentioned the MVP is not affected by the expected return at all, which improves the accuracy of the calculated risk but also means that it is of more limited interest to look at the achieved actual return (since optimization is only done to minimize risk and not to maximize return or using a fixed set of return). The picture below shows the daily risk for the MVP. The picture shows the same trends as for Portfolio(1), the risk is much higher for Sample than it is for Power and Filter. We also see that the occasion indicated by the arrow is now more different for the three methods (for Portfolio(1) this was almost identical). This might indicate that the previous identical result was a random occurrence and that if we increased the expected return they would surely differ even more.

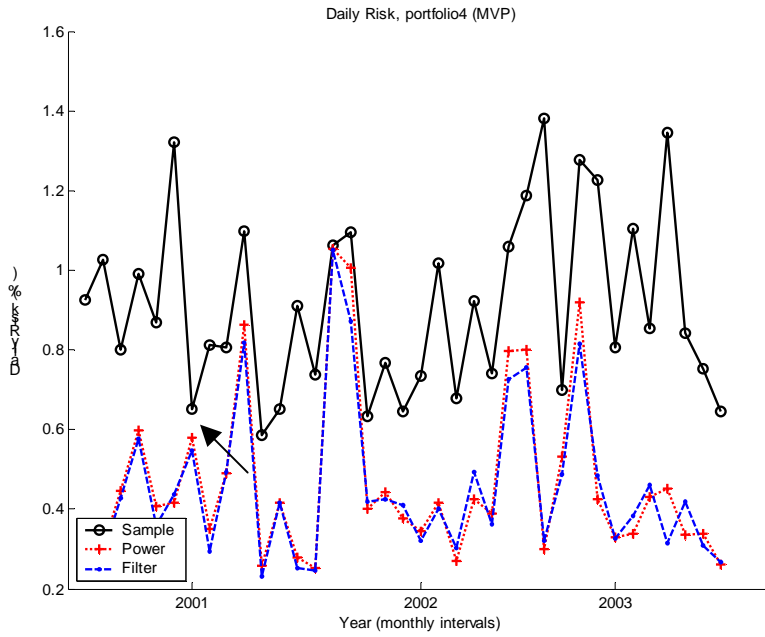


Figure 5.1.9. Daily risk for portfolio number four, the mean variance portfolio (MVP)

We look at the mean return for the MVP briefly and conclude that it is very similar to the figure for Portfolio(1).

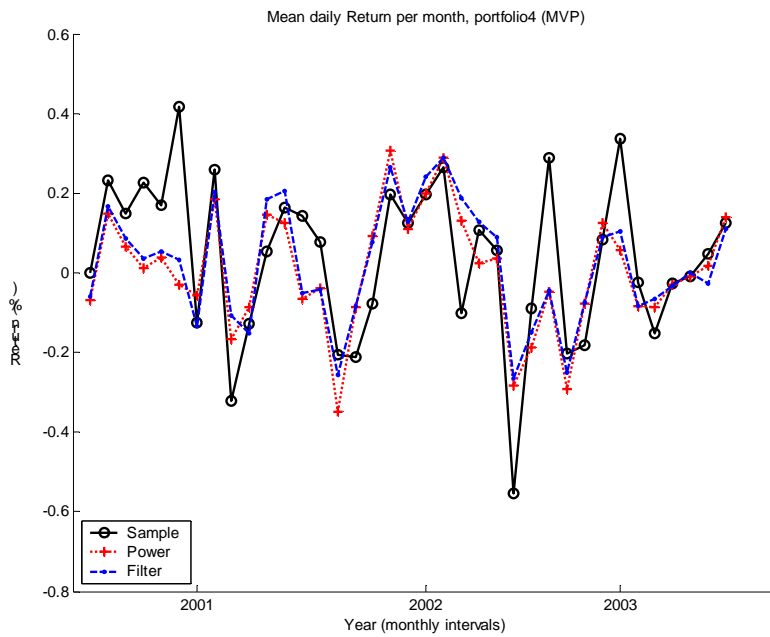


Figure 5.1.10. Mean daily return per month for portfolio four, the mean variance portfolio (MVP)

The actual return in the figure below looks near identical to the mean return in the previous figure.

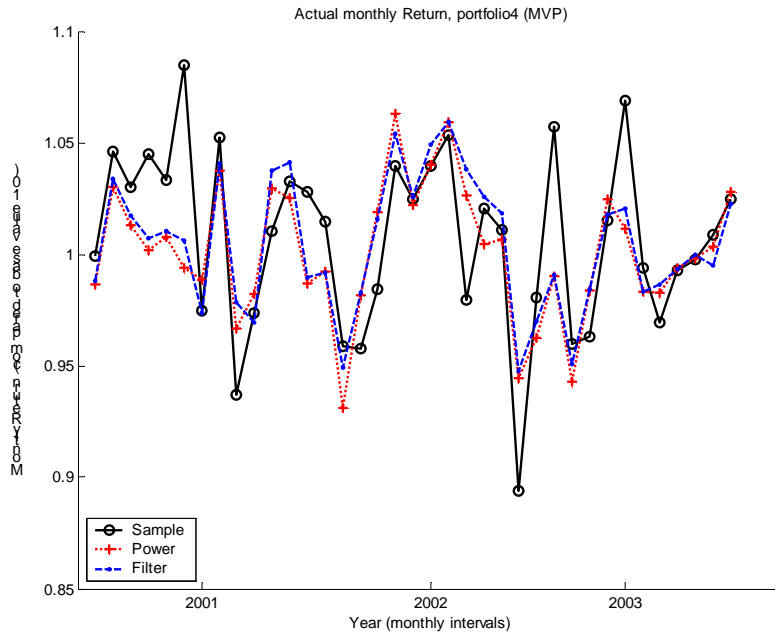


Figure 5.1.11. Actual monthly return for portfolio number four (MVP).

5.1.5 Correlation structure

The three figures below show C_{SAMPLE} , C_{POWER} and C_{FILTER} for the first year/period (this is the first correlation matrix created with $N = 197$ companies; we create 149 more matrices in the following periods). The bottom right picture shows a simulated correlation matrix done with Noh's model. As can be seen in the initial Sample matrix the branch structure is very weak. This is most likely a result from the large amount of smaller companies that are relatively unaffected by the larger companies and the obvious fact that the period used to create the correlation matrix isn't very long. The reason for using one-year data was the aim to create a correlation matrix that was dependent on recent stock fluctuations. If the in-sample period (the time used for creating the correlation matrix) is several years stronger patterns will emerge⁹³. The arrows in the top-left figure indicate companies in a branch which doesn't follow the general trend of the branch. The dark stripes mean that the companies have very low correlations or negative correlations with the different companies. Upon closer inspection the strong branch in the figure proved to be IT-companies which for all but for the "dark companies" had a very bad year at the time. In both RMT-filtering and Power mapping it is clear that the random noise visible in the Sample matrix is less in C_{POWER} and C_{FILTER} .

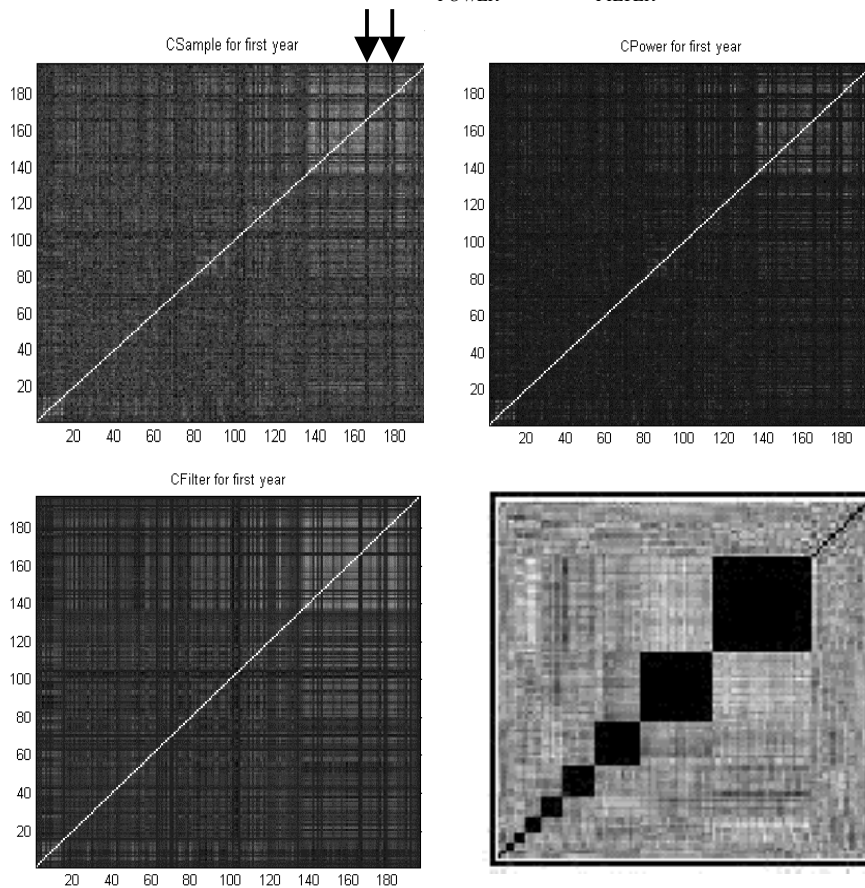


Figure 5.1.12. Correlation matrices generated by Sample, Power mapping and RMT-filtering using a 250 day in-sample period and 197 Swedish companies.

⁹³ As seen in Andersson, P. Öberg, A. 2003. Noise Reducing Methods for Correlation Matrices New Techniques to Minimize Risk in Portfolio Management?

Visually RMT-filtering seems to create a somewhat too strong line-structure in the above figure which should come from the fact that in reality very little of the information contained in the initial C_{SAMPLE} is really used. Also the RMT-filtering should suffer from the absence of clear braches and large eigenvalues, as is shown in the previous figure. Figure 5.1.13 shows the eigenvalue spectra for C_{SAMPLE} for the last portfolio⁹⁴. One very large eigenvalue exist which should represent the overall market information⁹⁵. So this is actually the general market “trend” we are seeing. We had anticipated finding larger eigenvalues outside the main bulk of the distribution since we use so many companies when creating the correlation matrix. The “missing” eigenvalues can be explained by the absence of a clear branch structure. As noted previously this probably depends on the fact that we have a short in-sample period for creating the correlation matrices and also that we have many smaller companies in the portfolio. It might even be that there is no strong structure of branches on the Swedish market, and that companies have a tendency to move between branches or not belong to a certain branch. If we had studied all the branches on the U.S. market we would expect to get larger eigenvalues from strong branch structures, this is also the case as we will see later on in this thesis. Though the main issue here, that is really puzzling, is that the RMT-filtering does not seem to suffer from the absence of clear branch structures. The portfolios created by RMT-filtering produced remarkably good results (as shown above). The reason why this is strange is that the method is created in such way that it should work when there is a clear structure, which in not the case in this situation. Though a somewhat comparable result can be observed in the study by Bengtsson et al 2002⁹⁶, although Bengtsson et al use a RMT-filtering approach in combinations with another different technique. It is hard to find any suitable explanation for this result, and further investigations can hopefully shed light upon these questions.

⁹⁴ The reason for choosing the last portfolio is only to simplify computations. Spectra for other periods show similar pictures and this one represents a standard situation. An animation of the eigenvalue spectrum would be interesting but was excluded due to time constraints.

⁹⁵ Vasiliki Plerou et al, 22 February 1999. Universal and Non universal Properties of Cross Correlations in Financial Time Series

⁹⁶ Bengtsson, Christoffer. & Holst, Jan, Nov 15th 2002, Two Shrinkage Based Covariance Matrix Estimators for Improved Efficiency of Mean-Variance Optimization

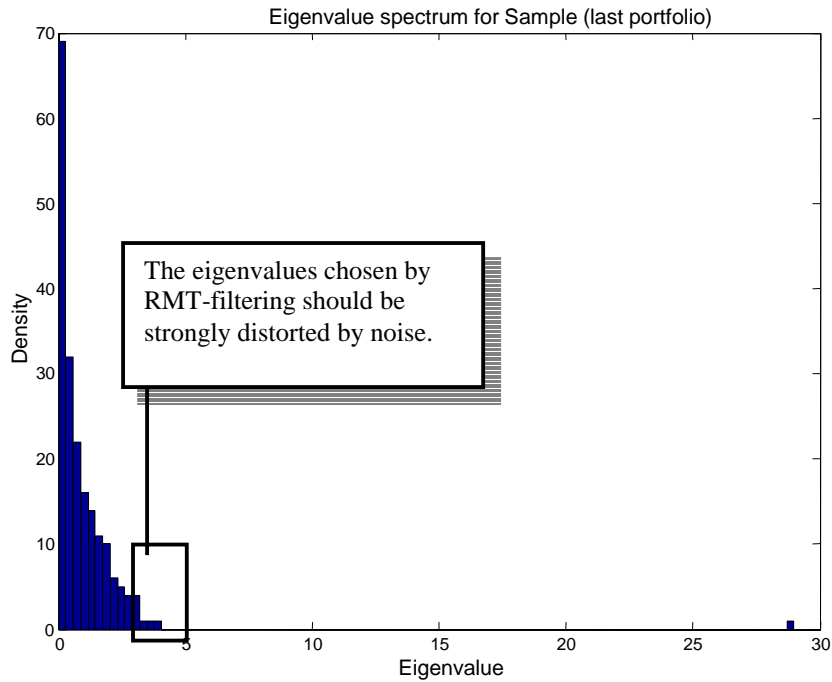


Figure 5.1.13. Eigenvalue spectra for Sample correlation matrix

The figure below shows the eigenvalue spectrum for C_{POWER} . This figure looks much better compared to the results of Sample especially if we compare it to the eigenvalue spectrum created by Noh's model⁹⁷ that represents a near ideal situation.

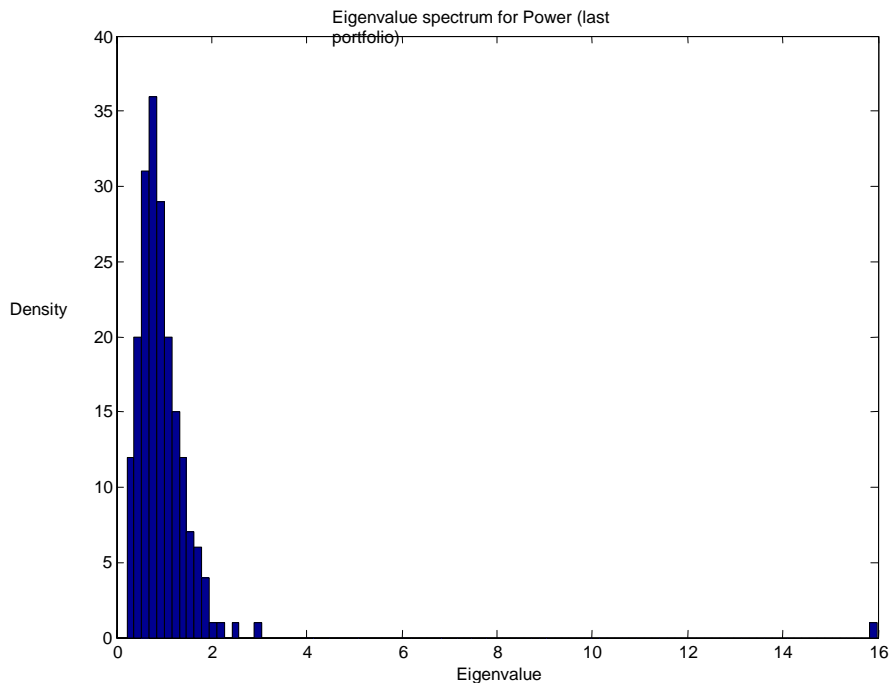


Figure 5.1.14. Eigenvalue spectra for Power Mapping correlation matrix

⁹⁷Andersson, P. Öberg, A. 2003. Noise Reducing Methods for Correlation Matrices New Techniques to Minimize Risk in Portfolio Management?

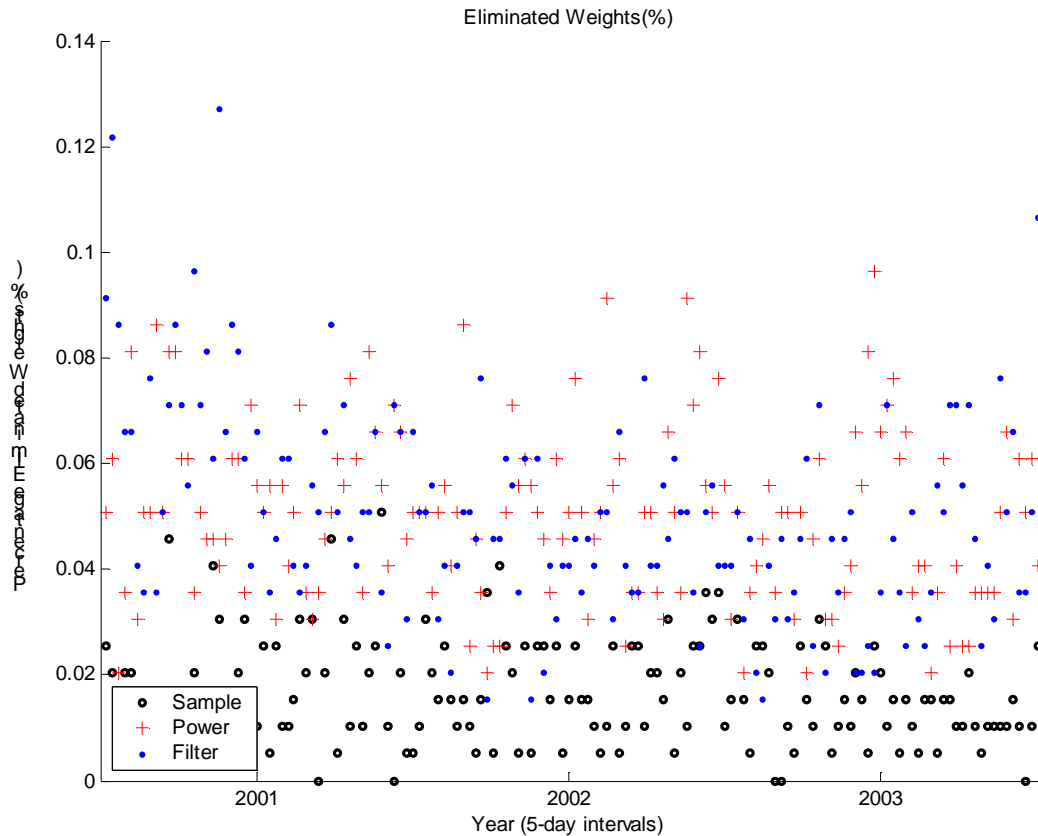


Figure 5.1.15. Percentage of eliminated weights for Sample, Power mapping and RMT-filtering

As suggested by tutor Göran Andersson unrealistically small weights were eliminated. The cut-of value chosen was 0.01%. For a portfolio of one million units this would mean that investments below 100 units would be excluded from the calculations. The picture above shows the percentage eliminated weights (for the first portfolio). As can be seen Sample has the lowest number of weights eliminated, not surprisingly since Sample generally have noticeably larger values. The correlation matrices and the resulting portfolios should be affected very little since most of the eliminated weights are ranging from approximately 10^{-10} to 10^{-20} (i.e. are almost zero). In the picture discrete steps can also be noted, meaning Sample have for example most of the time 3-4 weights eliminated.

5.1.5 Main Results - Constrained Weights

The tables below contain result from a study performed in almost the exact same way as the study presented in section 5.1, but with one important difference, we now introduce constraints forbidding the use of non-positive weights (i.e. no short selling is allowed). Though we don't introduce any upper bounds referring to the positive size of the weights. The reasons why we implement constraints are several, first of all the real investor often has limited or no opportunities/authorization to sell short due to various reason, secondly the introducing of constraints have some interesting theoretical and practical implications as we will soon see.

In the first portfolio (again the one closest to the MVP, using expected return = 0.001) we can see that RMT-filtering produces the smallest risk followed close by Power Mapping, Sample again produces the highest risk.

Total Return is the highest for Power Mapping, lower for RMT-filtering and still lower for Sample (Sample and RMT-filtering are relatively similar in aspects of return). The total Return is far lower for all methods compared to what we have seen in the study performed using no constraints. The reason is most likely that the Swedish stock market has declined during recent years (which we are studying), not allowing short selling presents serious limitations to the optimization process and for the portfolios to generate high returns in comparison to a scenario when short selling is allowed.

Looking at the ratio total Return divided by total Risk reveals that Power produces the best ratio, though closely followed by Filter and Sample.

Table 5.1.6. Results of portfolio number one, constraints imposed

Portfolio 1 (Exp Return= 0.001)

Total Actual Risk (Sample)= 0.64051% (daily), Yearly= 10.1274%
 Total Actual Risk (Power) = 0.6289% (daily), Yearly= 9.9438%, (Vs Sample= -1.8125%)
 Total Actual Risk (Filter) = 0.62405% (daily), Yearly= 9.8672%, (Vs Sample= -2.5693%)

Total Mean Return (Sample)= 0.0040221% (daily return)
 Total Mean Return (Power)= 0.0064344% (daily return)
 Total Mean Return (Filter)= 0.0045589% (daily return)

Total Mean Return div. Total Risk (Sample)= 0.0062795 (daily)
 Total Mean Return div. Total Risk (Power) = 0.010231 (daily) (Vs Sample= 62.9308%)
 Total Mean Return div. Total Risk (Filter) = 0.0073053 (daily) (Vs Sample= 16.3366%)

Total Return Actual (Sample)= 1.49% (whole period) Yearly= 0.494%
 Total Return Actual (Power)= 3.4% (whole period) Yearly= 1.12%
 Total Return Actual (Filter)= 1.98% (whole period) Yearly= 0.66%

Total Return Actual div. Total Risk (Sample)= 0.049
 Total Return Actual div. Total Risk (Power) = 0.1126 (Vs Sample= **130%**)
 Total Return Actual div. Total Risk (Filter) = 0.067 (Vs Sample= 33.67%)

The second portfolio (moving up the efficient frontier, using expected return = 0.00125) we can again see that Filter produces the smallest risk followed closely by Power, Sample produces the highest risk. The difference between the risks are fairly small, especially when looking at a yearly measured risk.

Total Return is the highest for Power, Filter is still lower but is now very close to the return of Power, Sample this time produces a relatively lower return compared to those of Power and Filter. Looking at the ratio total Return divided by total Risk reveals that Power again produces the best ratio followed by Filter. Sample produces a somewhat lower ratio compared to the other methods.

Table 5.1.7. Results of portfolio number two, constraints imposed

Portfolio 2 (Exp Return= 0.00125)

Total Actual Risk (Sample)= 0.70111% (daily), Yearly= 11.0856%
 Total Actual Risk (Power) = 0.69079% (daily), Yearly= 10.9224%, (Vs Sample= -1.472%)

Total Actual Risk (Filter) = 0.68811% (daily) ,Yearly= 10.88%, (Vs Sample= -1.855%)

Total Mean Return (Sample)= 0.0082275% (daily return)

Total Mean Return (Power)= 0.013865% (daily return)

Total Mean Return (Filter)= 0.011823% (daily return)

Total Mean Return div. Total Risk (Sample)= 0.011735 (daily)

Total Mean Return div. Total Risk (Power) = 0.020071 (daily) (Vs Sample= 71.038%)

Total Mean Return div. Total Risk (Filter) = 0.017182 (daily) (Vs Sample= 46.4199%)

Total Return Actual (Sample)= 4.43% (whole period) Yearly= 1.46%

Total Return Actual (Power)= 8.99% (whole period) Yearly= 2.88%

Total Return Actual (Filter)= 7.35% (whole period) Yearly= 2.39%

Total Return Actual div. Total Risk (Sample)= 0.132

Total Return Actual div. Total Risk (Power) = 0. 264 (Vs Sample= 99.7%)

Total Return Actual div. Total Risk (Filter) = 0.219 (Vs Sample= 65.9%)

In the third portfolio (using expected return = 0.0015) we can see that Filter (now produces the smallest risk though) and Power produces almost the same risk. Sample produces a somewhat higher risk.

The differences between the risks are smaller this time, when looking at yearly measured risk they deviate little from each other. Of course when investing large amounts of capital in volatile assets even a small change can have a large impact. Total Return is the highest for Power, Filter is lower but close to Power, Sample this time produce a lower return compared to those of Power and Filter.

Looking at the ratio total Return divided by total Risk reveals that Power yet again produces the best ratio, though as before Filter produces a quite similar ratio. Sample yet again produce a somewhat lower ratio compared to the other methods.

Table 5.1.8. Results of portfolio number three, constraints imposed

Portfolio 3 (Exp Return= 0.0015)

Total Actual Risk (Sample)= 0.79061% (daily) ,Yearly= 12.5007%

Total Actual Risk (Power) = 0.78187% (daily) ,Yearly= 12.3624%, (Vs Sample= -1.1059%)

Total Actual Risk (Filter) = 0.78112% (daily) ,Yearly= 12.3507%, (Vs Sample= -1.2001%)

Total Mean Return (Sample)= 0.010783% (daily return)

Total Mean Return (Power)= 0.01799% (daily return)

Total Mean Return (Filter)= 0.015507% (daily return)

Total Mean Return div. Total Risk (Sample)= 0.013639 (daily)

Total Mean Return div. Total Risk (Power) = 0.023009 (daily) (Vs Sample= 68.7018%)

Total Mean Return div. Total Risk (Filter) = 0.019852 (daily) (Vs Sample= 45.5547%)

Total Return Actual (Sample)= 5.92% (whole period) Yearly= 1.94%

Total Return Actual (Power)= 11.86% (whole period) Yearly= 3.81%

Total Return Actual (Filter)= 9.8% (whole period) Yearly= 3.16%

Total Return Actual div. Total Risk (Sample) = 0.1552

Total Return Actual div. Total Risk (Power) = 0.307 (Vs Sample= 97.8%)

Total Return Actual div. Total Risk (Filter) = 0.256 (Vs Sample= 64.9%)

The fourth portfolio is the Minimum Variance portfolio; we can here see that Filter produces the lowest risk. Power produces a higher risk and Sample produces the highest risk.

Table 5.1.9. Results of portfolio number four (MVP), constraints imposed

Portfolio 4 (MVP)

Total Actual Risk (Sample)= 0.55115% (daily) ,Yearly= 8.7145%
 Total Actual Risk (Power)= 0.54095% (daily) ,Yearly= 8.5532%, (Vs Sample= -1.8512%)
 Total Actual Risk (Filter)= 0.52181% (daily) ,Yearly= 8.2506%, (Vs Sample= -5.3236%)

Total Mean Return (Sample)= -0.016305% (daily return)
 Total Mean Return (Power)= -0.0072824% (daily return)
 Total Mean Return (Filter)= -0.01106% (daily return)

Total Mean Return div. Total Risk (Sample)= -0.029583 (daily)
 Total Mean Return div. Total Risk (Power)= -0.013462 (daily) (Vs Sample= -54.493%)
 Total Mean Return div. Total Risk (Filter)= -0.021196 (daily) (Vs Sample= -28.3499%)

Total Return Actual (Sample)= -12.51 % (whole period) Yearly = -4.36%
 Total Return Actual (Power)= -6.352 % (whole period) Yearly = -2.16%
 Total Return Actual (Filter)= -8.898 % (whole period) Yearly = -3.08%

Total Return Actual div. Total Risk (Sample)= -0.501
 Total Return Actual div. Total Risk (Power)= -0.253 (Vs Sample= 49.5%)
 Total Return Actual div. Total Risk (Filter)= -0.3733 (Vs Sample= 25.5%)

One interesting feature from the results above is that the risk produced by Sample when imposing constraints is far smaller than the risk for the same portfolios when allowing short selling (i.e. no constraints). The explanation why the risk produced by Sample decrease is because the imposing of constraints (no short selling) can in some aspects be interpreted as a method of shrinking the largest co-variances and variances (which are causing the non-positive weights) towards more standard values, since it can be argued that these extreme estimates are those most likely caused by estimation error.⁹⁸ The large reduction in risk when imposing constraints indicates that the Sample matrix suffers from estimation error. Looking at the changes in risk when imposing constraints for Power Mapping and RMT-filtering we can see that the risks have increased slightly, this is thus a good sign since this indicates that we do have good estimates of the matrices at the beginning, and that the increased risk is a product of errors introduced by imposing constraints.

Though one has to be careful not to jump to conclusions since studies have shown that it's difficult to dismiss the observed extreme negative and positive weights as entirely due to imprecise estimation of the inputs used to construct the portfolios, they might be the result of the presence of a single dominant factor in the covariance matrix of return. Imposing constraints also introduces specification error⁹⁹.

Obviously the net result will be a trade-of between the reduction in sampling error due to the imposed constraints and the increase in speciation error from the imposing, How the general outcome will be is hard to tell, but generally it seems

⁹⁸ Jagannathan Ravi. Tongshu Ma. Risk Reduction in Large Portfolios: A Role for Portfolio Weight Constraints. September 19, 2001

⁹⁹ Green, Richard C. and Burton Hollifield, 1992, When Will Mean-Variance Efficient Portfolios be Well Diversified? Journal of Finance 47, 1785-1809

like in most cases the reduction in estimation error is far larger the larger the specification error that is introduced¹⁰⁰ From the results above it seems likely that this is indeed the case in the test conducted. We can clearly see that Sample has much lower risk while Power Mapping and RMT-filtering have higher risk compared to the previous test with no constraints.

In the graph below one can see how many of the weights that were set to the default value zero (eliminated), this since they were far too small to be realistic (in a real world situation). The percentage of the weights eliminated is far greater than the equivalent eliminated percentage in the earlier study when implying no constraints. A rough estimate of the mean percentage eliminated among the three methods would be slightly above 2.3%. From the graph one can see that Power mapping generally generates a smaller number of very small weights especially in the beginning of the time window, while Filter and Sample generate approximately the same numbers of small weights (with a small favor to RMT-filter). One way to interpret the result concerning the large elimination of weights would be that these weights otherwise would have been negative of at least could be combined in such way that they could generate more negative weights.

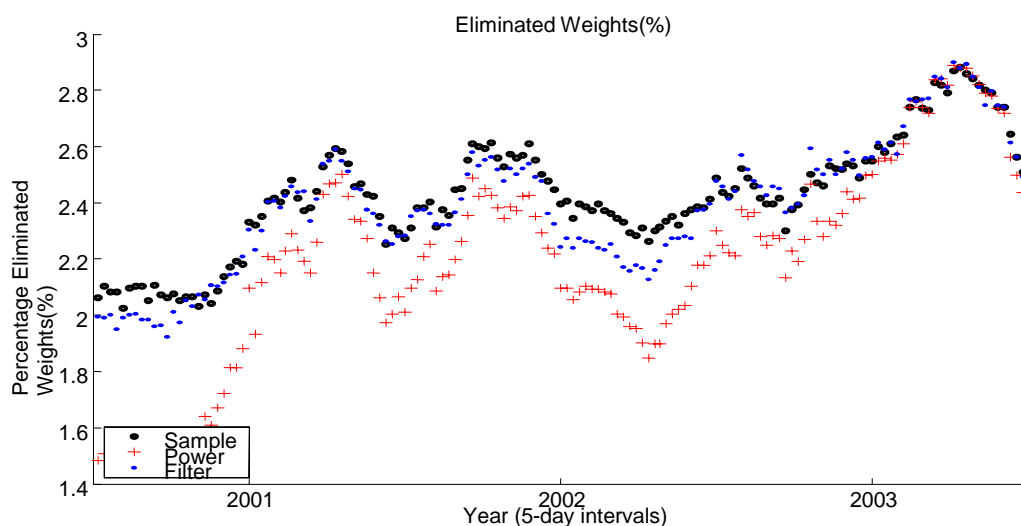


Figure 5.1.16. Percentage of eliminated weights for Sample, Power mapping and Sample

The next figure contains the results of the calculations of the estimated risk of portfolio number one implying constraints (it is worth mention once again that we get this estimated risk in the process of optimizing the portfolio weights). These risks are only meant to be compared to each other relatively, just like the estimated risks computed and calculated in the previous section when implying no constraints. The graph below shows that all estimated risks for Filter, Power and Sample are closer to each other when using constraints, though Power has a lower estimated risk. The figure should be compared to figure 5.1.2, the estimated risk do not behave in the same way as when no short selling is allowed.

¹⁰⁰ Jagannathan Ravi. Tongshu Ma. Risk Reduction in Large Portfolios: A Role for Portfolio Weight Constraints. September 19, 2001

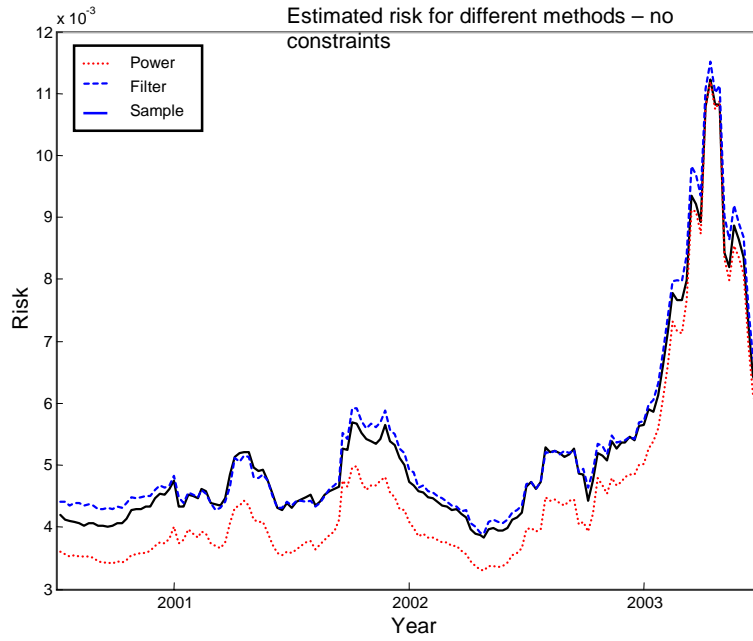


Figure 5.1.17. Estimated risk for Sample, Power mapping and Sample

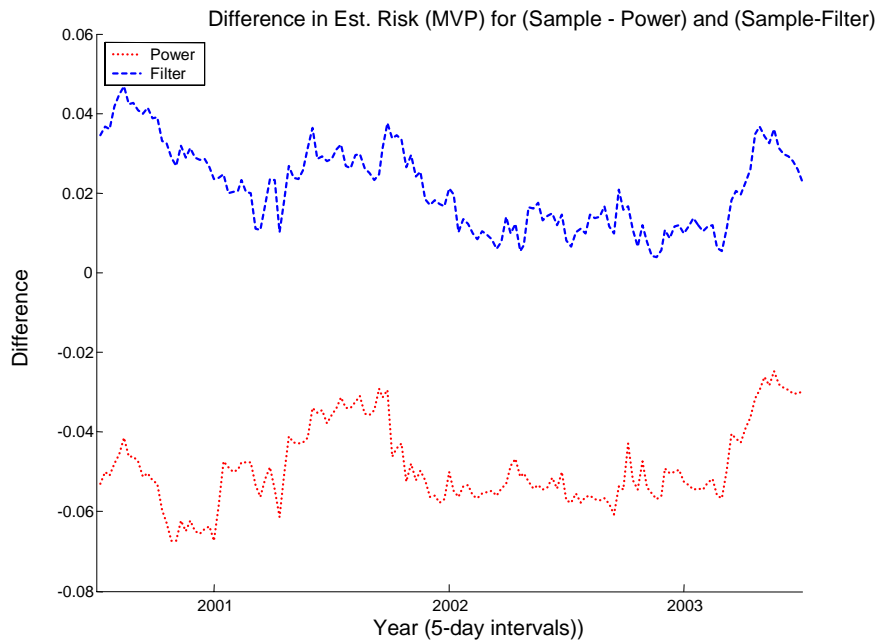


Figure 5.1.18. Difference in estimated risk between Power mapping/RMT-filtering and Sample for portfolio number four (MVP).

The graph below shows the difference in estimated risk for Power minus Sample for Portfolio(1) to Portfolio(3). The patterns are very similar to each other this since the expected returns for the portfolios are close to each others on the efficient frontier.

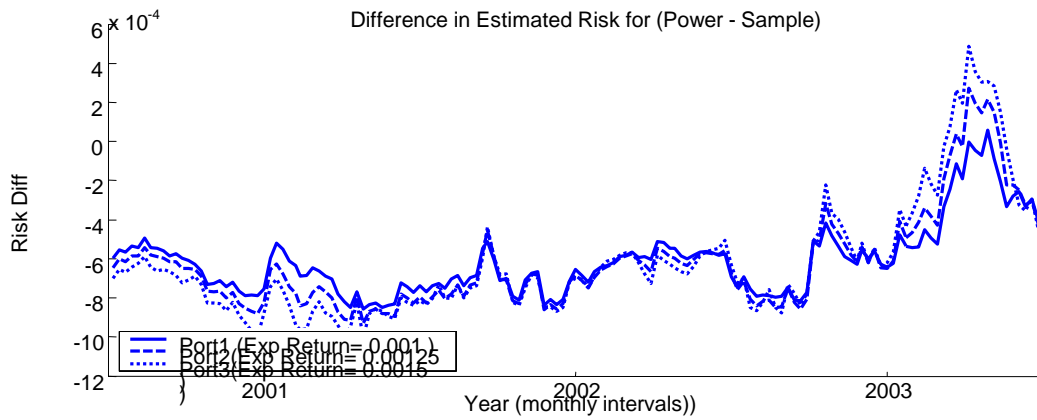


Figure 5.1.19. Difference in estimated risk between Power mapping and Sample for portfolios one to three.

The graph below shows the difference in estimated risk for Filter minus Sample for Portfolio(1) to Portfolio(3). The patterns are very similar, this since the expected returns for the portfolios are close to each others on the efficient frontier.

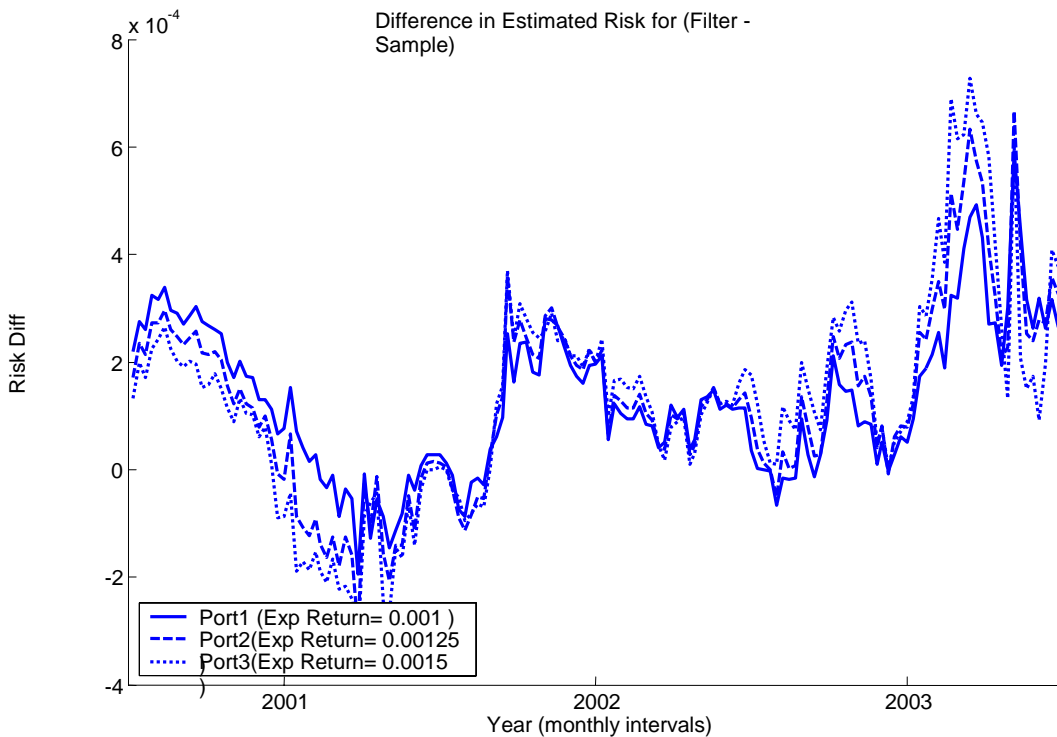


Figure 5.1.20. Difference in estimated risk between RMT-filtering and Sample for portfolios one to three.

The graph below shows the actual daily risk calculated at monthly intervals for number one i.e. Portfolio(1). One can see that the risks are almost the same for Sample, Power and Filter, perhaps with a minimal favor for the filtering techniques. This was not the case when using no constraints in the portfolio optimization process. The obvious reason for this is that short selling has been forbidden.

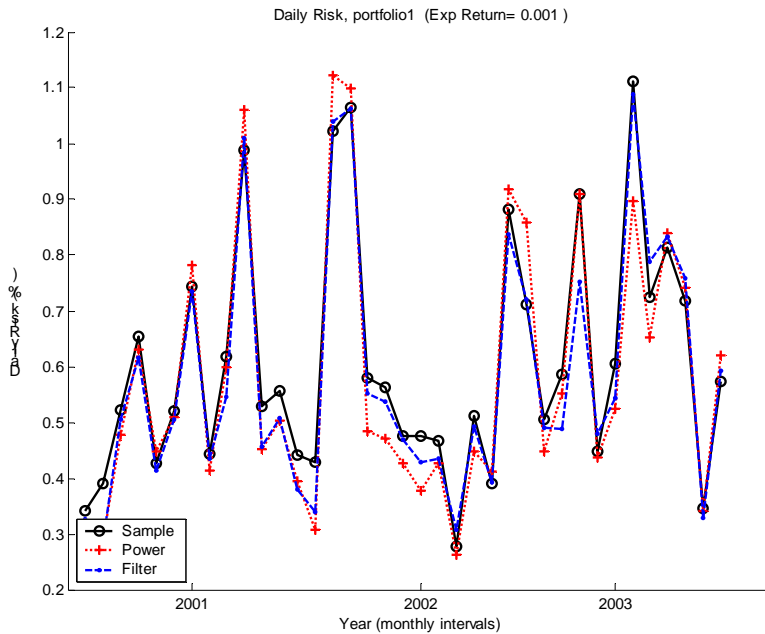


Figure 5.1.21. Daily risk for portfolio number one

The picture below shows the mean daily returns for the portfolio number one. The curves conform nicely, as one could expect after have seen the graph containing the daily risk above. Power mapping seems to produce more extreme spikes of return

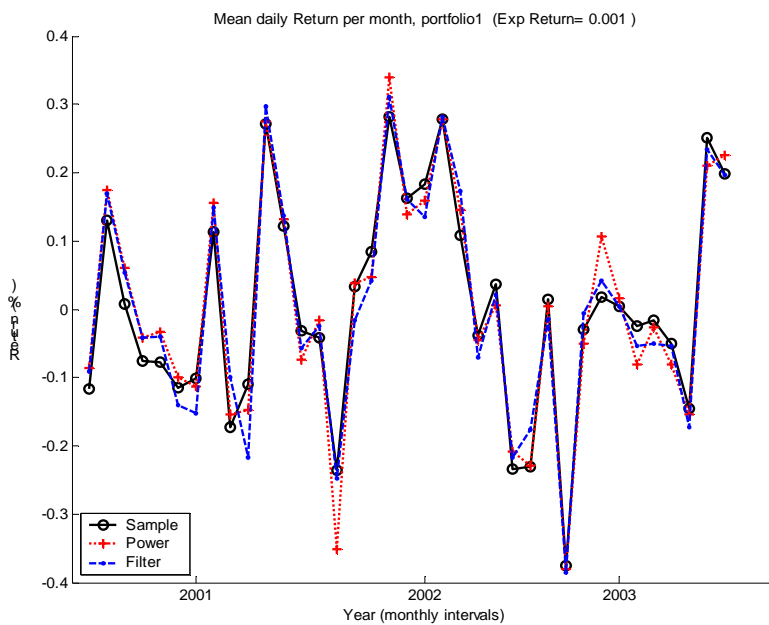


Figure 5.1.22. Mean daily return calculated each month for portfolio number one

The figure of the actual monthly return looks exactly the same as in the previous picture, except for the axis. Power Mapping manages to create the worst return at one location, where the market fluctuates strongly.

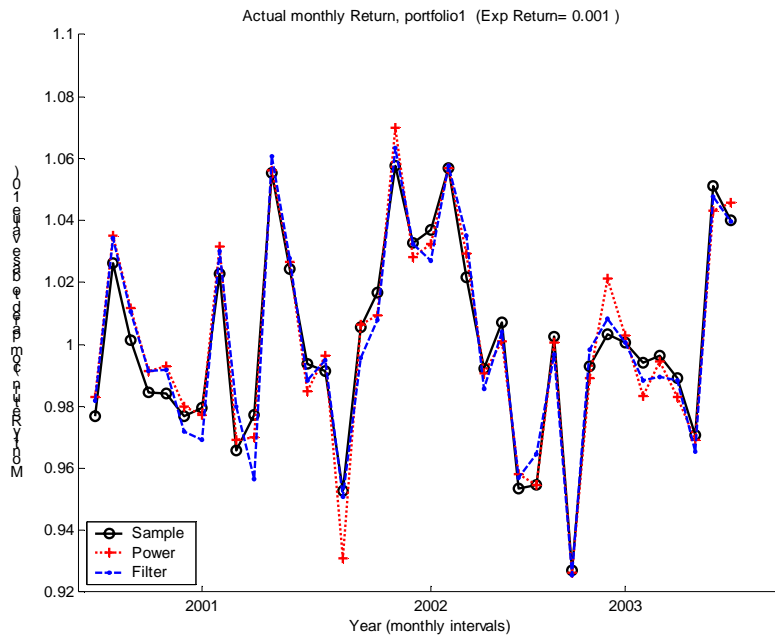


Figure 5.1.23. Actual monthly return for portfolio number one, constraints imposed.

For the MVP the risk is similar for many occasions. Slightly lower especially for RMT-filtering

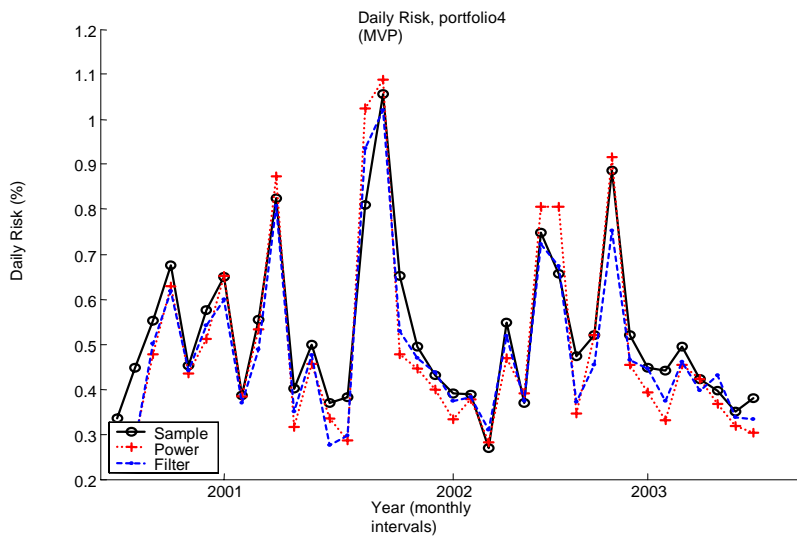


Figure 5.1.24. Daily risk for portfolio number four (MVP)

The mean daily return is fairly similar for all three methods as seen in the two figures below.

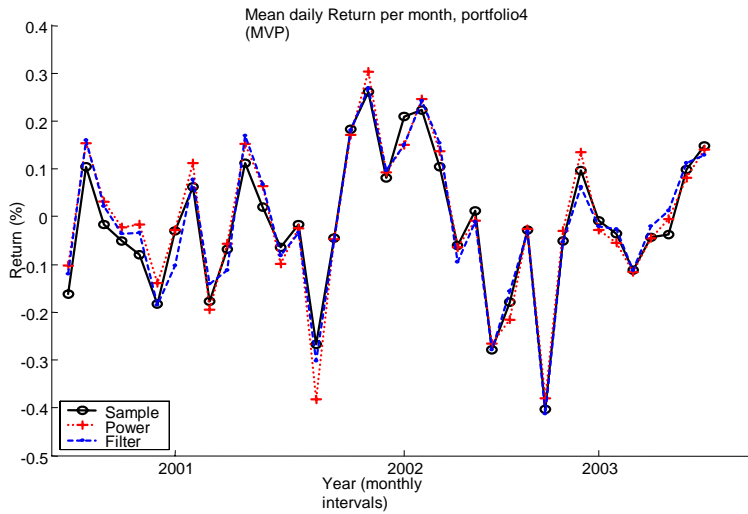


Figure 5.1.25. Mean daily returns calculated each month for portfolio number four (MVP)

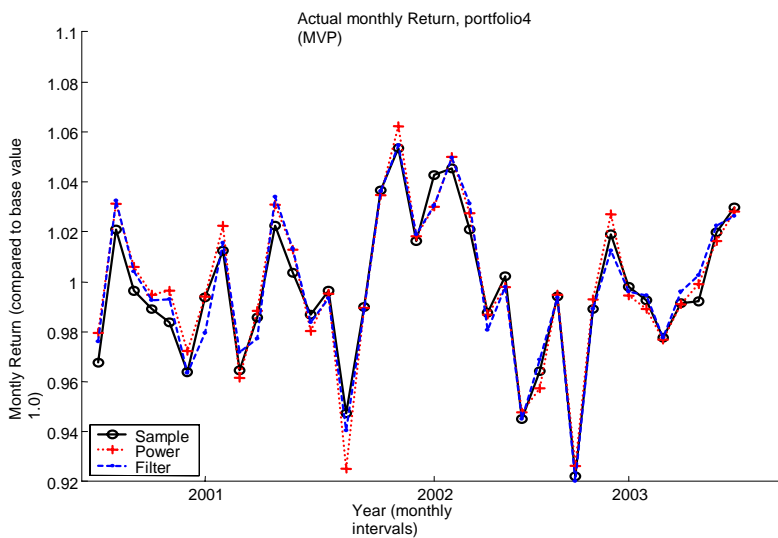


Figure 5.1.26. Actual monthly return for portfolio number four (MVP)

The graph below shows the weights for portfolio number one at the first period. Sample creates portfolios with larger weights than Sample and Power mapping, just as before. Though the weights don't fluctuate as much anymore since the constraints were implemented in the portfolio-creating model. As mentioned before we can see that a large number of the weights are close to zero. It's also very hard to see any general trends in the behavior of Power and Filter, possibly an indication that Power sometimes creates lower weights than the other methods.

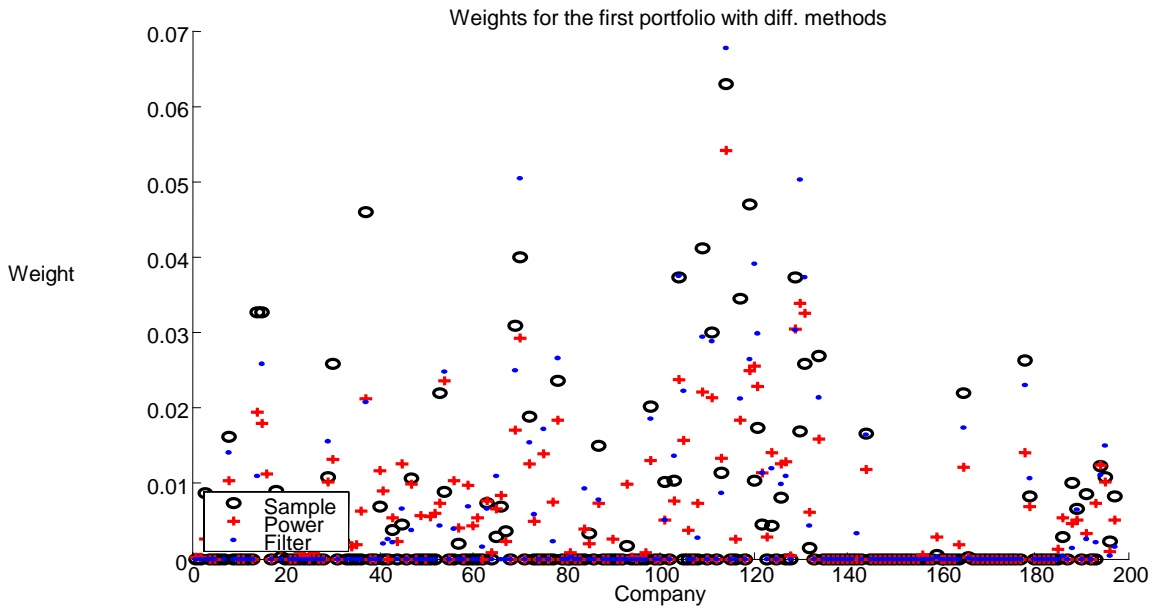


Figure 5.1.27. Overview of weights generated for portfolio number one by Sample, RMT-filtering and power mapping.

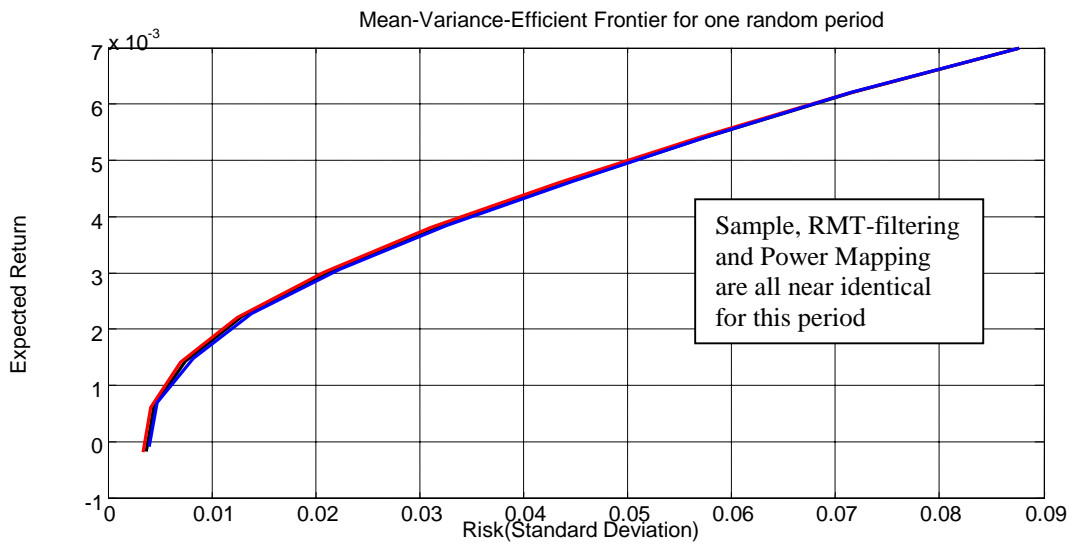


Figure 5.1.28. Efficient frontier for portfolio number one generated by Sample, power mapping and Sample. The period is randomly chosen.

The figure above shows the efficient frontier for one portfolio. As can be seen in the figure they are extremely similar for the three methods. This is probably an effect due to the imposing of constraints.

5.1.5 Correlation Structure

In the figure below the correlation matrices for Sample, Power Mapping and RMT-filtering are presented for the 3 years in the evaluation period. Every period is 5 days. As can be seen the correlation structure changes over the years. RMT-filtering also enhances the structure in the branches (but may lose some information in the process), Power Mapping have less extreme changes but reduce the noise slightly making the picture darker.

The RMT-filtering have a quite strong structure of lines, this is what the PM-filtering probably reduces making it better than the other methods (See Chapter 5.5 for more info about PM-filtering).

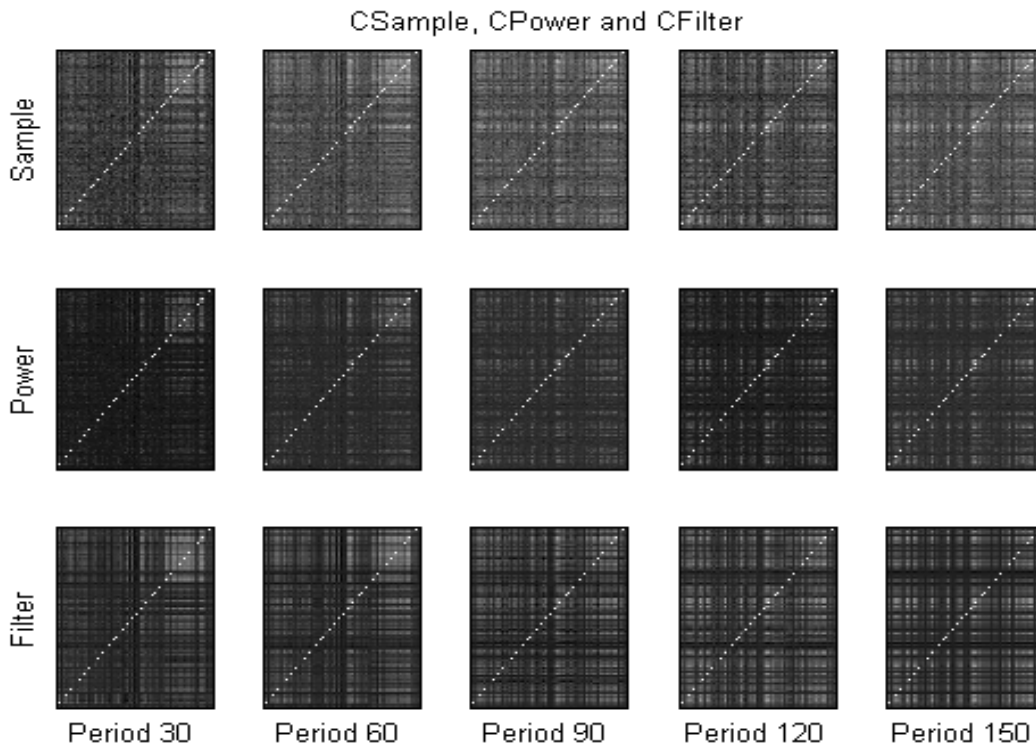


Figure 5.1.29. Correlation matrices generated by Sample, Power mapping and RMT-filtering.

5.1.6 Mean Values of the Correlation Matrices

One year correlation matrices created at periodic intervals, one period is 5 days. Some areas of interest exist in the picture. The sharp peak in the beginning corresponds to some very negative trends in the market; several stocks are going down fast at this occasion.

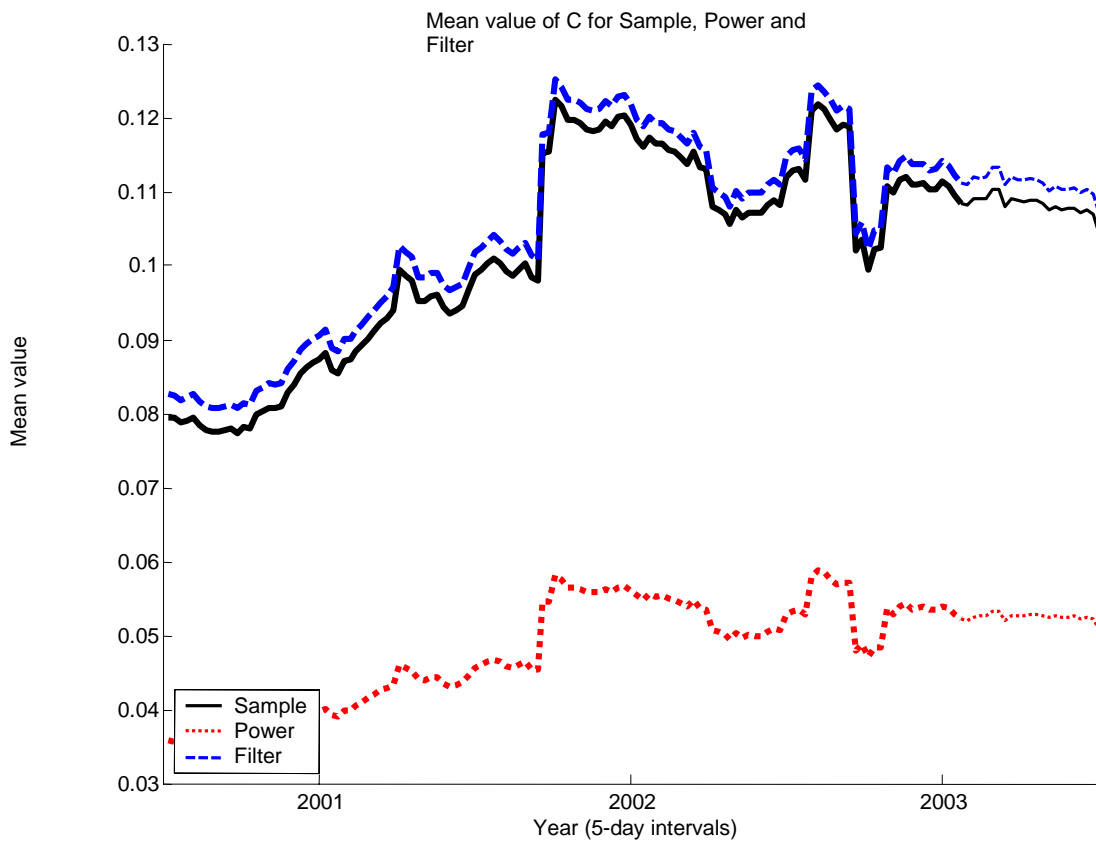


Figure 5.1.30. Mean value of correlation matrix elements for Sample, power mapping and RMT-filtering plotted versus time.

5.1.7 The Swedish Stock Data – 15 Years for 48 Companies

A quick calculation for the MVP was also performed on 48 companies for 15 years (Data1c). The results for Power Mapping and RMT-filtering were positive here as well.

Table 5.1.10. Results of portfolio number one

Total Actual Risk (Sample)= 0.77749% (daily) Yearly= 12.2932%
Total Actual Risk (Power)= 0.73072% (daily) Yearly= 11.5537%, (Vs Sample= **-6.0154%**)
Total Actual Risk (Filter)= 0.73053% (daily) Yearly= 11.5506%, (Vs Sample= **-6.0401%**)

5.2 A study on 48 U.S. Industry portfolios Stocks

5.2.1 Introduction

In this second subsection of the analysis we will examine the properties of portfolios created upon 48 U.S. industry portfolios (created upon stock returns of all companies listed on NYSE, AMEX and NASDAQ, for more detail regarding the data please see Chapter 4.1). In this chapter we also study how the length of the in-sample period (the period that we use to create the correlation matrix) affects the results.

5.2.2 Data Overview

These industry portfolios can from a statistical point of view be viewed in the same way as individual assets.¹⁰¹ The main difference would be that one can no longer talk about single companies (and their corresponding returns) or companies belonging to specific branches, one would have to refer to industry portfolios and to clusters of industry portfolios (i.e. clusters of branches). The first parts of the study are carried out in a similar way as the previous study based upon Swedish stock data returns.

The denotation is exactly the same as in the previous study concerning the Swedish stock data. Sample is used as an abbreviation for the initial unprocessed correlation matrix, the Power Mapping technique is denoted Power and the RMT-filtering technique is denoted Filter. In the table below Total Actual Risk is the actual risk that the portfolio creates from an investor's perspective. The Total Return Mean is the total actual mean return that the portfolio achieves. The portfolios consist of 48 U.S. industry portfolios ranging from 1973 to 2002.

¹⁰¹ Thomas Guhr Personal interview conducted 2003-09-02

5.2.3 Main Results – no Constraints Imposed

The tables below contain the main results of this part of the study; the optimizations have no constraints meaning negative weights are possible. The first portfolio is also the one closest to the MVP in aspect of “distance” on the efficient frontier. As can be seen in the following table both Power Mapping and RMT-filtering produce lower portfolio risks than the portfolio generated by Sample (the unfiltered original correlation matrix), though the difference between Sample and Filter is marginal. Especially Power Mapping creates a far lower risk than the other two methods. Though even the sun has spots, the total return (Total Return Actual) is the highest for Sample, quite high for RMT-filtering and lower for Power Mapping.

Total Return divided by total risk favors both Power Mapping and RMT-filtering which means that they both are capable of creating portfolios with a much better return/risk ratio than Sample.

The table below is calculated with an in-sample period of 5 years¹⁰².

It was possible to create larger steps in aspect of expected return for this industry portfolio than for the portfolios created upon the 197 Swedish companies. One explanation to this might be that the industry portfolios are created upon company stock returns, this creation of portfolios could have the impact that the portfolios are ill-behaved in a lesser extent compared to individual stock returns. Interesting to note is that RMT-filtering manages to create portfolios with little risk and superior return in this case, especially compared to Sample. A likely reason for this is that the branch structure is very strong in this portfolio which is essential for good results with RMT-filtering. This is the kind of environment that favors RMT-filtering.

Table 5.2.1. Results of portfolio number one, no constraints imposed

Portfolio 1 (Exp Return= 0.03) In-sample = 5 years

Total Actual Risk (Sample)= 8.6059% (monthly) ,Yearly= 29.8119%

Total Actual Risk (Power) = 5.539% (monthly), Yearly= 19.1875%, (Vs Sample= **-35.638%**)

Total Actual Risk (Filter) = 5.356% (monthly), Yearly= 18.5536%, (Vs Sample= **-37.7645%**)

Total Mean Return (Sample)= 1.0987% (monthly return)

Total Mean Return (Power)= 0.9891% (monthly return)

Total Mean Return (Filter)= 1.4011% (monthly return)

Total Mean Return div. Total Risk (Sample)= 0.12767 (monthly)

Total Mean Return div. Total Risk (Power) = 0.17857 (monthly) (Vs Sample= 39.867%)

Total Mean Return div. Total Risk (Filter) = 0.26159 (monthly) (Vs Sample= 104.8932%)

Total Return Actual (Sample)= 779.67% (whole period) Yearly= 9.08%

Total Return Actual (Power)= 1112.53% (whole period) Yearly=10.49 %

Total Return Actual (Filter)= 4152.99% (whole period) Yearly= 16.18%

Total Return Actual div. Total Risk (Sample)= 0.304

Total Return Actual div. Total Risk (Power) = 0.547 (Vs Sample= 79.9%)

Total Return Actual div. Total Risk (Filter)= 0.872 (Vs Sample= 186.8%)

¹⁰² 5 years (i.e. 60 months) is a commonly used in-sample period

The table below shows an even better picture for RMT-filtering but also slightly less improvement in risk.

Table 5.2.2. Results of portfolio number two, no constraints imposed

Portfolio 2 (Exp Return= 0.04) In-sample = 5 years

Total Actual Risk (Sample)= 10.0724% (monthly) ,Yearly= 34.892%
 Total Actual Risk (Power) = 6.9634% (monthly), Yearly= 24.1218%, (Vs Sample= -30.8672%)
 Total Actual Risk (Filter) = 6.7244% (monthly), Yearly= 23.294%, (Vs Sample= -33.2396%)

Total Mean Return (Sample)= 1.2058% (monthly return)
 Total Mean Return (Power)= 1.0381% (monthly return)
 Total Mean Return (Filter)= 1.5004% (monthly return)

Total Mean Return div. Total Risk (Sample)= 0.11972 (monthly)
 Total Mean Return div. Total Risk (Power) = 0.14907 (monthly) (Vs Sample= 24.523%)
 Total Mean Return div. Total Risk (Filter) = 0.22313 (monthly) (Vs Sample= 86.3852%)

Total Return Actual (Sample)= 690.57% (whole period) Yearly= 8.62%
 Total Return Actual (Power)= 970.77% (whole period) Yearly= 9.94%
 Total Return Actual (Filter)= 4363.49% (whole period) Yearly= 16.40%

Total Return Actual div. Total Risk (Sample)= 0.247
 Total Return Actual div. Total Risk (Power) = 0.412 (Vs Sample= 66.8%)
 Total Return Actual div. Total Risk (Filter) = 0.704 (Vs Sample= 185%)

The portfolio below with the highest expected return for the portfolio shows slightly less return for all methods. The risk is also lower here. One reason for the negative trend could be that the error our expected returns introduce distort the results.

Table 5.2.3. Results of portfolio number three, no constraints imposed

Portfolio 3 (Exp Return= 0.05) In-sample = 5 years

Total Actual Risk (Sample)= 11.4432% (monthly) ,Yearly= 39.6404%
 Total Actual Risk (Power) = 8.4494% (monthly), Yearly= 29.2696%, (Vs Sample= -26.1622%)
 Total Actual Risk (Filter) = 8.255% (monthly) ,Yearly= 28.5961%, (Vs Sample= -27.8611%)

Total Mean Return (Sample)= 1.2992% (monthly return)
 Total Mean Return (Power)= 1.103% (monthly return)
 Total Mean Return (Filter)= 1.5998% (monthly return)

Total Mean Return div. Total Risk (Sample)= 0.11354 (monthly)
 Total Mean Return div. Total Risk (Power) = 0.13054 (monthly) (Vs Sample= 14.9728%)
 Total Mean Return div. Total Risk (Filter) = 0.1938 (monthly) (Vs Sample= 70.6932%)

Total Return Actual (Sample)= 560.47 % (whole period) Yearly= 7.84%
 Total Return Actual (Power)= 816.88 % (whole period) Yearly= 9.27%
 Total Return Actual (Filter) = 4149.62 % (whole period) Yearly= 16.18%

Total Return Actual div. Total Risk (Sample)=0.198
 Total Return Actual div. Total Risk (Power) = 0.316 (Vs Sample=59.6 %)
 Total Return Actual div. Total Risk (Filter) = 0.566 (Vs Sample= 185.9%)

The MVP (Mean Variance Portfolio) shows the greatest improvements in risk for RMT-filtering and Power Mapping compared to Sample. In this case Power Mapping presents a slightly smaller risk than RMT-filtering.

Table 5.2.4. Results of portfolio number four (MVP), no constraints imposed

Portfolio 4 (MVP) In-sample = 5 years

Total Actual Risk (Sample)= 6.7344% (monthly) ,Yearly= 23.3287%
 Total Actual Risk (Power) = 3.8655% (monthly) ,Yearly= 13.3903% , (Vs Sample= **-42.6015%**)
 Total Actual Risk (Filter) = 3.9676% (monthly) ,Yearly= 13.7441% , (Vs Sample= **-41.0848%**)

Total Mean Return (Sample)= 0.90138% (monthly return)
 Total Mean Return (Power)= 0.8414% (monthly return)
 Total Mean Return (Filter)= 1.2383% (monthly return)

Total Mean Return div. Total Risk (Sample)= 0.13385 (monthly)
 Total Mean Return div. Total Risk (Power) = 0.21767 (monthly) (Vs Sample= 62.6283%)
 Total Mean Return div. Total Risk (Filter) = 0.31211 (monthly) (Vs Sample= 133.1862%)

Total Return Actual (Sample)= 646.44% (whole period) Yearly = 8.37%
 Total Return Actual (Power)= 889.19% (whole period) Yearly = 9.60%
 Total Return Actual (Filter)= 3080.66% (whole period) Yearly = 14.8%

Total Return Actual div. Total Risk (Sample)= 0.359
 Total Return Actual div. Total Risk (Power) = 0.716 (Vs Sample= 99.4%)
 Total Return Actual div. Total Risk (Filter) = 1.077 (Vs Sample= 200%)

The figure below shows the estimated risk for the MVP. It is notable that RMT-filtering and Power Mapping have higher values. It should be remembered that the expected return for the MVP is changing for all the portfolios, compared to Portfolio(1)-Portfolio(3) which all have fixed expected returns for the portfolio.

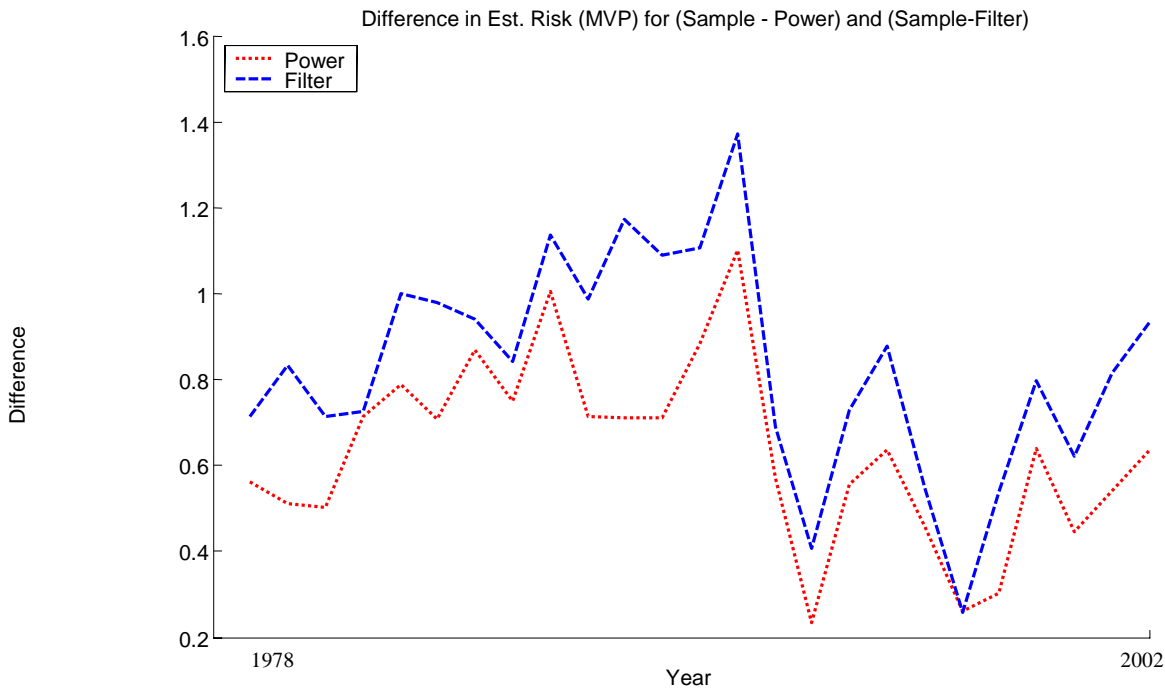


Figure 5.2.1. Difference between Sample and Power mapping/RMT-filtering in estimated risk for portfolio number four (MVP).

The figure below with the estimated risk for the Sample, Power Mapping and RMT-filtering shows the same structure that was visible in figure 5.1.2. This is encouraging and gives further reason to believe the previous discussion that Power Mapping and RMT-filtering creates information that is closer to reality, while Sample creates an incorrect optimization. It is also interesting to note that the Power Mapping and RMT-filtering are very close to each other; this could mean that they are converging to the “true” situation.

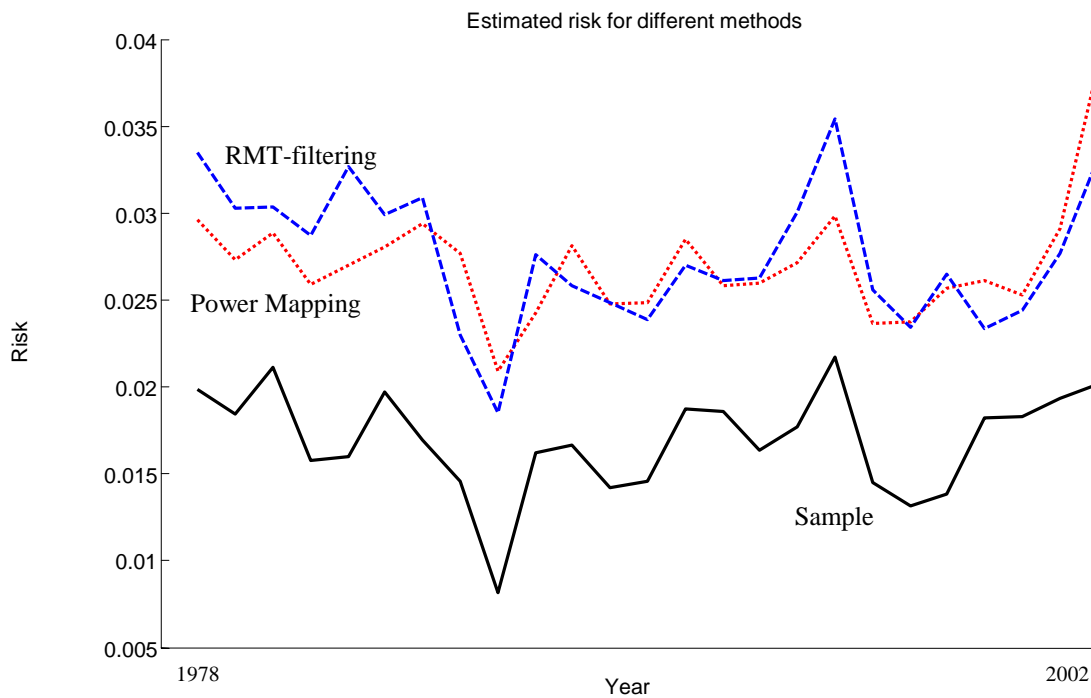


Figure 5.2.2. Estimated risk for Sample, Power mapping and RMT-filtering

The figure below shows an example of the efficient front for the first calculated period. As can be seen in the figure below the RMT-filtering and Power Mapping are estimated to have higher risk than Sample.

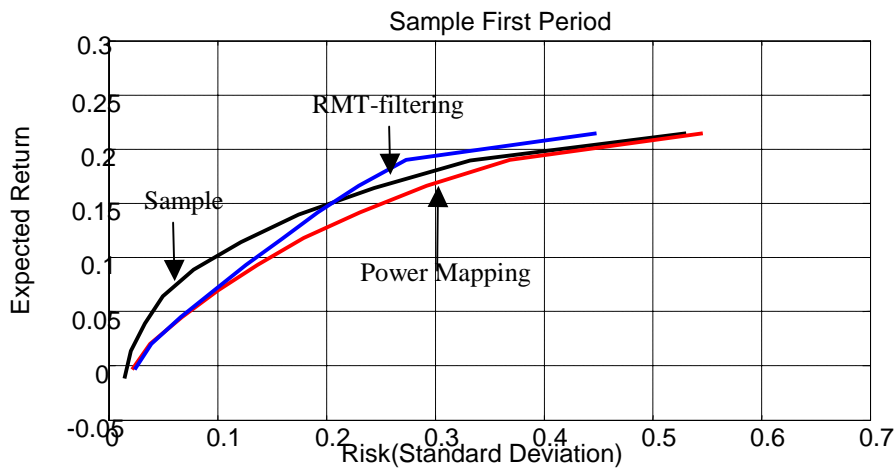


Figure 5.2.3. Efficient frontiers generated by Sample, Power mapping and RMT-filtering

The estimated risk for the Portfolio(1)-Portfolio(3) also shows the same general trend with higher estimated risk for Power Mapping.

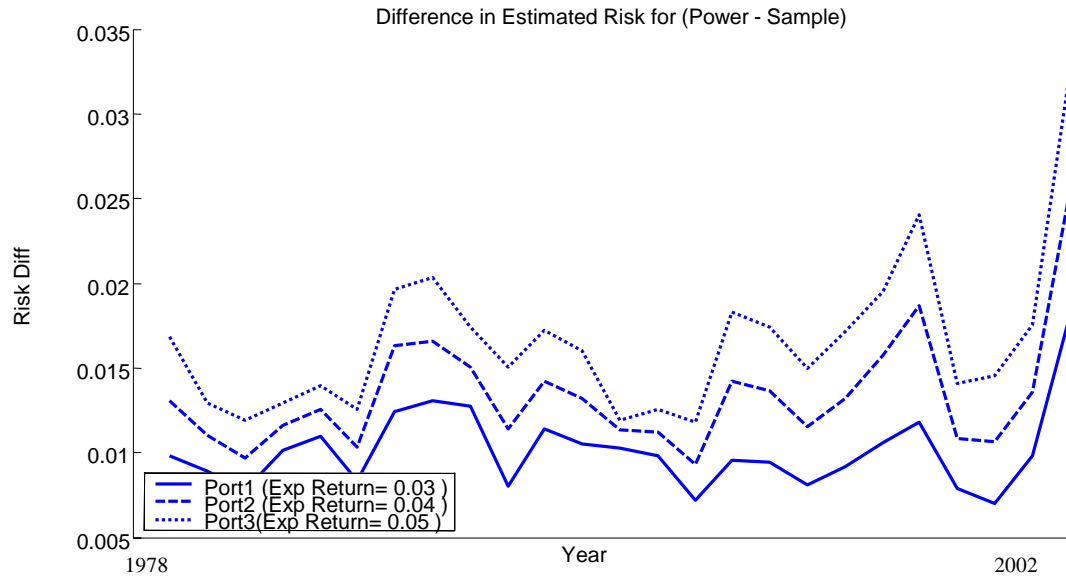


Figure 5.2.4. Difference in estimated risk between Power mapping and Sample for portfolios one to three.

The estimated risk in the next figure for RMT-filtering shows a more interesting pattern. The different portfolios are fluctuating heavily at some occasions but are stable at others. Turmoil in the market is a possible explanation. It seems like the expected return have less importance in the areas where the portfolios converge.

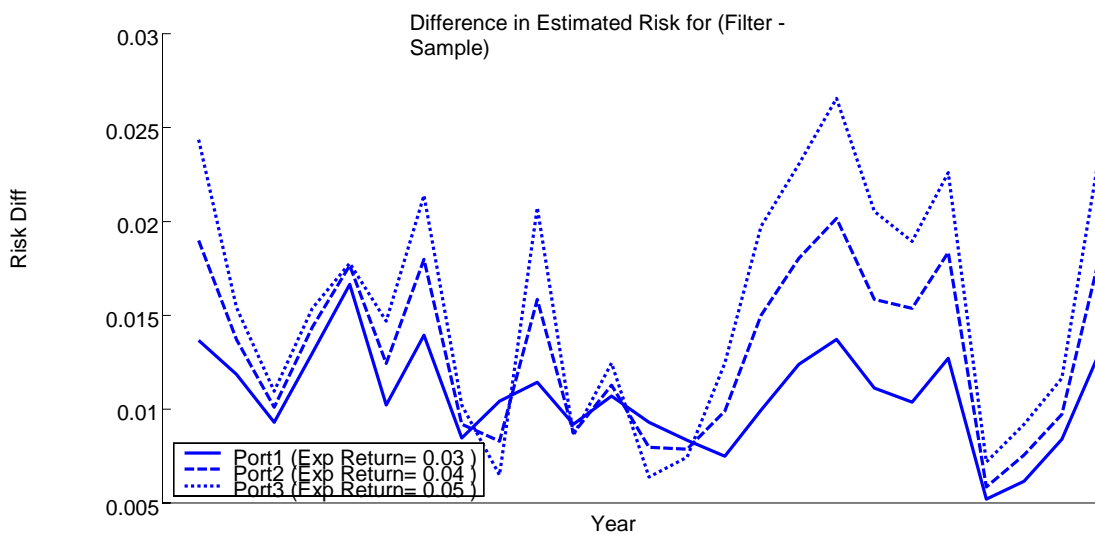


Figure 5.2.5. Difference in estimated risk between Power mapping and Sample for portfolios one to three.

In the figure below one can see the portfolio weights created for portfolio number one (expected return=0.02), where short positions are allowed (i.e. no constraints introduced). One can observe that Sample creates more extreme weights than Power Mapping and RMT-filtering, this is consistent with the results and discussion from the study when using Swedish stock data. There also seems to be a pattern that RMT-Filter produce the most restrictive distributions, Power Mapping creates fairly similar weights at some of the more extreme occasions.

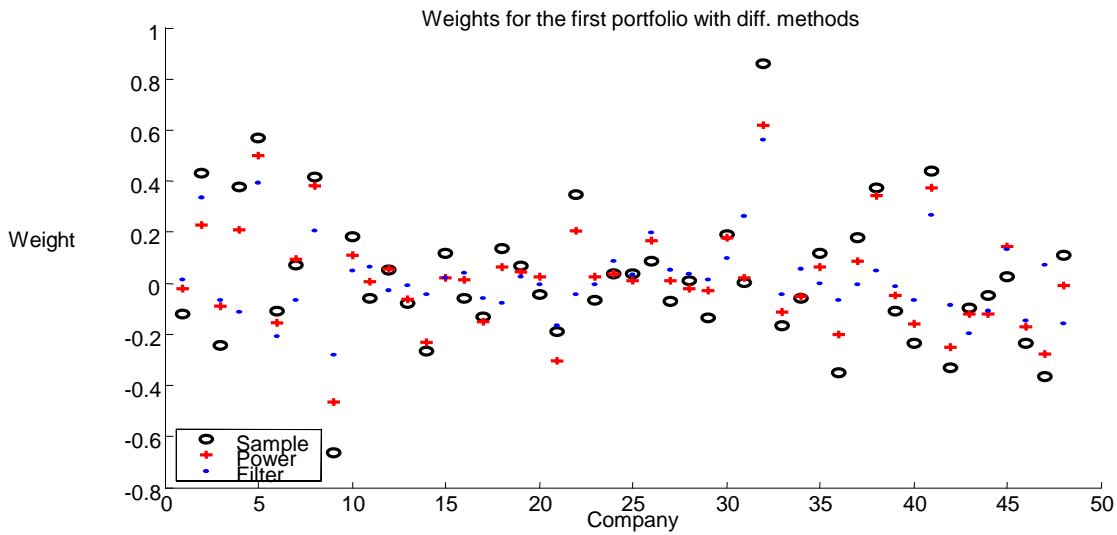


Figure 5.2.6. Weights generated by Power mapping, RMT-filtering and Sample for portfolio number one.

The daily risk for portfolio(1) presented in the figure below shows results similar to the ones we have seen earlier, in this occasion Sample always shows a higher risk than for Power Mapping and RMT-filtering. This difference is larger than the one we have seen in Chapter 5.1. One reason for this is that we are working with monthly data and this figure has portfolios created yearly, while the data in Chapter 5.1 had a higher frequency and the portfolios were created every day. Yet another reason is of course that Power Mapping and RMT-filtering are working with data which is more suited for them with a clearer branch structure.

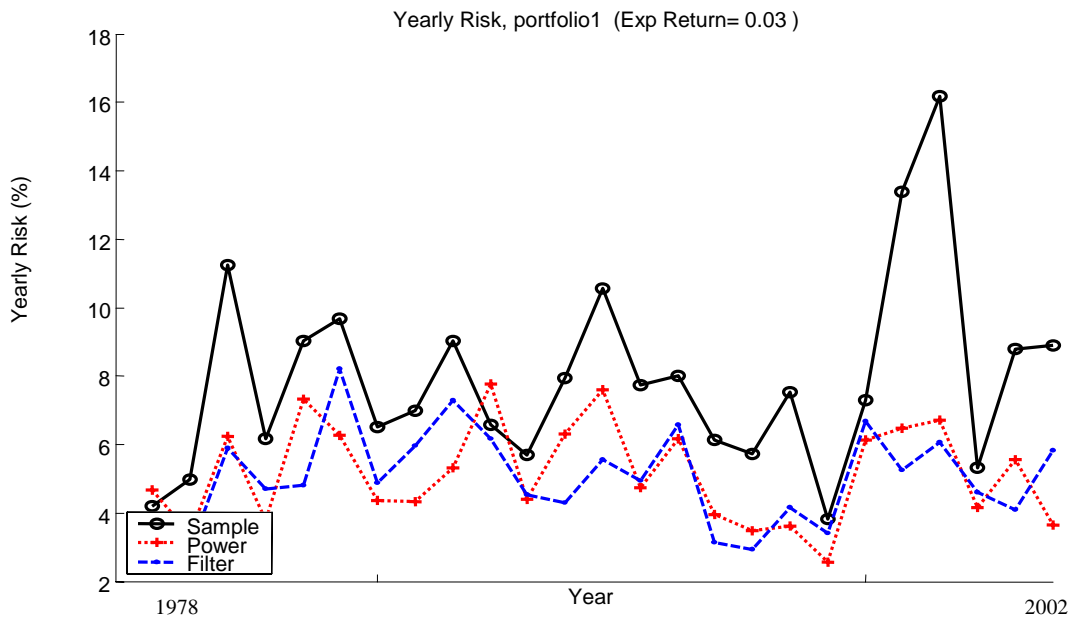


Figure 5.2.7. Yearly risk generated by Power mapping, RMT-filtering and Sample for portfolio number one.

The figure with the mean yearly return presented shows some interesting information also. In Chapter 5.1 the mean return was most of the time lower for Power Mapping and RMT-filtering, here it's often the opposite.

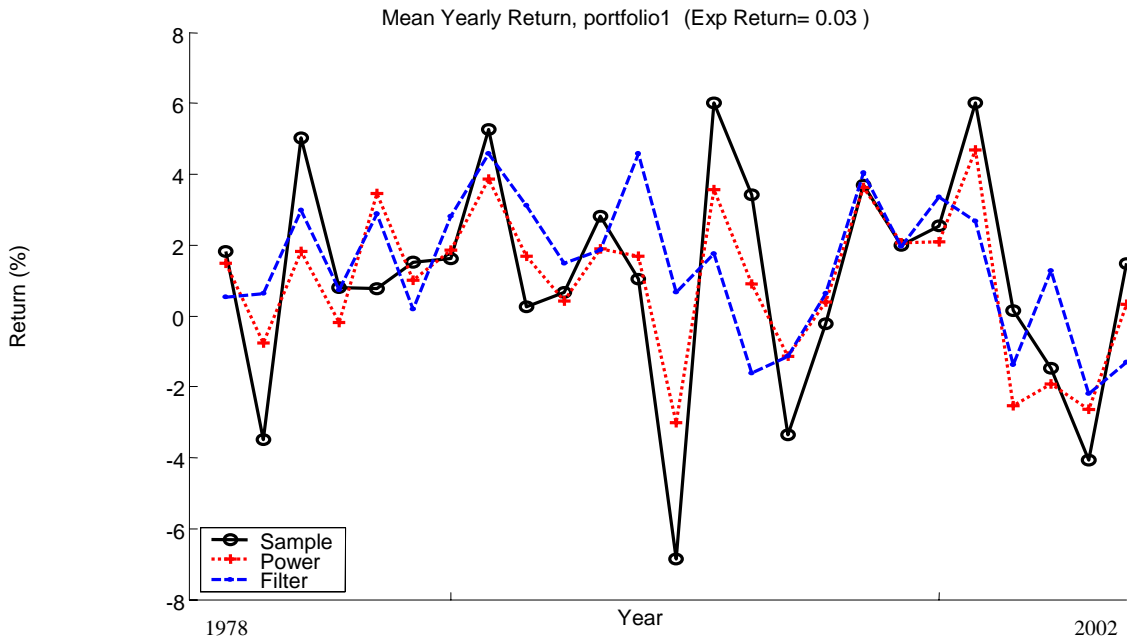


Figure 5.2.8. Mean Yearly return generated by Power mapping, RMT-filtering and Sample for portfolio number one.

The actual return in the figure below is near identical to the mean return, unless you look at the axis. Interesting is that the actual return for some years are very high, mostly for Sample. In the middle of the figure a very large drop in actual yearly return is visible; this is an effect due to a recent major stock market crash. Very interestingly though is that the RMT-filtering actually manages to stay afloat with little loss. One has to remember that this situation has no constraints and negative weights are permitted. When a stock market crash happens there is little chance that negative weights will be possible in reality, meaning that these very positive results are hardly realistic.

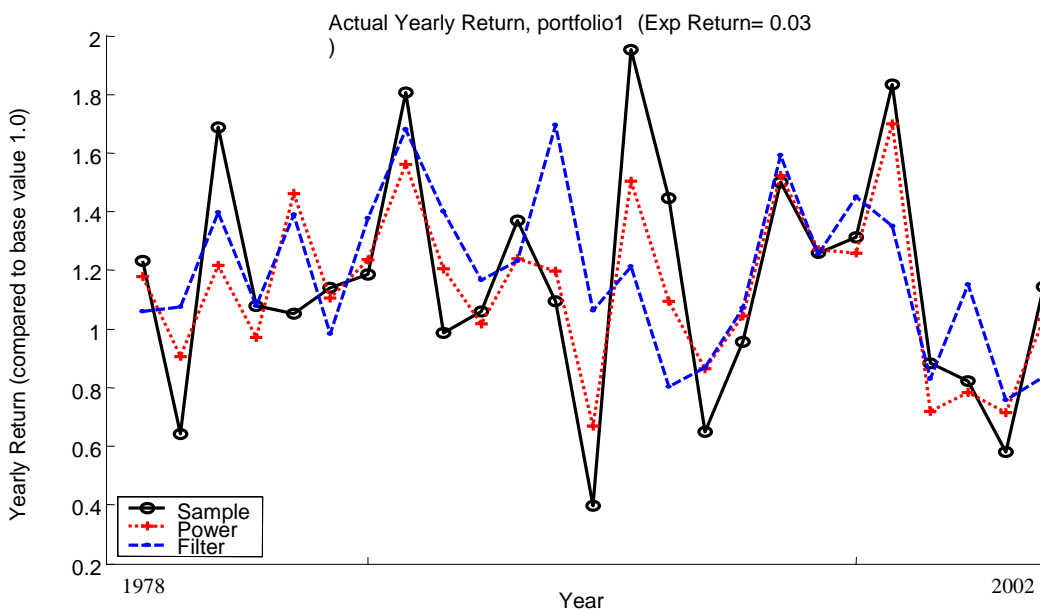


Figure 5.2.9. Actual Yearly return generated by Power mapping, RMT-filtering and Sample for portfolio number one.

The figure below shows the third Portfolio (Portfolio(2) was moved to Appendix, it shows similar trends like the third portfolio). Portfolio(3). The risk is obviously much higher in the later years for Sample. The reason for this is not really clear, but is probably due to negative market trends and the inability for Sample to decrease the risk effectively. It could also mean that the expected returns are introducing stronger fluctuations now when the expected return for the portfolio is higher.

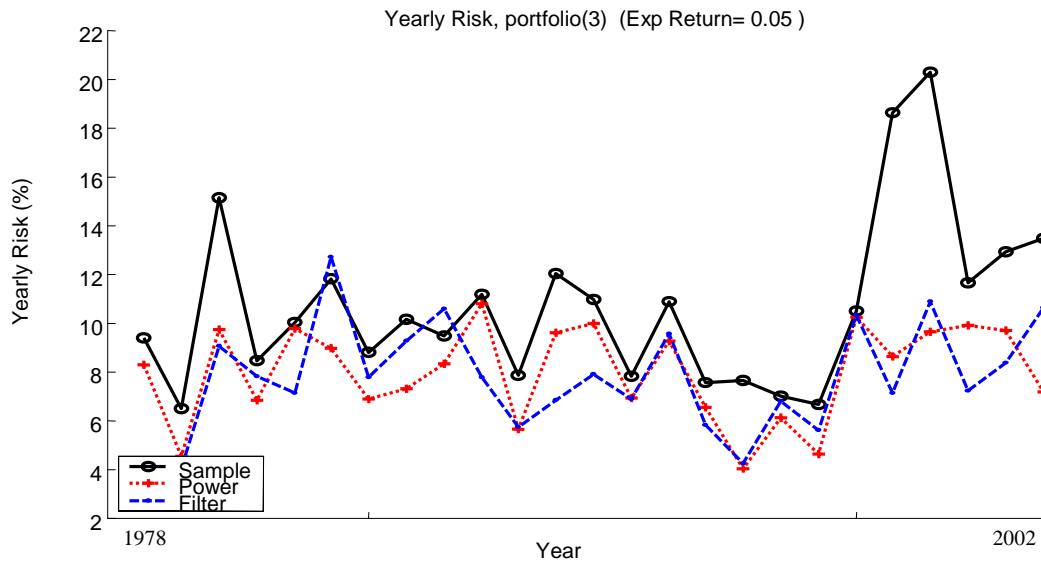


Figure 5.2.10. Yearly actual risk for power mapping, RMT-filtering and Sample for portfolio number three.

The actual monthly return in the picture below (mean return was cut out since it is practically identical) show some interesting patterns. The Power Mapping and sometimes RMT-filtering are surprisingly close to Sample at occasions.

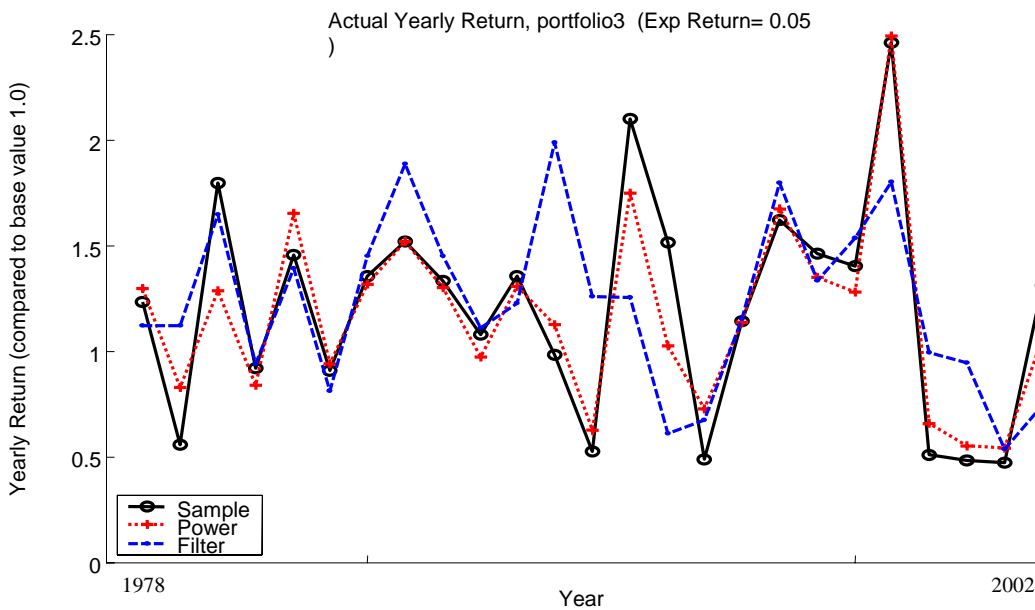


Figure 5.2.11. Yearly actual return for power mapping, RMT-filtering and Sample for portfolio number three.

The portfolio for the MVP shows some very interesting patterns. We have a severely jagged rim especially for Sample. Power Mapping and RMT-filtering is generally following Sample but have much lower values. We are probably seeing the different fluctuations in the market in the 25 years.

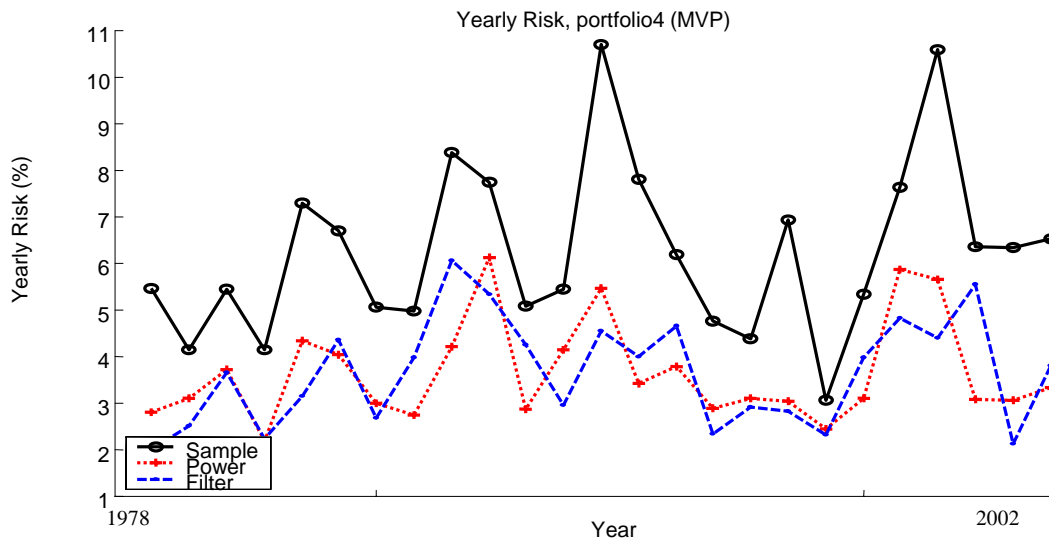


Figure 5.2.12. Yearly actual risk for power mapping, RMT-filtering and Sample for portfolio number four, the minimum variance portfolio (MVP).

The mean yearly returns show the same sharp decline as we have seen for the other portfolios (the mean return was moved to Appendix).

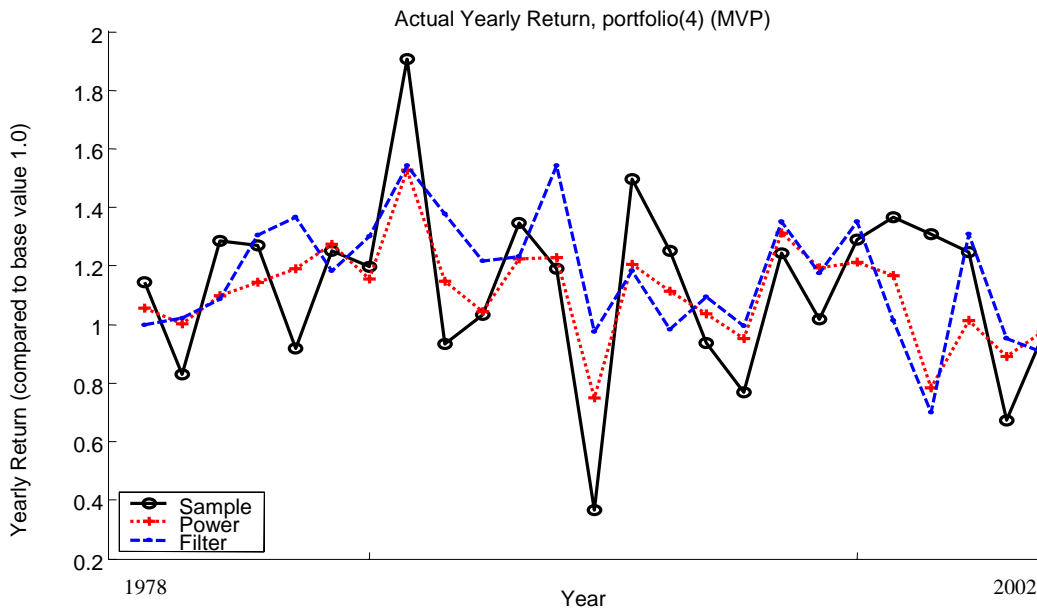


Figure 5.2.13. Yearly actual return for power mapping, RMT-filtering and Sample for portfolio number four, the minimum variance portfolio (MVP).

5.2.4 Main Results – Constraints Imposed

To keep the calculations within reasonable levels, only the MVPs are studied with constraints imposed. The in-sample period used is 5 years, just like before. In this situation Power Mapping is still better than Sample but RMT-filtering is slightly worse which is surprising. As previously explained this should be a promising situation for RMT-filtering. One problem here is that we do not allow short selling which will eliminate many possibilities to reduce risk effectively, this could possibly eliminate some of the benefits of the noise reducing methods. Comparing the risks below to the same risks obtained when not imposing constraints shows a decrease in risk for all three methods. This indicates that not only Sample suffers from estimation error, since both Power Mapping and RMT-filtering have lower risks, which indicates that we start with matrices contaminated with estimation error. The reason to this error could either lie in the fact that we use a too long time period (5 years) to estimate the covariance matrix i.e. including information that no longer is valid, or in the fact that we do not use that many time steps T compared to the number of assets N when creating the matrix ($Q=1.25$). Though increasing the time steps leads to the problem that one will have to use longer time period or switch to data with higher frequency. If Q approaches the value one (1) the determination of the covariance matrix is likely to be noisy, therefore the correlation matrix become random to a certain extent¹⁰³. This is something one has to be careful about since this kind of noise introduced affects the correlation matrix's smallest eigenvalues, which corresponding eigenvectors determine the least risky portfolio¹⁰⁴.

The information about the return has a smaller font in the table since we only optimize for risk, we still leave the results since there might be weak patterns to be found.

If we look at the portfolio with the in-sample period of 5 years RMT-filtering produces a solution that is slightly worse than for the previous portfolio. This is to be expected since we now have less time steps, and it is very important for RMT-filtering to have as clear a structure in the eigenvalue spectrum as possible.

Table 5.2.5. Results for portfolio number four (MVP), constraints imposed.

Portfolio 4 (MVP) In-sample = 5 years (Starting point is the same)

Total Actual Risk (Sample)= 3.561% (daily) ,Yearly= 12.3357%

Total Actual Risk (Power)= 3.5282% (daily) ,Yearly= 12.2221%, (Vs Sample= **-0.92095%**)

Total Actual Risk (Filter)= 3.6504% (daily) ,Yearly= 12.6452%, (Vs Sample= 2.5087%)

Total Mean Return (Sample)= 0.90539% (daily return)

Total Mean Return (Power)= 0.94603% (daily return)

Total Mean Return (Filter)= 0.96811% (daily return)

Total Mean Return div. Total Risk (Sample)= 0.25425 (daily)

Total Mean Return div. Total Risk (Power)= 0.26813 (daily) (Vs Sample= 5.4597%)

Total Mean Return div. Total Risk (Filter)= 0.26521 (daily) (Vs Sample= 4.3105%)

¹⁰³ Laloux L. et al, 1998, Noise dressing of financial correlation matrices.

¹⁰⁴ Ibid.

Total Return Actual (Sample)= 647.35% (whole period) Yearly = 8.37%
 Total Return Actual (Power)= 725.1% (whole period) Yearly = 8.81%
 Total Return Actual (Filter)= 761.16% (whole period) Yearly = 8.99%

Total Return Actual div. Total Risk (Sample)= 0.678
 Total Return Actual div. Total Risk (Power)= 0.72 (Vs Sample= 6.20%)
 Total Return Actual div. Total Risk (Filter)= 0.711 (Vs Sample= 4.87%)

5.2.5 Correlation structure

The pictures below show 10-year correlation matrices for Sample, Power and RMT-filter at the top row and 5-year correlation matrices at the bottom row chosen for the first period. The pictures are very similar which is somewhat surprising. As in Chapter 5.1 the RMT-filtering creates a very clear picture; probably losing some information in the process but keeping the essential main trends.

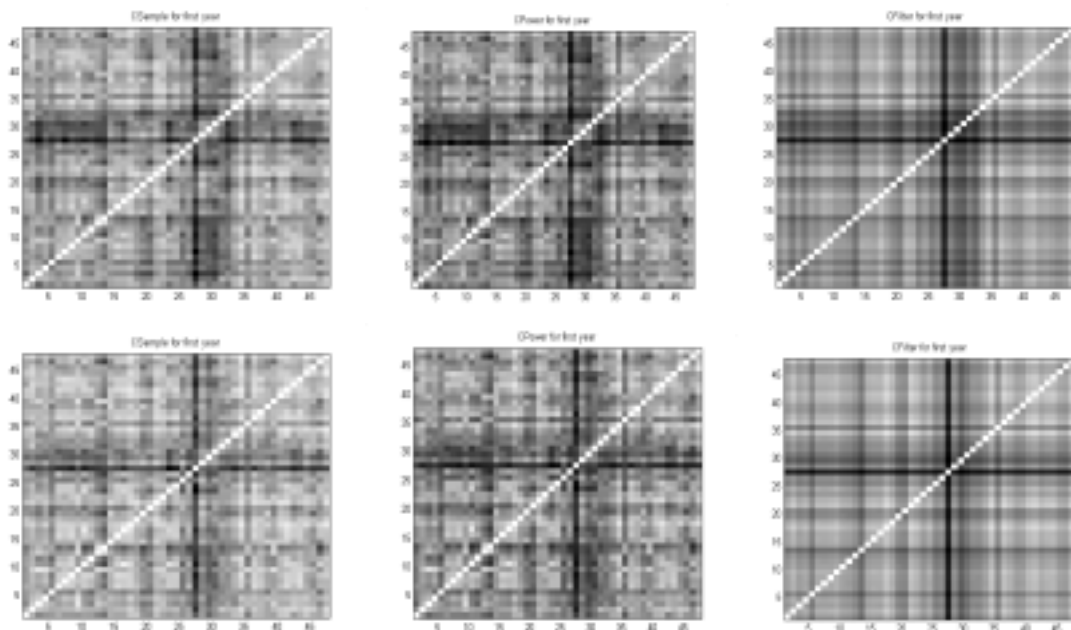


Figure 5.2.14. Illustrating 10-year correlation matrices for Sample, Power mapping and RMT-filtering. At the top row and 5-year correlation matrices at the bottom row chosen for the first period.

In the two following figures we study how the correlation structure changes over time. One Period is one year. If one looks closer at the industry around the “dark cross”-region, one finds that these industries are Gold, Coal, Oil and Mining. Especially Gold has a very low correlation of approximately 0.3 while most of the others have values around 0.6-0.7 at this time period. This is not surprising since this period include the 1973 oil crisis which might cause these stocks to move differently than the rest of the market.

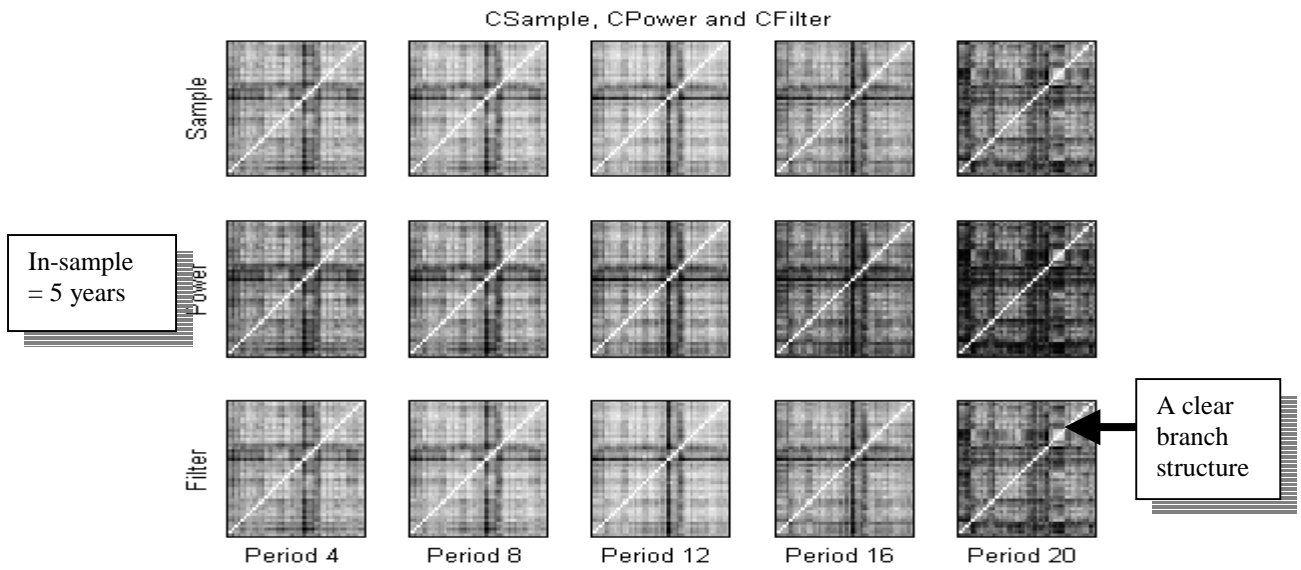


Figure 5.2.15. Changing correlation structures, using 5 years in-sample period.

One notable difference between the 10-year correlation matrices and the 5-year correlation matrices is that there seem to be smaller structures in the 10-year correlation matrices. This is unexpected, more trends should be found in 10 years than in 5 years though. Another interesting detail that is clearest in the last 10-year RMT-Filter matrix is that the smaller branch is completely left out.

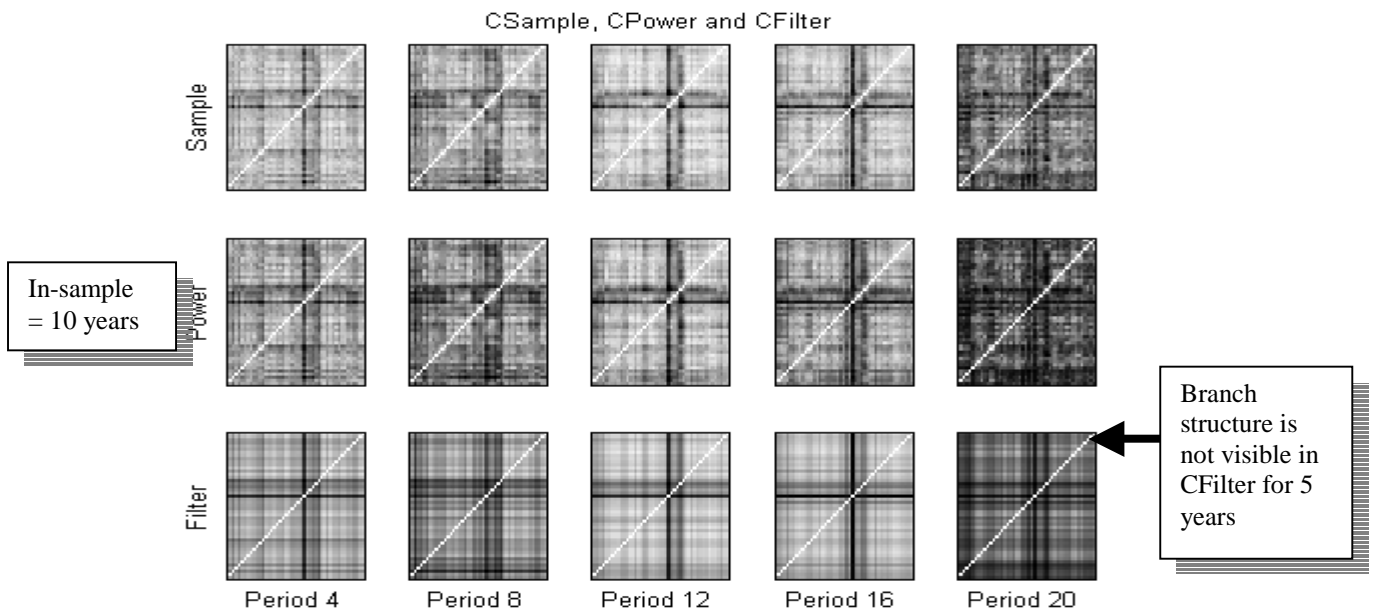


Figure 5.2.16. Changing correlation structures, using 10 years in-sample period.

One would expect to see correlation matrices that change more over the years, but these are industry portfolios and the correlations will be stronger¹⁰⁵ and more stable over time than individual stocks, something that the pictures clearly show.

¹⁰⁵ Thomas Guhr, personal interview conducted at 2003-09-02

The eigenvalue spectra for Sample below shows some large eigenvalues indicating several strong correlations (associated with branches). This will lead to better performance of the RMT-filtering which also has been shown in the previous tests. In these tests with an in-sample period of 5 years both RMT-filtering and Power Mapping also produces satisfying results.

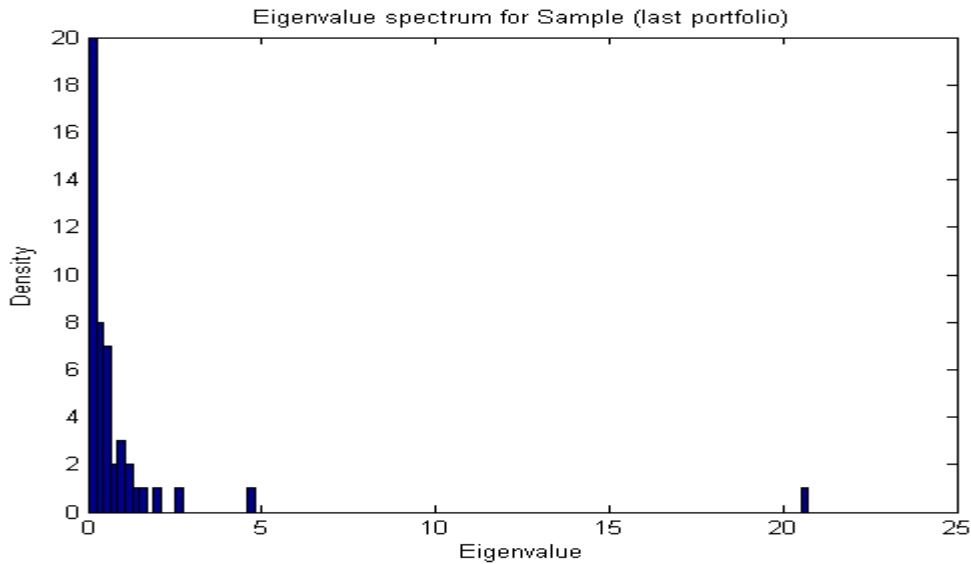


Figure 5.2.17. Eigenvalue spectra for sample correlation matrix

In the figure above we can see several clear branches. In the tests in Chapter 5.1 we had no clear branch structure like in this case. Here several large eigenvalues are visible which makes it more similar to the ideal situation created by Noh's model¹⁰⁶.

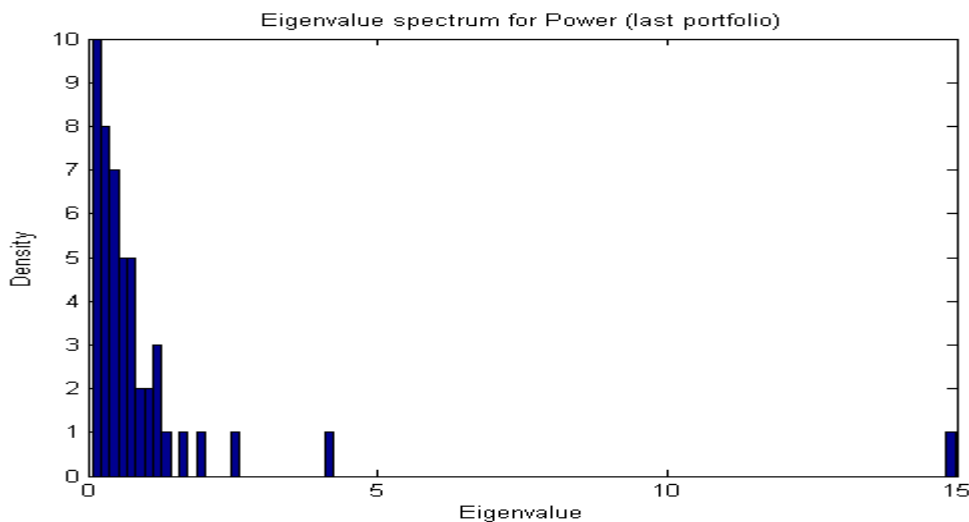


Figure 5.2.18. Eigenvalue spectra for Power mapping correlation matrix

¹⁰⁶ Thomas Guhr, personal interview conducted at 2003-09

5.2.5 Mean values of the correlation matrices

Below are several figures describing the mean value of the correlation matrices. The mean value of the correlation matrices can show general trends. The 48 U.S. industry portfolios showed some clear peaks at stock market crashes for example and also indicated that there were less general correlations after a stock market crash¹⁰⁷. The picture below shows the 10-year correlation matrices. As can be seen the trend is very smooth since we are using data from such a wide period. “Ghost”-effects are also visible in the data¹⁰⁸.

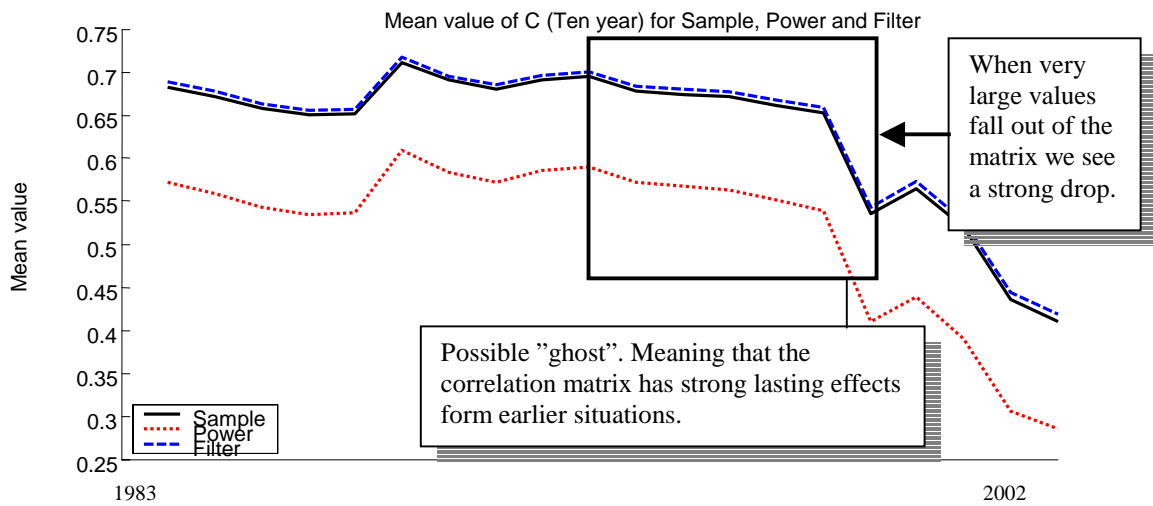


Figure 5.2.19. Mean value of correlation matrix elements (created under a 10 year in-sample period) for Sample, power mapping and RMT-filtering plotted versus time.

Below the mean value of the 5-year correlation matrices is visible. The trends are not as smooth as before.

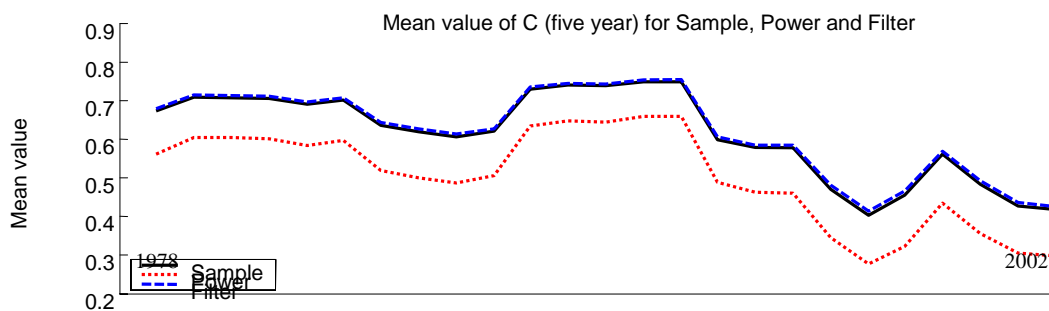


Figure 5.2.20. Mean value of correlation matrix elements (created under a 5 year in-sample period) for Sample, power mapping and RMT-filtering plotted versus time.

¹⁰⁷ Andersson, P. Öberg, A. 2003. Noise Reducing Methods for Correlation Matrices New Techniques to Minimize Risk in Portfolio Management?

¹⁰⁸ Alexander, C., Market models, 2001, John Wiley & Sons

The picture below with the one-year correlation matrices shows the market trend clearly.

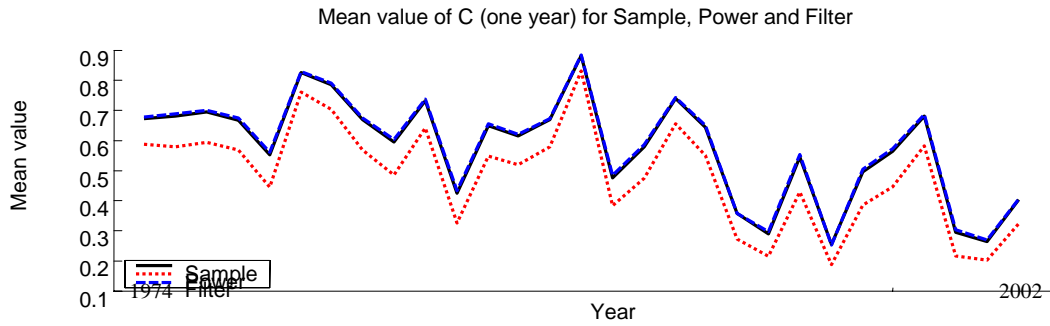


Figure 5.2.21. Mean value of correlation matrix elements (created under a one year in-sample period) for Sample, power mapping and RMT-filtering plotted versus time.

The picture below shows the mean value of the correlation matrices for a very long time period, from 1931-01 to 2002-12, almost 70 years of data for 43 different industries. Mean values of Sample, Power Mapping and RMT-filtering are also shown. This is the same data that was used for the previous industry analysis, but with a slightly lower number of industries since some industries didn't exist around 1930. The occasions with very high mean value indicate a period where many companies follow each other strongly. The clearest periods with this feature are naturally stock market crashes. Some stock markets are clearly visible, for example the stock market crash at 1987 (as seen in the picture below). More stock market crashes are easy to spot with the same method.

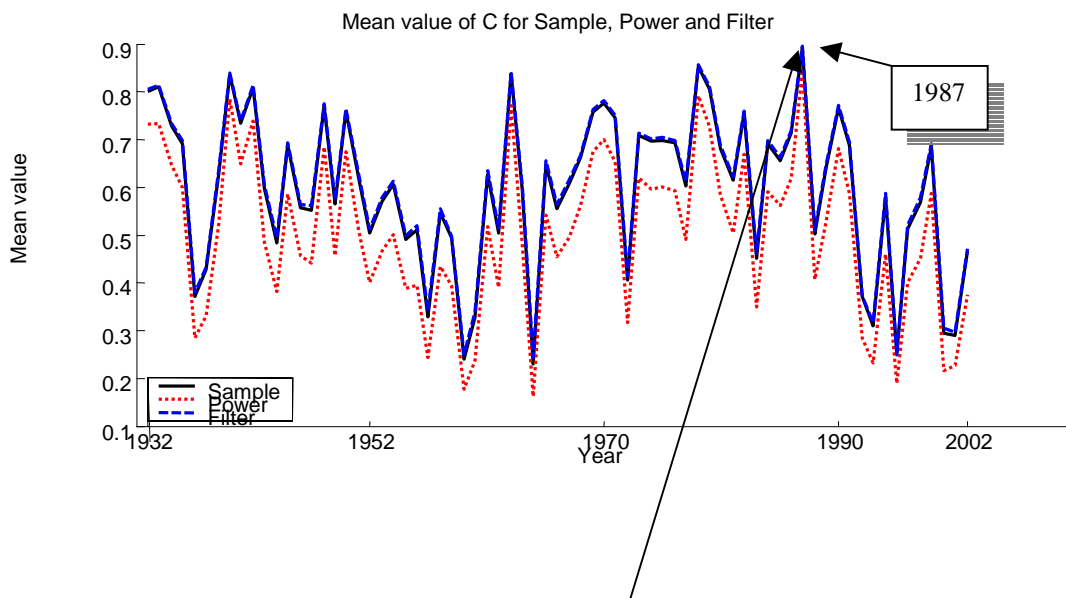




Figure 5.2.22. Illustration of the increase of positive correlations during a market crises

¹⁰⁹ Low Risk <http://www.lowrisk.com/crash/crashcharts.htm> (2003-08-10)

5.3 Standard deviation tests

In this part of the analysis the large portfolio containing 197 Swedish companies was divided into 7 smaller portfolios, each containing 28 companies. The sorting into the 7 smaller portfolios was based upon the increasing standard deviation (ranging from 0.01418606 0.31843002) for the individual company. The idea was to study if different risk preferences (in the portfolios creation process) would affect the abilities of Power Mapping and RMT-filtering.

The figure below shows the improvement of Power Mapping and RMT-filtering compared to Sample. Generally the improvement seems to be greater in the last portfolios with the largest risk but the portfolios are small thus limiting the accuracy of the data.

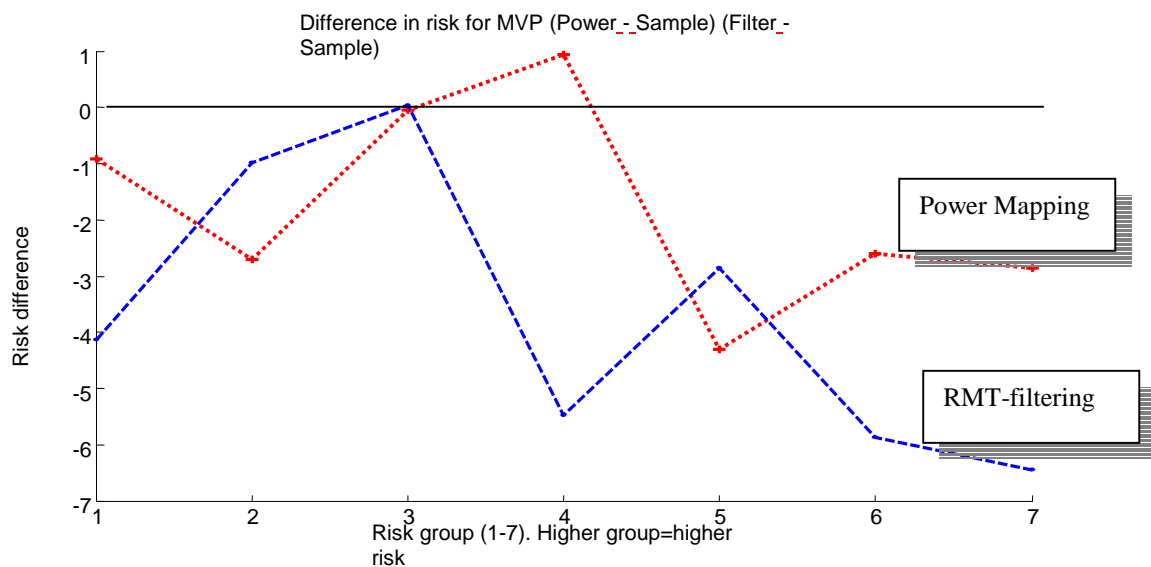


Figure 5.3.1. Difference in risk between Power mapping/RMT-filtering and Sample for portfolio number four (MVP).

Table 5.3.1. Results for all seven portfolios sorted after standard deviation.

Portfolio – STD Group 1

Total Actual Risk (Sample)= 0.72619% (daily) ,Yearly= **11.482%**
 Total Actual Risk (Power)= 0.69623% (daily) ,Yearly= 11.0083%, (Vs Sample= **-4.1255%**)
 Total Actual Risk (Filter)= 0.71956% (daily) ,Yearly= 11.3772%, (Vs Sample= **-0.9126%**)

Portfolio – STD Group 2

Total Actual Risk (Sample)= 0.7412% (daily) ,Yearly= 11.7195%
 Total Actual Risk (Power)= 0.73396% (daily) ,Yearly= 11.605%, (Vs Sample= **-0.97693%**)
 Total Actual Risk (Filter)= 0.72116% (daily) ,Yearly= 11.4026%, (Vs Sample= **-2.7041%**)

Portfolio – STD Group 3

Total Actual Risk (Sample)= 0.82281% (daily) ,Yearly= 13.0097%
 Total Actual Risk (Power)= 0.82306% (daily) ,Yearly= 13.0138%, (Vs Sample= **0.031414%**)
 Total Actual Risk (Filter)= 0.82236% (daily) ,Yearly= 13.0026%, (Vs Sample= **-0.054428%**)

Portfolio – STD Group 4

Total Actual Risk (Sample)= 1.1348% (daily) ,Yearly= 17.9427%
Total Actual Risk (Power)= 1.0727% (daily) ,Yearly= 16.9605%, (Vs Sample= **-5.4743%**)
Total Actual Risk (Filter)= 1.1455% (daily) ,Yearly= 18.1118%, (Vs Sample= **0.9422%**)

Portfolio – STD Group 5

Total Actual Risk (Sample)= 1.5058% (daily) ,Yearly= 23.8092%
Total Actual Risk (Power)= 1.4627% (daily) ,Yearly= 23.1278%, (Vs Sample= **-2.862%**)
Total Actual Risk (Filter)= 1.441% (daily) ,Yearly= 22.7843%, (Vs Sample= **-4.3047%**)

Portfolio – STD Group 6

Total Actual Risk (Sample)= 1.7432% (daily) ,Yearly= 27.5622%
Total Actual Risk (Power)= 1.6409% (daily) ,Yearly= 25.9442%, (Vs Sample= **-5.8704%**)
Total Actual Risk (Filter)= 1.6977% (daily) ,Yearly= 26.8435%, (Vs Sample= **-2.6077%**)

Portfolio – STD Group 7

Total Actual Risk (Sample)= 2.5675% (daily) ,Yearly= **40.5962%**
Total Actual Risk (Power)= 2.4022% (daily) ,Yearly= 37.9816%, (Vs Sample= **-6.4405%**)
Total Actual Risk (Filter)= 2.4942% (daily) ,Yearly= 39.4374%, (Vs Sample= **-2.8546%**)

5.4 Image Filter Test

A short test was also performed using an image filter; this filter is a contrast filter. It cuts off all values lower than a certain threshold and rescales the remaining values so that the smallest values are as low as the lowest one before the cut-off procedure¹¹⁰. In this test 5% and 10% were used. Since Power Mapping visually is similar to a contrast filter the idea was born to try what effect an image filter would have. The data used were the same as used in Chapter 5.1 (197 Swedish companies). As can be seen in both tables below the method did not give better results than Sample or Power Mapping. This is not so surprising since we are actually cutting away data from the correlation matrix. It would be more interesting to study more image filters that don't use as harsh methods to change the data. Since we are focusing on the ability to reduce risk we do not look at the return for this method.

Table 5.4.1. Results for portfolio number four (MVP) implementing image filter 10% cut

Portfolio 4 (MVP) Image filter 10% cut-of value

Total Actual Risk Sample= 0.9315% (daily) Yearly= 14.7283%
Total Actual Risk Power= 0.52033% (daily) Yearly= 8.2271%, (Vs Sample= -44.141%)
Total Actual Risk Image Filter= 2.0788% (daily) Yearly= 32.8688%, (Vs Sample= 123.1671%)

Table 5.4.2. Results for portfolio number four (MVP) implementing image filter 5% cut

Portfolio 4 (MVP) 5% cut-of value

Total Actual Risk Sample= 0.9315% (daily) Yearly= 14.7283%
Total Actual Risk Power= 0.52033% (daily) Yearly= 8.2271%, (Vs Sample= -44.141%)
Total Actual Risk Image Filter= 1.7374% (daily) Yearly= 27.4705%, (Vs Sample= 86.5145%)

¹¹⁰ The mean value of detail that is lost when the data is converted to 16 bit data for image filtering is 1.533×10^{-5}

5.5 “PM-Filtering” – a new Tool for Improving Correlation Matrices?

5.5.1 Introduction

A “new hybrid method” that merged Power Mapping and RMT-filtering was developed by Andersson & Öberg¹¹¹ combined both methods. The new hybrid method was called “PM-filtering” (as in Power-Mapping-Filtering) The idea was that since RMT-filtering creates a promisingly clear picture of the correlations it still seemed to introduce a pattern of stripes which shouldn’t be there. To suppress mostly these stripes the Power Mapping was used with a higher q-value (1.7, the normal value is 1.5). The results were promising with noticeably lower risk for the portfolio. One of the main goals of this thesis has been to study if the PM-filtering still holds with more extensive data. Due to time limits only portfolios with no constraints have been studied (when constraints are imposed there are also smaller differences in the results, which might lead to less accuracy).

5.5.2 Results for 197 Swedish Companies (no Constraints)

The first table below shows the normal methods, Sample, RMT-filtering and Power Mapping. In-sample is one year (250 data points).

Table 5.5.1. Results of the minimum variance portfolio (MVP)

Portfolio 4 (MVP) NORMAL SETTINGS, NO PM-filtering

Total Actual Risk (Sample)= 0.9315% (daily) ,Yearly= 14.7283%

Total Actual Risk (Power)= 0.52033% (daily) ,Yearly= 8.2271%, (Vs Sample= **-44.141%**)

Total Actual Risk (Filter)= 0.50039% (daily) ,Yearly= 7.9119%, (Vs Sample= **-46.281%**)

Total Mean Return (Sample)= 0.03715% (daily return)

Total Mean Return (Power)= 0.0079063% (daily return)

Total Mean Return (Filter)= 0.024293% (daily return)

Total Mean Return div. Total Risk (Sample)= 0.039882 (daily)

Total Mean Return div. Total Risk (Power)= 0.015195 (daily) (Vs Sample= -61.9009%)

Total Mean Return div. Total Risk (Filter)= 0.048549 (daily) (Vs Sample= 21.7294%)

Total Return Actual (Sample)= 27.9% (whole period), Yearly = 8.54%

Total Return Actual (Power)= 5.03% (whole period), Yearly = 1.65%

Total Return Actual (Filter)= 18.86% (whole period), Yearly = 5.92%

Total Return Actual div. Total Risk (Sample)= 0.580

Total Return Actual div. Total Risk (Power)= 0.200 (Vs Sample= **-65.5%**)

Total Return Actual div. Total Risk (Filter)= 0.748(Vs Sample= 29.0%)

RMT-filtering in this normal setup (with no constraints) creates the portfolios with the lowest risk.

¹¹¹ Andersson, P. Öberg, A. 2003. Noise Reducing Methods for Correlation Matrices
New Techniques to Minimize Risk in Portfolio Management?

Table 5.5.2. Results of the minimum variance portfolio (MVP) using q=1.5

Portfolio 4 (MVP) PM-filtering, q=1.5

TotalActualRiskPMFilter= 0.52852% (daily) ,Yearly= 8.3566%, (Vs Sample= **-43.2618%**)
 TotalReturnMeanPMFilter= 0.0077325% (daily return)
 TotalReturnMeanDivRiskPMFilter= 0.014631 (daily) (Vs Sample= -63.3158%)
 TotalReturnActualPMFilter= 4.86% (whole period) Yearly = 1.59%
 TotalReturnActualDivRiskPMFilter= 0.190 (Vs Sample= -67.2%)

The PM-filtering obviously didn't produce as good results as the other methods on daily data. For PM-filtering with a q-value of 1.5 the improvement in risk compared to Sample was 43.3%, compared to Power Mapping that had 44.1% and RMT-filtering that had 46.3%.

Table 5.5.3. Results of the minimum variance portfolio (MVP) using q=1.7

Portfolio 4 (MVP) PM-filtering, q=1.7

TotalActualRiskPMFilter= 0.55001% (daily) ,Yearly= 8.6964%, (Vs Sample= **-40.9547%**)
 TotalReturnMeanPMFilter= 0.002858% (daily return)
 TotalReturnMeanDivRiskPMFilter= 0.0051962 (daily) (Vs Sample= -86.9711%)
 TotalReturnActualPMFilter= 1.01% (whole period) , Yearly = 0.336%
 TotalReturnActualDivRiskPMFilter= 0.0386(Vs Sample= -93.3%)

With the q-value of 1.7 even lower results emerged for RMT-filtering, with a lower risk of 41.0%. One explanation could be that there is actually very little information in the correlation matrix for the Swedish stock market; many companies seem to have little correlation to other companies. If we use two filtering methods there is a risk that we eliminate too much information.

5.5.3 Results for 48 U.S. Industries (no constraints)

Below the results from Power Mapping and RMT-filtering are displayed.

Table 5.5.4. Results of the minimum variance portfolio (MVP)

Portfolio 4 (MVP) INDUSTRIES . normal settings

Total Actual Risk (Sample)= 7.475% (monthly) ,Yearly= 25.8941%
 Total Actual Risk (Power)= 3.8655% (monthly) ,Yearly= 13.3903%, (Vs Sample= **-48.2882%**)
 Total Actual Risk (Filter)= 3.9676% (monthly) ,Yearly= 13.7441%, (Vs Sample= **-46.9218%**)

Total Mean Return (Sample)= 0.70701% (monthly return)
 Total Mean Return (Power)= 0.8414% (monthly return)
 Total Mean Return (Filter)= 1.2383% (monthly return)

Total Mean Return div. Total Risk (Sample)= 0.094583 (monthly)
 Total Mean Return div. Total Risk (Power)= 0.21767 (monthly) (Vs Sample= 130.1382%)
 Total Mean Return div. Total Risk (Filter)= 0.31211 (monthly) (Vs Sample= 229.9861%)

Total Return Actual (Sample)= 243.98% (whole period) , Yearly =5.07%
 Total Return Actual (Power)= 889.19% (whole period) , Yearly = 9.6%
 Total Return Actual (Filter)= 3080.66% (whole period) , Yearly = 14.8%

Total Return Actual div. Total Risk (Sample)= 0.196
 Total Return Actual div. Total Risk (Power)= 0.716 (Vs Sample= 265%)
 Total Return Actual div. Total Risk (Filter)= 1.07(Vs Sample= 446%)

To reduce calculation time fewer q-values were used. With a q-value of 1.5 the PM-Filtering is better than the other methods here as well.

Table 5.5.5. Results of the minimum variance portfolio (MVP), q=1.5

INDUSTRIES PM-FILTERING q=1.5

Total_Actual_Risk_PMFILTER= 3.6392% (monthly) ,Yearly= 12.6067%, (Vs Sample= **-51.3146%**)
Total_Return_Mean_PMFILTER= 1.1832% (monthly return)
Total_Return_Mean_Div_RiskPMFILTER= 0.32511 (monthly) (Vs Sample= 243.7347%)
Total_Return_Actual_PMFILTER= 2700.82% (whole period), Yearly = 14.3%
Total_Return_Actual_Div_Risk_PMFILTER=) 1.13 (Vs Sample= **476%**)

In the table below the q-value of 1.7 provided the best result.

Table 5.6.5. Results of the minimum variance portfolio (MVP), q=1.7

Portfolio 4 (MVP) q=1.7

Total Actual Risk PMFILTER= 3.5823% (monthly) ,Yearly= 12.4093%, (Vs Sample= **-52.0768%**)
Total Return Mean PMFILTER= 1.1643% (monthly return)
Total Return Mean Div Risk PMFILTER= 0.32501 (monthly) (Vs Sample= 243.6214%)
Total Return Actual PMFILTER= 2563.28% (whole period), Yearly = 14.0%
Total Return Actual Div. Risk PMFILTER= 1.128 (Vs Sample= **502%**)

5.5.4 Conclusions for PM-Filtering

For a q-value of 1.7 PM-filtering is even better than with a q-value of 1.5. This is interesting, since a q-value of 1.7 also provided the best result in previous studies performed by Andersson & Öberg. One similar property that the 48 industry portfolio shares with the previous result (for the 22 U.S. companies) is that we are using a lower frequency on the data. Perhaps monthly data works better if the q-value is higher in the PM-Filtering since the data is more limited.

To further study hybrid methods would be interesting since the results are more than promising. It would perhaps be possible to mix existing methods to create even better ones.

5.6 The Mean Values of the Columns of the Correlation Matrix

5.6.1 Introduction

In the previous chapters we have studied the mean values of the complete correlation matrices. In this chapter we will look at a somewhat refined measurement, the mean values of the correlation matrices' columns. What does this value really describe? The column mean value would actually describe how much the company is correlated to other companies in the matrix; the higher the value the more similar it follows other companies. This analysis is but a brief exploration looking for possible patterns in the "column-mean-values".

5.6.2 Data Description

The stock data used spans over 13 years with an in-sample period of 3 years (750 values). With the data for the Swedish companies (Data1c) 48 were included since they were the only ones that had data for such a long period. The long in-sample period was used since we wanted the correlation matrices to have more time steps which should create a clearer correlation structure. The Sample matrix is used in the studies, unless otherwise mentioned. The correlation matrices are calculated for every day so we get as high frequency as possible.

5.6.3 Column Mean Values

Below a large picture of the correlation matrix is shown. The white area shows an example of a columns-mean-value that we are studying. As can be seen in the picture some column-mean-values are high indicating that they are in branches that behave similarly.

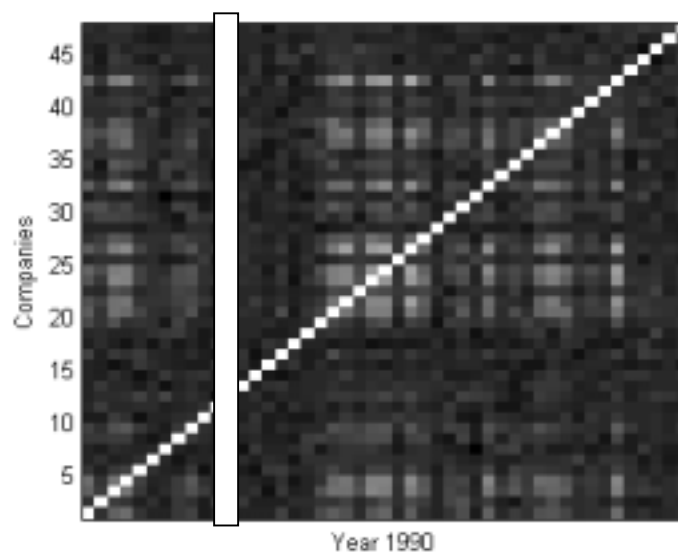


Figure 5.6.1. Illustration of column mean values.

The figure below shows how the correlation matrix changes over time for the studied data.

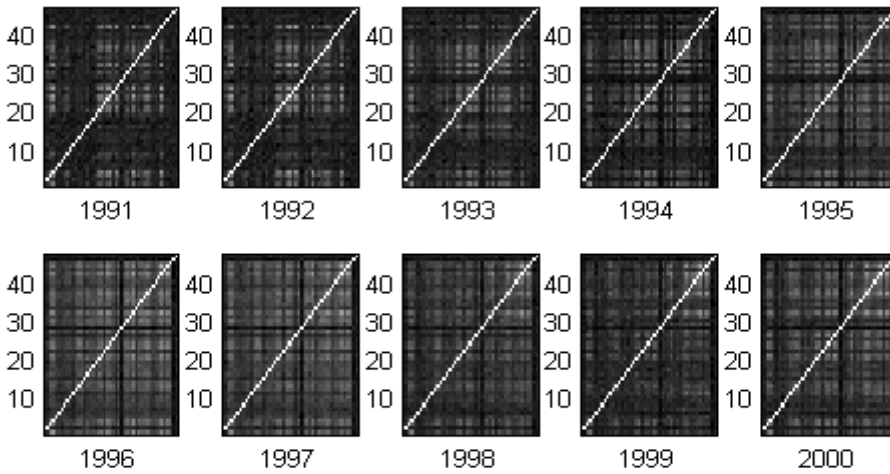


Figure 5.6.2. Changing correlation structure during the period 1991 to 2000.

We choose to study only the first 10 companies (to keep the number of figures within a reasonable amount). As can be seen in the picture below there are strong similarities between some companies at certain times. There seem to be “steps” at certain times, which is a common pattern seen for all companies. The three areas, a, b and c are visible in some of the companies. Some similarities are not surprising, since we should see the same effect here as we do in the mean of the whole correlation matrix; at certain periods of time the stocks move in a similar way.

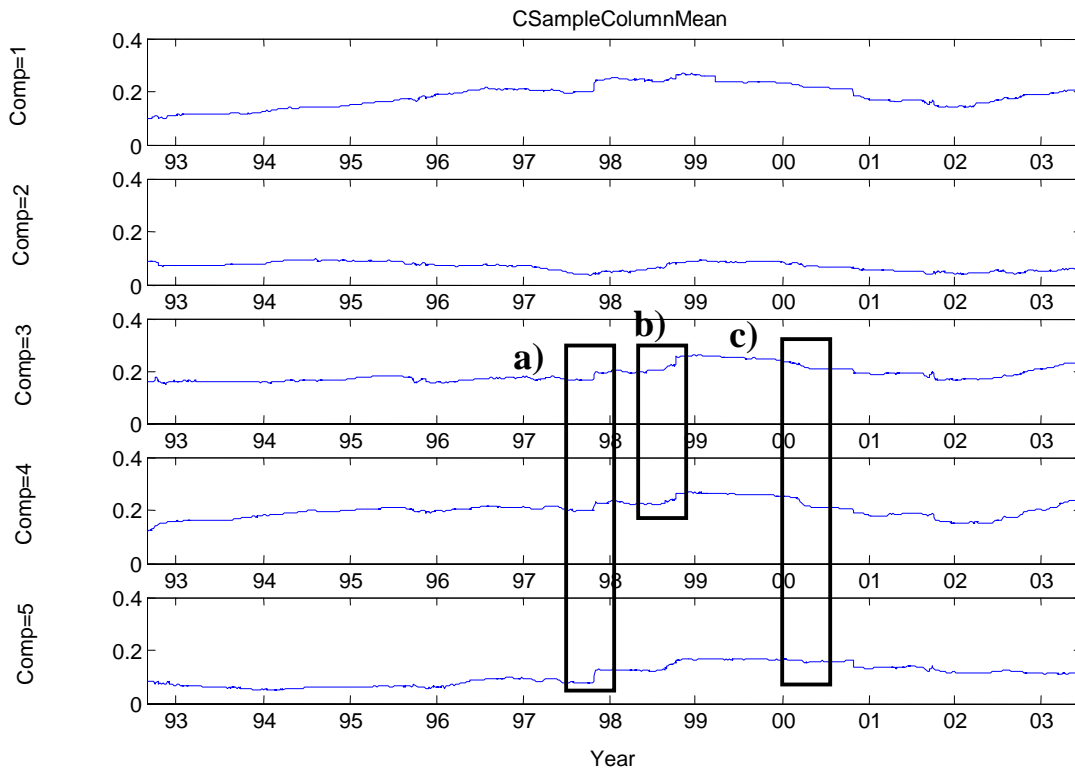


Figure 5.6.3. Mean values of the columns.

The following figure shows similar patterns for the next 5 companies as the ones seen in the previous figure. The observant reader can surely find more patterns in the presented pictures, a, b and c should mainly be viewed as examples or points of interested.

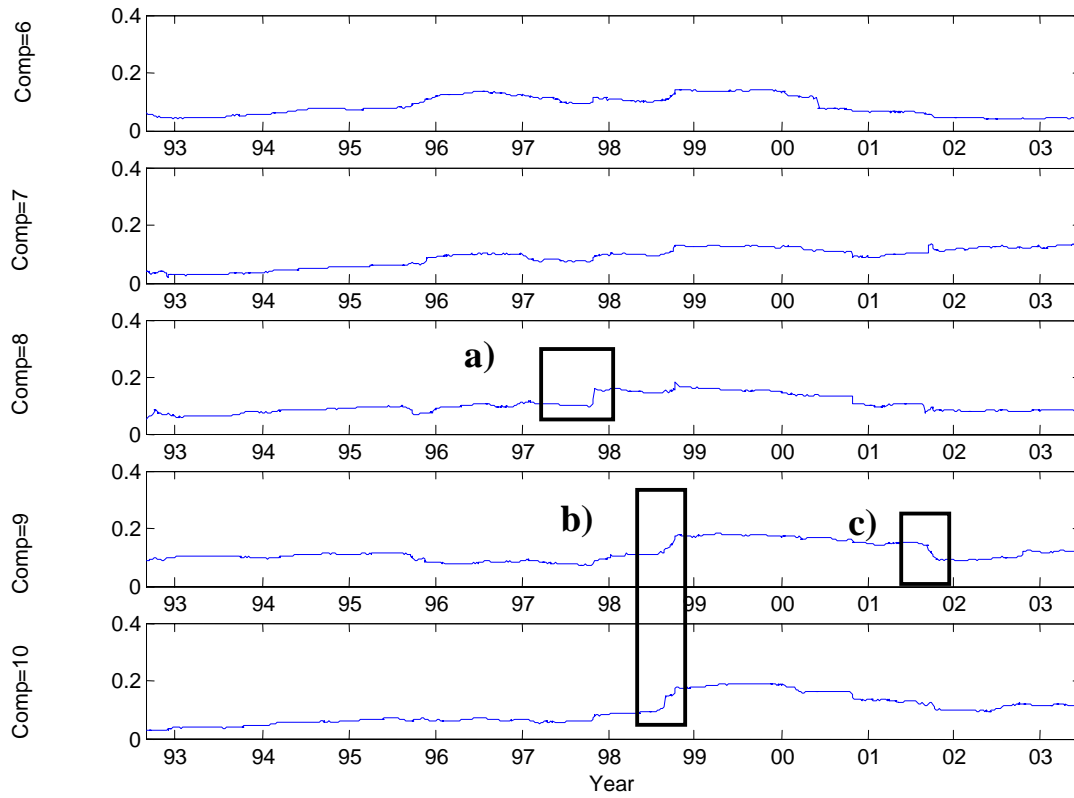


Figure 5.6.4. Mean values of the columns.

To compare the return more directly to the mean value of the correlation matrix the first 4 companies are shown in the two following figures. It is very hard to find any similar patterns, even if one looks at several more companies. The column-mean-values have general trends as we have seen above, but a connection between the return and the column mean values is hard to find.

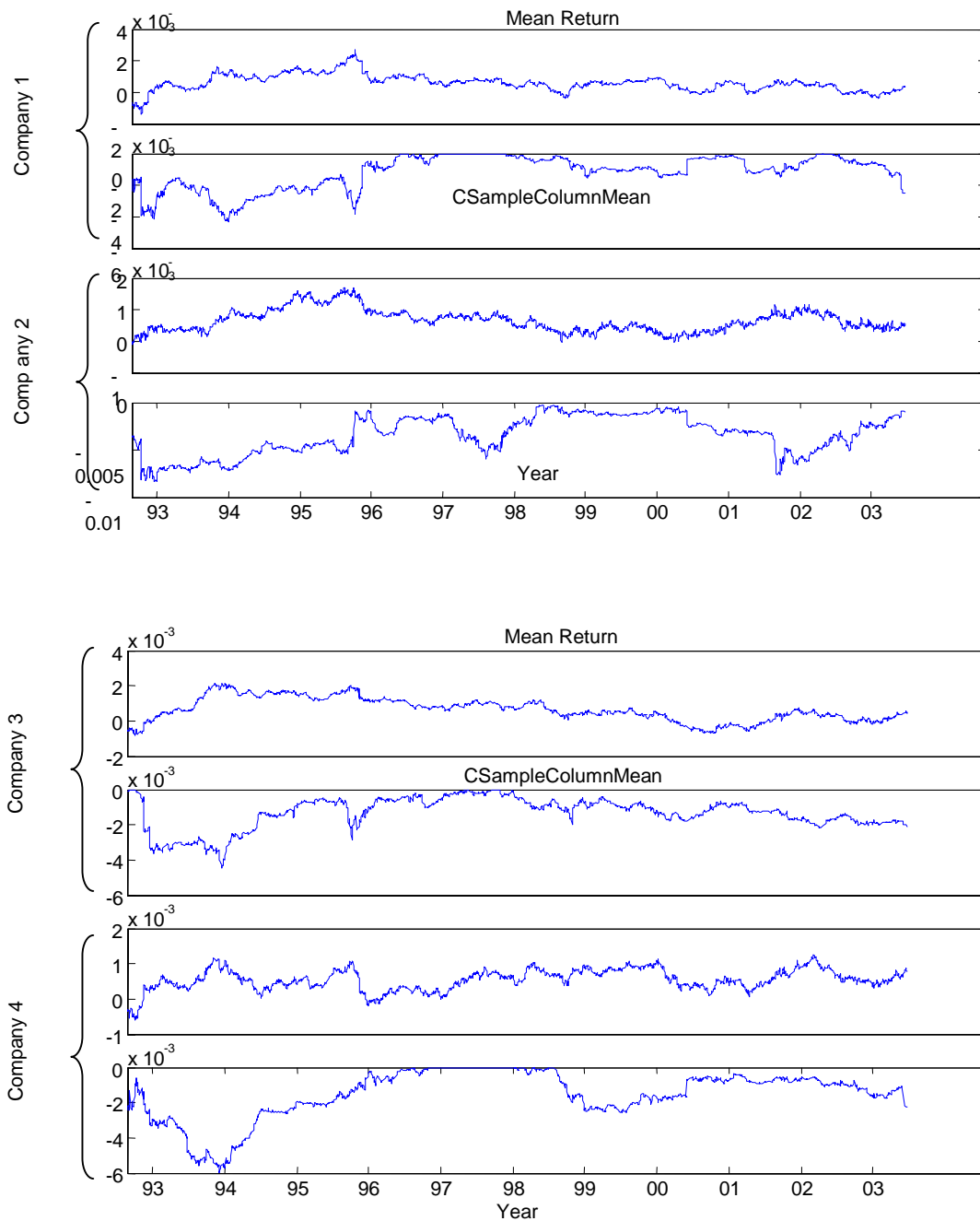


Figure 5.6.5. Column mean values and return.

5.5.3 Correlation Between Column Mean Values and the Return

A last attempt to find a pattern in the correlation between the stocks' returns and the column mean values were calculated. For Volvo this picture is shown below. The pattern looks quite random and moves from the extreme values -1 to 1 very rapidly. One might question what this correlation actually means. The answer would be that we are studying if there is any connection between "if the company follows the market" and its return.

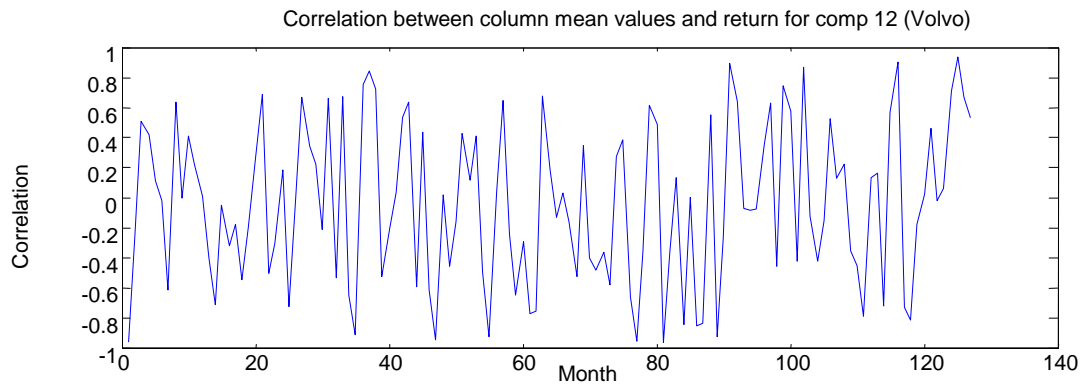


Figure 5.6.6. Correlation between column mean values and return for Volvo.

The figure below again shows the correlation between the returns and the column mean values, but this time for the first 6 companies. Possible patterns are close to impossible to find indicating that there is no clear relationship between the column-mean-values and the return.

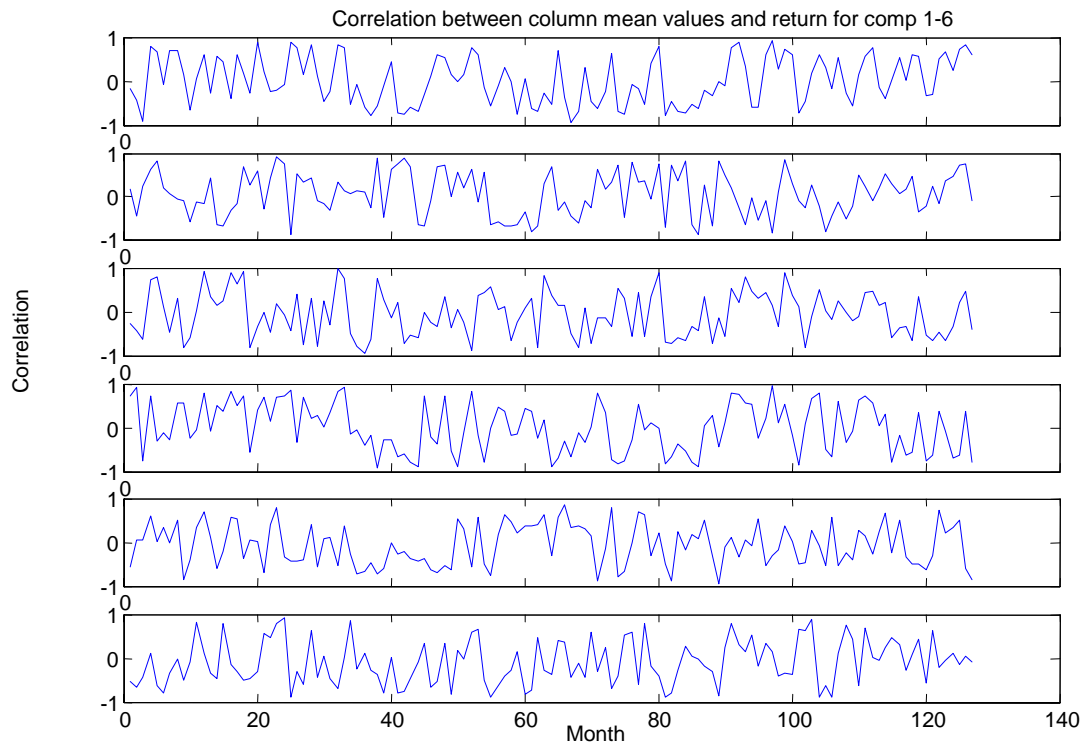


Figure 5.6.7. Correlation between column mean values and return for company number 1 to 6.

A final analysis of the column-mean-values was to compare the columns of Sample and Power Mapping. The idea suggested by Thomas Guhr¹¹² was that the column-mean-values might indeed be similar for Sample and Power Mapping. It could be that the noise introduced might even out. The test revealed that the difference is quite large; the mean value is around 30 % smaller for Power Mapping. If one considers a standard correlation matrix this is in reality not a strange result. Tests have shown that very few of the values in the correlation matrix are actually smaller than zero, roughly 10 % for our data. Since Power Mapping “pushes” all numbers closer to zero and we have mostly positive values this means that the mean value of all of the mostly positive values must decrease. If we had a more idealized situation with 50% positive and 50% negative values we would most likely have more similar mean values of the columns.

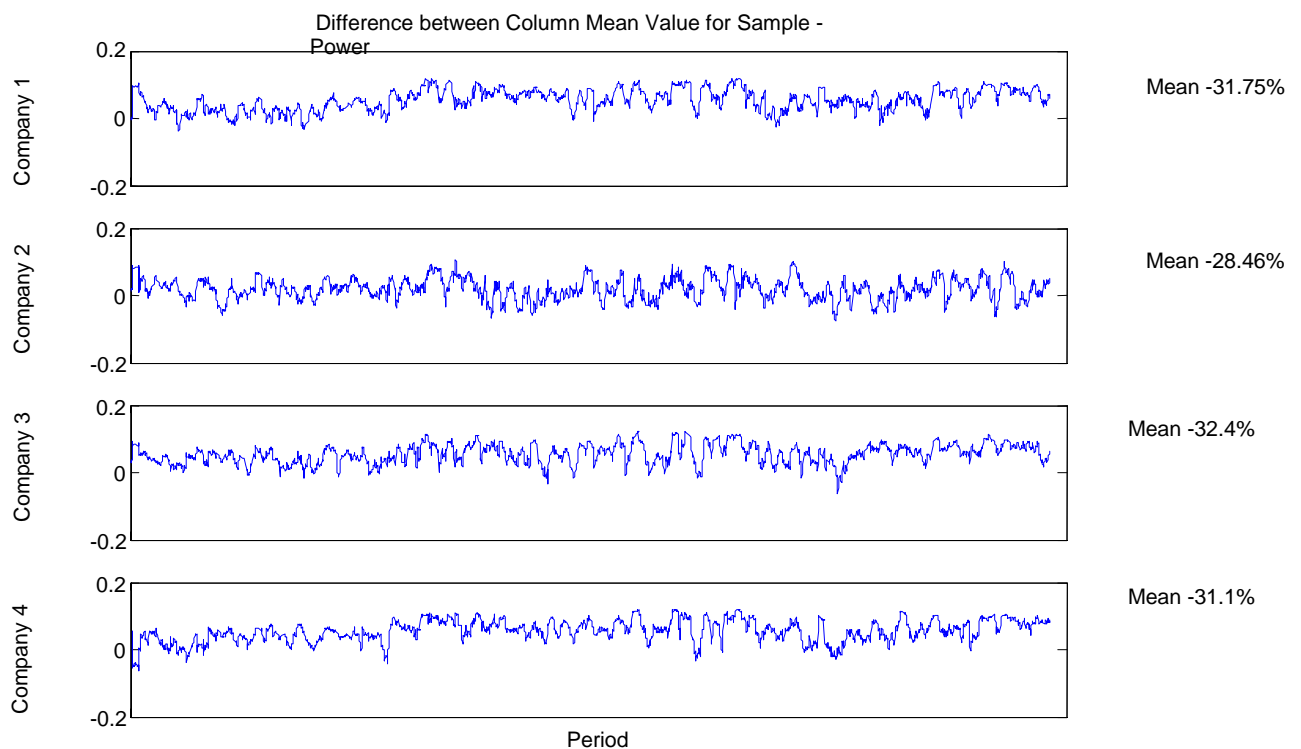


Figure 5.6.8. Difference in percent between Column mean value for Sample minus Power mapping

5.6.4 Conclusions

Several aspects of the matrices’ column mean values have been studied. The main result one can conclude is that some common patterns in the mean columns can be found, but no connection to the return is visible. With more studies performed more useful patterns might be found.

¹¹² Thomas Guhr, personal interview conducted during 2003-09

Chapter 6

Summary and Conclusions

The main focus of this thesis has been to investigate if Power Mapping and RMT-filtering are able to improve the correlation matrices so that the risk can be reduced in portfolio management. A wide array of tests has been performed and the two methods produced promising results in most of the tests.

The 197 Swedish Companies

The Power Mapping and RMT-filtering manages to reduce the risk for Portfolio(1) with about **-45%** compared to Sample when negative weights are allowed. Obviously this is a strong improvement when looking only at the risk. If we look at the return to risk ratio for the first portfolio for Power Mapping it is **-43%** (with risk a negative value is naturally good in comparison, but with return a positive value is beneficial) compared to Sample and **+55.1%** for RMT-filtering compared to Sample. This means that RMT-filtering produces the highest return for the least risk in this case. In reality however negative weights are not always possible. RMT-filtering produces splendid results; something quite surprising since there is no clear branch structure and one would expect it to work unsatisfactory in this case. Power Mapping does not produce a good return, but if we increase our expected return we get much better values. For Portfolio(3) Power Mapping had increased its value to **-17.4%** which is much better than **-43%**. This is interesting; it is important to note that all portfolios are actually very close to the MVP, which is something that we had to endure with our method, but is actually not something we want. If we could select higher expected returns (with another dynamic selection method) it seems like Power Mapping would produce better results.

With only positive weights Power Mapping has **-1.8%** less risk compared to Sample and RMT-Filter has **-2.6%** for the first Portfolio. The return to risk ratio was **3.85%** and **3.1%**. Obviously the improvements aren't as strong with negative weights.

An interesting observation is that the actual risk is much lower for Sample with constraints imposed, but actually slightly higher for Power Mapping and RMT-filtering. This indicates that the estimates for the correlation matrices for RMT-filtering and Power Mapping are good from the beginning and that they do not benefit significantly from a reduction in estimation error. Instead the imposing of constraints introduces specification error that slightly increases the risks of Power Mapping and RMT-filtering (see previous discussion in Chapter 5.1.5). In simple terms the results above indicate that the RMT-filtering and Power mapping techniques are useful for an investor who wants to reduce noise and increase the likelihood of creating accurate portfolios, when dealing with portfolio optimization according to Markowitz. Using both techniques reduces risk, and in some cases even boosts the return, which of course is favourable for the investor.

For the portfolio with no constraints it was found that the estimations for risk were incorrect for Sample while RMT-filtering and Power Mapping seem to have more

realistic estimates for risk. A sound conclusion from this result is that Sample is, due to noise, not showing a correct picture of reality, leading the Markowitz optimization process to “believe” that it’s possible to create portfolios with less risk than is actually true. This means that the investor have to be very careful when using raw stock data in the process of finding the optimal portfolios of assets without applying e.g. noise reduction and shrinkage techniques, this since the noise and other disturbances in the raw data might create faulty portfolios. One important thing to remember is that the estimations of the expected return introduce far bigger errors if they are not correct compared to the errors in the correlation matrix¹¹³. However it is far easier to improve the correlation matrices (which isn’t to say that it is easy), this is one reason why so many methods exist to improve the correlation matrices and the reason that this thesis focus on the correlation matrices.

U.S. industry Portfolio

Several variables were studied in the U.S. Industry portfolio. For the MVP Power Mapping had -42.6% less risk and RMT-filtering had -41.1%. RMT-filtering also produced a very good return to risk ratio on all the portfolios, a 186.8% improvement for RMT-filtering compared to Sample for Portfolio(1). Power Mapping produced much better results compared to the calculations on the Swedish daily data and got 79.9% (return to risk ratio). In this portfolio it is not a surprise that the RMT-filtering works good since a clear branch structure exists. Another finding is that the estimated risk behaves in exactly the same way as for the Swedish portfolio. Sample is always estimated to have very little risk when the actual risk is far higher than for Power Mapping and RMT-filtering. This is an important finding, since it further indicates that Power Mapping and RMT-filtering are showing estimated risks that are much closer to the “true” estimated risk. Yet another interesting pattern is that the estimated risk for both Power Mapping and RMT-filtering are very similar which is reasonable if the previous discussion is correct. For the investor this indicates that noise reduction techniques, such as Power mapping and PM-filtering on data with strong branch structure or industry structure, can have beneficiary effects concerning the portfolios created. There will also be some positive effects on the accuracy of the created portfolios.

With constraints, RMT-filtering does not manage to create a portfolio with less risk than Sample, Power Mapping is slightly better than Sample in this example. The return to risk ratio is still slightly higher for both Power Mapping and Sample.

For the U.S. portfolio we can see that the actual risk is much lower for Sample, Power Mapping and RMT-filtering compared to the risk for the portfolios without constrains. This is different compared to the no-constraints structure for the Swedish portfolio. The explanation for this is most likely that since we are looking at monthly data we have a larger error in the estimated correlation matrix (compare with the discussion in this Chapter about the Swedish companies). One interpretation of this can be that the investor should use high-frequency data since

¹¹³ Chopra, V. K., and William T. Ziemba, “The Effect of Errors in Means, Variances and Covariances on Optimal Portfolio Choice,” *Journal of Portfolio Management*, July1993.

this can reduce estimation error, but the investor must also examine the data carefully in order to detect important trends in the data before using optimization techniques.

Portfolios with Different Standard Deviations

There seem to be a pattern in how good the Power Mapping and RMT-filtering works on portfolios with varying risk. Both methods worked even better on the portfolios that had the highest risk. One possible explanation is that there is more noise to be suppressed in the portfolios with the highest risk; another explanation could be that when the risk is higher it's even more important that the information in the correlation matrices is correct. This is an important result for investors, since the noise reducing techniques can have a larger effect on high-risk investments. A large reduction of noise also make the noise that the noise reducing methods possibly might introduce themselves small in comparison to the noise already present in the raw data. Our conclusion in this area is that Power Mapping and RMT-filtering works even better compared to Sample on portfolios with high risk (when no constraints are imposed) in our study, but more tests should be performed on more portfolios before this statement could be generalized.

Image Filter Tests

Some simple tests with image filter algorithms were also conducted. The results were negative, but many more advanced image filter methods are waiting to be tested, they should hopefully work better.

PM-filtering Results

The PM-filtering works fairly good, but not as good as Power Mapping and RMT-filtering for the MVP for the Swedish portfolio. For the portfolio with the 197 Swedish companies the best value for Power Mapping was -44.1% (compared to Sample) RMT-filtering -46.3% and **-43.3%** for PM-filtering. The return to risk ratio was quite extreme with -67.2% for a q-value of 1.5 and -93.3% for PM-filtering.

The tests performed on the 48 U.S. industries portfolios gave much more promising results: Power Mapping -48.3% RMT-filtering -46.9%, PM-filtering q=1.5 **-51.3%** q=1.7 **-52.1%**. The return to risk ratio was excellent with 476% with a q-value of 1.5 and 502% with a q-value of 1.7. This was a trend that was noted in our previous thesis¹¹⁴ For the investor this means that applying the PM-filtering technique on stock data can lower the risk substantially for portfolios but even more important the return may be increased substantially.

Interestingly we got the best result for a q-value of 1.7, just like in our previous thesis, perhaps this could be because the data were fairly similar in structure since it

¹¹⁴ Andersson, P. Öberg, A. 2003. Noise Reducing Methods for Correlation Matrices New Techniques to Minimize Risk in Portfolio Management?

was monthly data for a fairly small portfolio. Our conclusion is that PM-filtering works well on monthly data, but do not work too well on higher frequency data. PM-filtering also works very good when using data with a clear industry and branch structure. Further studies should be done on daily date before a more certain conclusion can be made on higher frequency data.

The Mean Values of the Columns of the Matrices

The mean values of the columns of the correlation matrices have been studied, revealing some smaller patterns between companies. If one studies the mean values of the complete correlation matrices, high values occur at stock market crashes, these have to be visible in the mean values of the columns also (otherwise the total mean value cannot be high). Our conclusion on the mean values of the columns is that there are patterns, one can easily see if some companies do not follow the general trend of for example the current branch. Most of the companies (mean values of the columns) will also follow the very high mean values in a stock market crash. There seem to be other patterns as well, but the reasons for these are unclear. The attempt to connect the mean values of the columns to the companies return did not prove successful; the conclusion is that in our data no visible connection was to be found.

Smaller weights for Power Mapping and RMT-filtering

It has been shown in Chapter 5.1 and Chapter 5.2 that the weights fluctuate less strongly when Power Mapping and RMT-filtering have been applied. This indicates that the noise is actually less in the correlation matrices leading to more “stable” and less extreme weights. This shows the usefulness of using the techniques as noise reducers for the concerned investor or portfolio manager.

Problem Statement Summary

It seems clear that Power Mapping and RMT-filtering produce portfolios with less risk than the initial Sample matrix for almost all cases. The somewhat rough return to risk ratio also seem to indicate that especially RMT-filtering can achieve higher return with lower risk than Sample. The methods also seemed to work differently on different segments of the market. The study was fairly limited in this area but indicated that the noise reducing methods worked even better for portfolios with higher risk. Adding constraints like forbidding negative weights limited the benefits for Power Mapping and RMT-filtering. No new strong patterns were found in the mean values of the correlation matrices, and they were not possible to connect to the return.

Many properties of the correlation matrices have been studied and many new questions have arisen. Further studies will surely be needed before a clearer picture of the so elusive true correlations will emerge.

Chapter 7

Suggestions for Future Studies

In this thesis we have studied different noise reducing methods and the general structure of the correlation matrices. Many areas still need even more expansions, especially the image filtering methods and the mean values of the column matrices should be addressed in more detail. In this chapter we give some suggestions for interesting areas for future studies.

- Although the Swedish portfolio is fairly large, at least compared to the portfolio used in our previous tests, larger portfolios would be even more interesting. Portfolios from larger markets in Europe or in the U.S for example.
- Higher-frequency data. Most of the studies have been performed on daily data, but extremely high frequencies, minutes or even seconds would surely reveal some usable patterns.
- Studies of portfolios on selected markets with a more limited number of branches, perhaps the branches for IT-companies.
- Studies with a focus on the time around large fluctuations on the market. How will these, very fast and complex changes, affect the noise reduction of the portfolios?
- The effects of different lengths of the in-sample period have been studied and the differences in the results are sometimes large. More tests with different in-sample periods would be a good idea to study; perhaps there are patterns to be found. Different markets might need different length of the period, as well as different market conditions. Also the different noise reducing methods could work differently with different length on the in-sample period; it is not impossible that some methods work better with more/less information.
- More models like Noh's model could be used to further study the noise reducing aspect of Power Mapping and RMT-filtering.
- Since this is the second study where PM-filtering produces good results it gives further weight to its actual usability. Naturally more studies are important. Especially the understanding for why the PM-filtering works with monthly data needs to be developed. Some rough seasonings exist but these are not mathematically securely founded.
- An image filter has been tried in limited scale in this thesis with negative results. More advanced image filtering methods could bring more positive results. For example image noise reducing methods and more accurate contrast filters.

- The computations behind the methods analyzed in this thesis can be quite burdensome for the computer when the data is high-frequent. What effect would the delays for the calculations have on how the companies work with these tools? Some of the calculations initially planned would have taken 40 hours for every procedure. Now it was reduced to 8 hours/procedure. The time to finish the minimization procedure grows exponentially, indicating that with very large portfolios problems might arise.
- The mean values of the columns of the matrices provided some interesting patterns when compared to each others. Not surprisingly they behaved similarly for companies in the same branches. There are several more things that could be studied for the mean values of the correlations. This would be even more interesting if the data were more high-frequent since the correlations would have more information.
- To study the mean values of the correlations for data with high frequency and long time periods, especially at times of financial unrest, e.g. at stock market crashes could bring more insight into these times of turmoil.
- Right now the in-sample period collects data evenly from a period, say 5 years, with this methods we catch the long trends in the market. It could be possible to divide this area into several different groups and construct correlation matrices for these shorter periods, these correlation matrices would then include the shorter financial trends. Perhaps it could be possible to weight these correlations together into a correlation matrix that more selectively chooses what trends to include.
- To use more statistical methods like hypothesis testing would further increase the reliability of the tests.
- It is surprising that the estimated risk always is better for Sample than it is for RMT-filtering and Power Mapping. To further investigate the reason for this could produce **very** important results for how Power Mapping and RMT-filtering affect the optimization procedure. It could also help explain more in detail the problems with the Sample matrix.
- The improvements for the noise reducing methods aren't as strong when no short selling is allowed. It would be meaningful to allow short selling for some companies so that a more realistic situation would be created. Also to limit short selling at market crashes would eliminate some unrealistic scenarios. With the possibility of "selected" short selling (as well as looking at e.g. the Sharpe quote), a more realistic study of a real investor situation would emerge.
- Learning algorithms. It should be possible to create correlation matrix algorithms that improve the structure depending on how the market looks, as well as actually learn from the history. Perhaps the methods studied in this thesis could be dynamically combined into an even more powerful tool.

References

Literature

- Alexander, C., *Market models*, 2001, John Wiley & Sons
- Brealy R. & Myers S., 1996, *Principles of corporate finance*, 5th edition.
- Böiers, L-C & Persson, A., 2000, *Analys i en variabel*, Studentlitteratur
- Brooks, C., 2002, *Introductory Econometrics for Finance*, Cambridge University Press
- Elton, E., Gruber, M., 1995, *Modern Portfolio Theory and Investment analysis*, John Wiley & Sons Inc.
- Fabozzi, F. J., 1999, "*Investment management*", 2nd ed., Prentice-Hall Inc.
- Guhr, Thomas, 2001, *Econophysics*, Lund University Department of mathematical physics
- Guhr, Thomas, 2002, Lecture notes in *Econophysics*, Lund University Department of mathematical physics
- Haugen R., 2001, *Modern Investment Theory*, 5th edition, Prentice Hall
- Johnson, R.A and Wichern D.W, 2003,. *Applied Multivariate Statistical Analysis* 5th ed, Prentice-Hall.
- Kee, Hong Bae, 2000, *Portfolio Diversification* Korea University Seoul.
- Körner S, Wahlgren L, Wahlgren, 2000, *Statistisk dataanalys*, third edition
- Muirhead R, 1982, *Aspects of Multivariate Statistical Theory*, New York, Wiley, cop.
- Solnik, Bruno H., 1987, *International Investments*
- Spanne, Sven. 1995, *Lineära System*,. Lund
- Åberg, Svante, 2002, *Quantum Chaos*, Lund University Department of mathematical physics

Articles

- Andersson, P. Öberg, A. 2003. Noise Reducing Methods for Correlation Matrices New Techniques to Minimize Risk in Portfolio Management?

- Bengtsson Christoffer. & Holst Jan, Nov 15th 2002, Two Shrinkage Based Covariance Matrix Estimators for Improved Efficiency of Mean-Variance Optimization
- Black, Fisher and Robert Litterman, 1992, Global Portfolio Optimization, Financial Analysts Journal, September-October, 28-43
- Bookstaber, Richard , Spring 1997. Global risk management: are we missing the point?, Journal of Portfolio Management Vol. 23, No. 3.
- Byström Hans NE, January 3, 2002. Cross-Sectional Correlation: New Evidence on Changing Correlations and Correlation Breakdown in Equity Markets
- Chopra, V. K., and William T. Ziemba, "The Effect of Errors in Means, Variances and Covariances on Optimal Portfolio Choice," *Journal of Portfolio Management*, July 1993.
- Green, Richard C. and Burton Hollifield, 1992, When Will Mean-Variance Efficient Portfolios be Well Diversified? *Journal of Finance* 47, 1785-1809
- Guhr, T & Kälber, B. 28 Jun, 2002, arXiv: cond-mat/0206577 v1
- Jagannathan Ravi. Tongshu Ma. Risk Reduction in Large Portfolios: A Role for Portfolio Weight Constraints. September 19, 2001
- Krengel, U 1991, Einführung in die Wahrscheinlichkeitstheorie und Statistik (Viewweg, Branunschweig) referred to by Thomas Guhr & Bernd Kälber, 28 Jun 2002, arXiv: cond-mat/0206577 v1
- Kullmann, L, Kerteszand, J and R.N. Mantegna, cond-mat 0002238 referred to in Thomas Guhr & Bernd Kälber, 28 Jun 2002, arXiv:cond-mat/0206577 v1
- Longin, F., Solnik, B., 1995. Is the correlation in international equity returns constant: 1960-1990?, *Journal of International Money and Finance* 14, pp 3-26.
- Laloux L. et al, 1998, Noise dressing of financial correlation matrices.
- Markowitz, Harry M. 1952. Portfolio selection: *Journal of Finance* 7
- Marsili, M, Data clustering and noise undressing of correlation matrices, cond-mat/0003241
- Noh., J.D Phys. Rev. E61, 5981, 2000, cond-mat/9912076
- Rey, David Michael, 2000, Time-varying Stock Market Correlations and Correlation Breakdown. *Finanzmarkt und Portfolio Management*, Nr. 4, S. 387-412
- Rosenow, B, Vasiliki Plerou, Parameswaran Gopikrishnan, H. Eugene Stanley. Jan 7th, 2002, Portfolio Optimization and the Random Magnet Problem

Ross, S, 1976, Jo. Econ. Theory 13, 341

Rubinstein, Mark, "Comments on the 1987 Stock Market Crash: Eleven Years Later, published in Risks in Accumulation Products, Society of Actuaries, 2000

Schwartz Tal, March 15, 2000, How to Beat the S&P 500 with Portfolio Optimization

Plerou Vasiliki, Parameswaran Gopikrishnan et al. July 31, 2001. A Random Matrix Approach to Cross-Correlations in Financial Data.

Plerou Vasiliki et al, 22 February 1999. Universal and Non universal Properties of Cross Correlations in Financial Time Series

Internet

Carlson School of management, University of Minnesota
<http://legacy.csom.umn.edu/wwwpages/faculty/AKapoor/4321/FALL02/24>
(2003-01-14)

Finance, College of Business & Information Systems Dakota State University.
<http://www2.dsu.edu/>
(2002-12-24)

Fusion Wave Design (data for this thesis is available here for download)
www.fusion-wave.com

Jeffrey Dodgson, Napier University, PCA-Principal Component analysis 2001
<http://www.maths.napier.ac.uk/~jeff/>
(2003-06-21)

G. David Garson, 2002, Lecture notes: Correlation, Quantitative Research in Public Administration.
<http://www2.chass.ncsu.edu/garson/pa765/correl.htm>
(2003-01-21)

Low Risk
<http://www.lowrisk.com/crash/crashcharts.htm>
(2003-08-10)

Högskolan i Skövde. Lektionsunderlag finansiell ekonomi
<http://www.his.se/ish/kurser/Kursplaner/Ht02/Nek/Overhead01.doc>
(2002-12-15)

Philip H. Dybvig, 1997. Mean variance Theory Lecture notes Washington University in Saint Louis
<http://dybfin.wustl.edu/teaching/inv/slides/invl3.html>
(2002-12-14)

Philip H. Dybvig, 1997. Investments, Washington University in Saint Louis
<http://dybfin.wustl.edu/teaching/inv/slides/invl2html>
(2003-03-13)

Seligson & Co, Risk och Portföljteori, 2002
www.seligson.fi
(2003-04-24)

Trading-Glossary
www.trading-glossary.com
(2002-12-24)

U.S. Research Data Library of Kenneth R. French, University of Dartmouth
http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html
(2003-08-25)

Personal Interviews

Thomas Guhr, Lund Sweden, conducted on several occasions during the period 2003-01-15 to 2003-09-06

Anders Johansson, Vinga System AB, Gothenburg Sweden , conducted on several occasions during the period 2003-05-20 to 2003-08-30

Databases

Affärsvärlden Financial Statistics
www.affarsvarlden.se

Stockholmsbörsen
www.stockholmsborsen.se

Affärsdata Financial Database
www.affarsdata.com

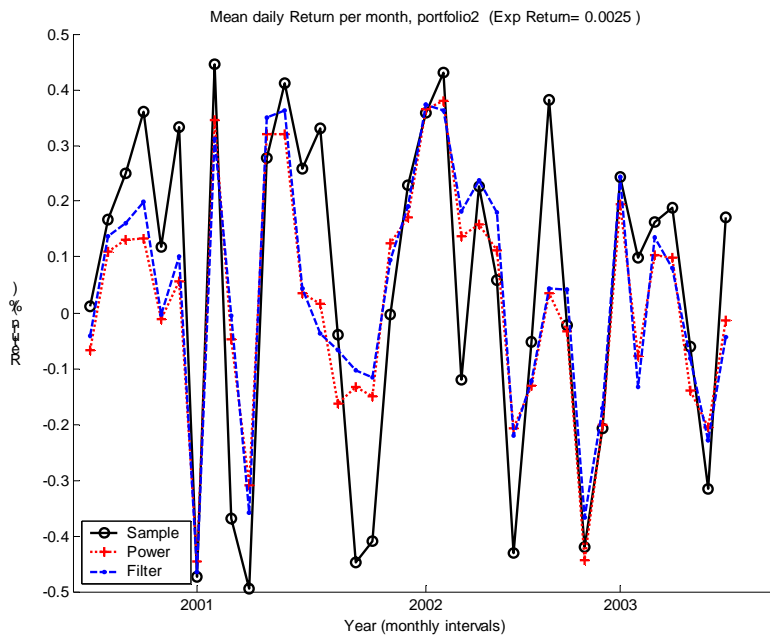
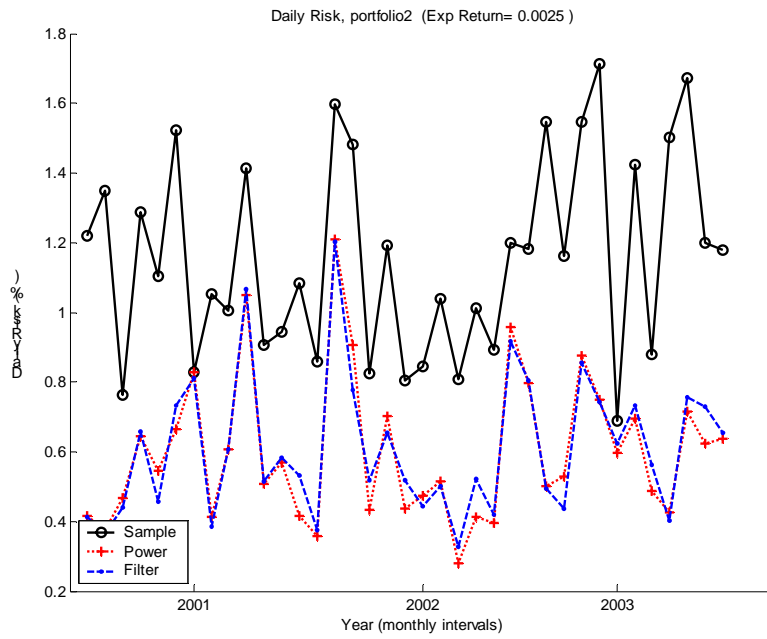
Frontpage

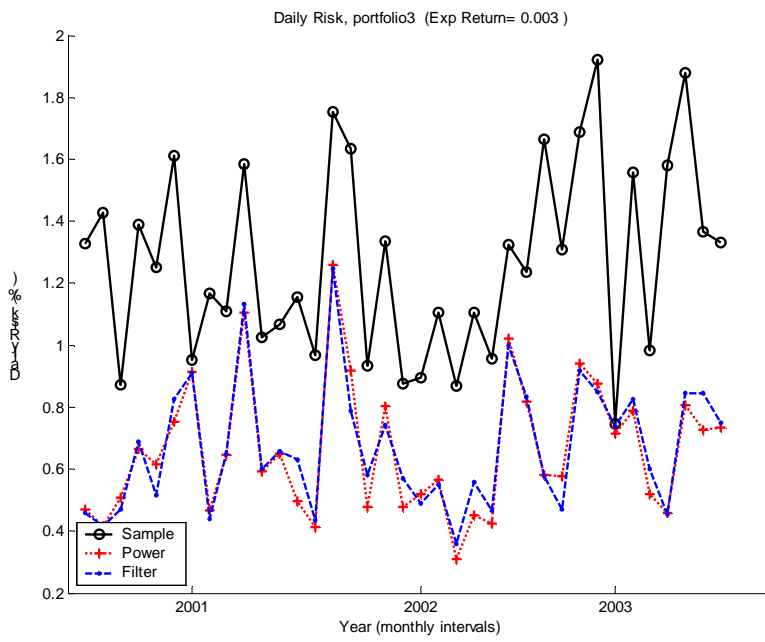
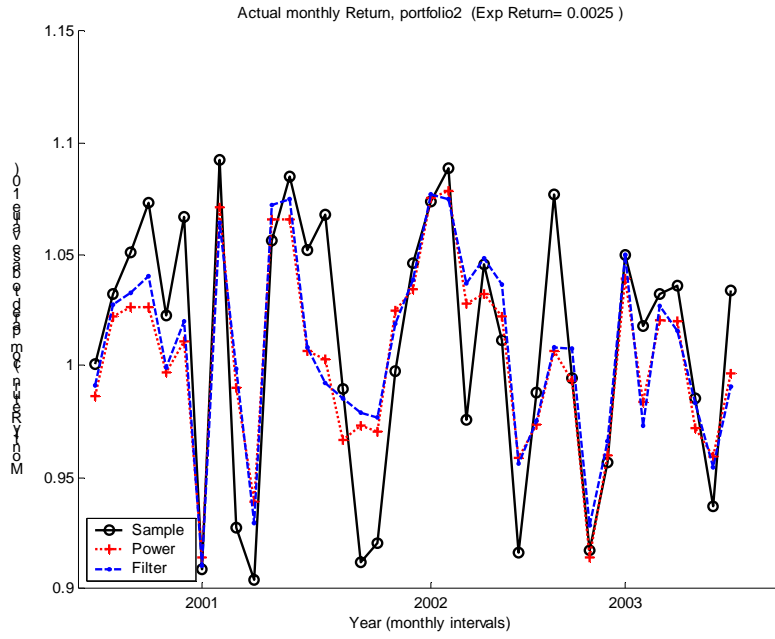
Original photo from Tonystone.com, additional design by Andreas Öberg

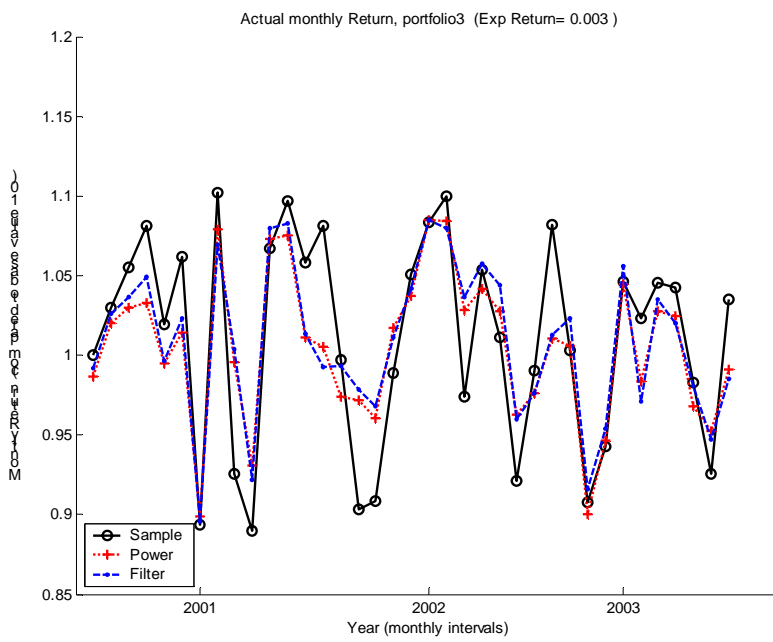
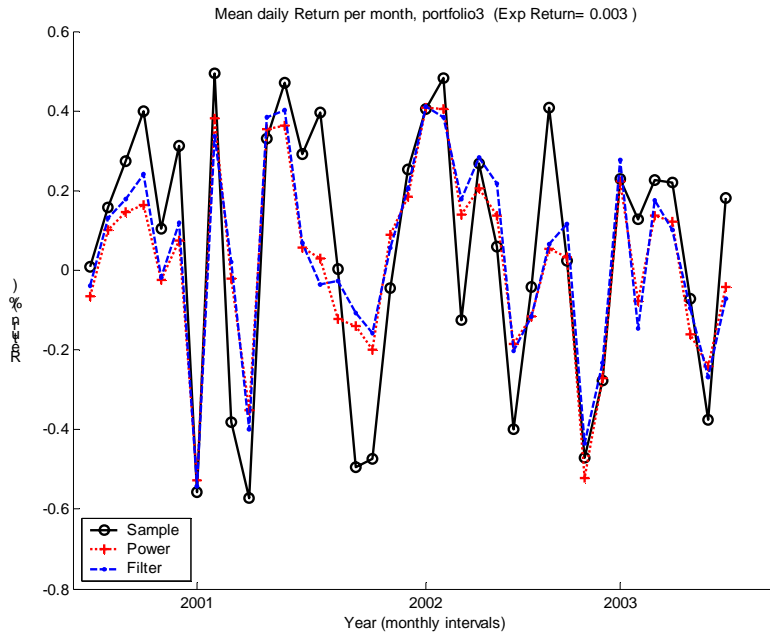
Appendix

A.1 Portfolios Created using Swedish Stock Data

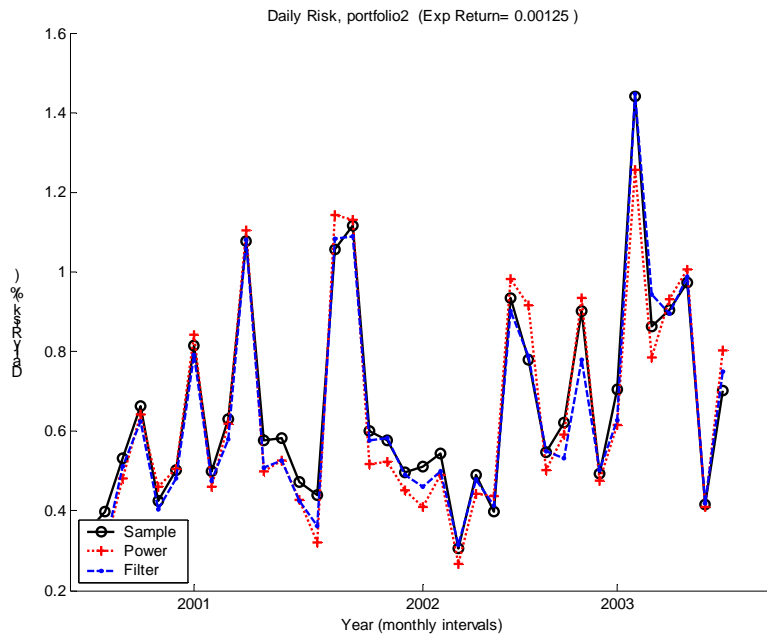
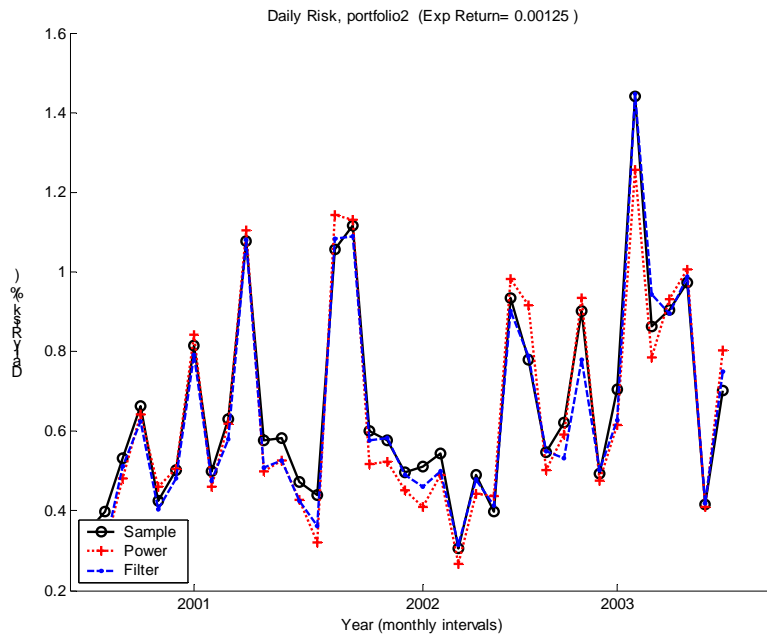
A.1.1 Portfolios Number 2 and 3 no Constraints Imposed

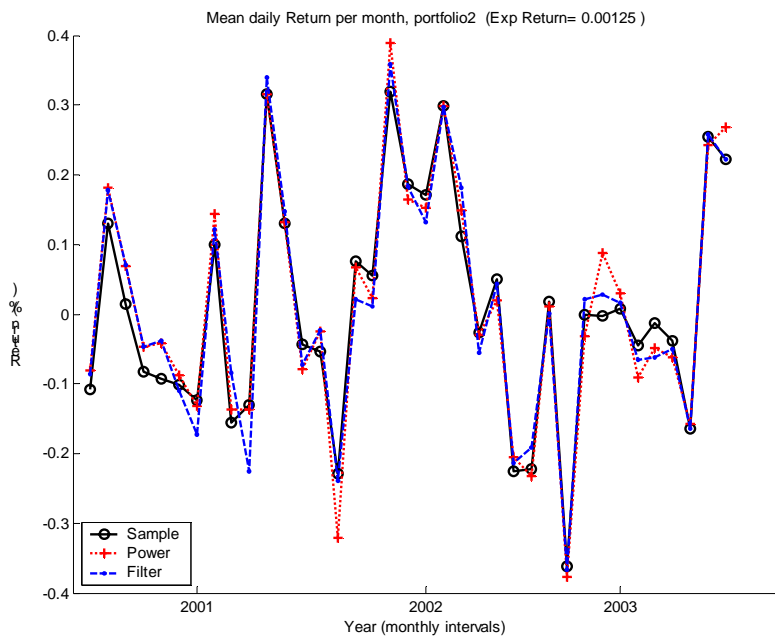
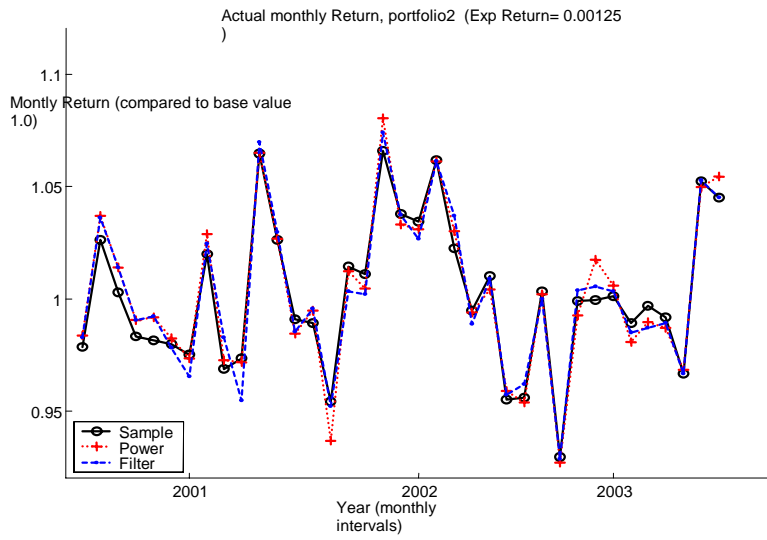


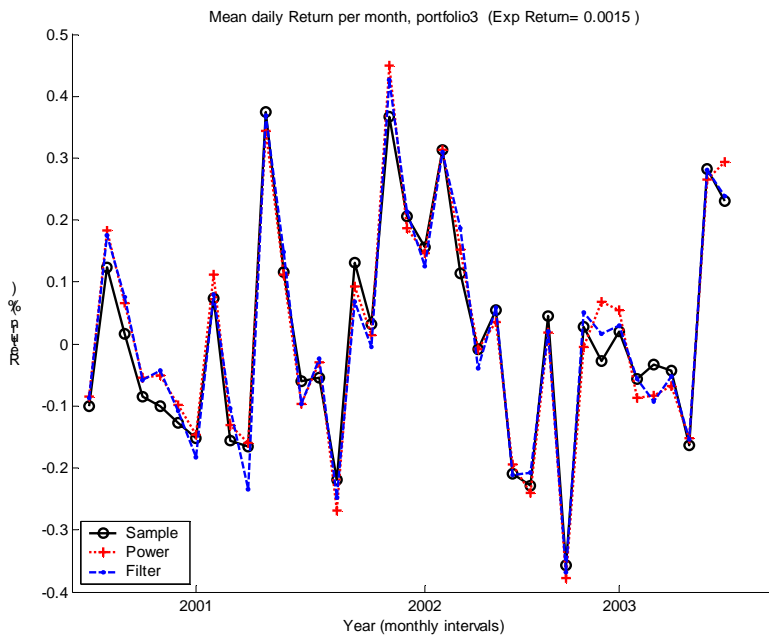
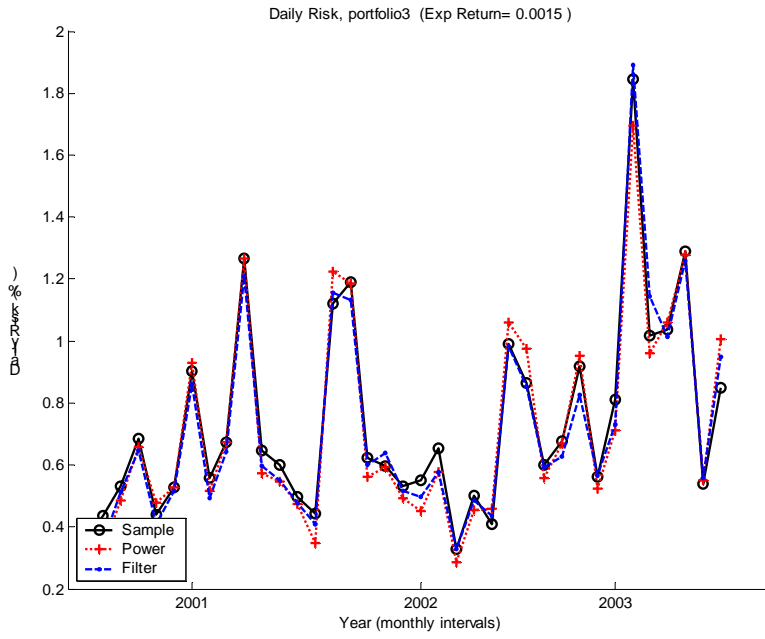


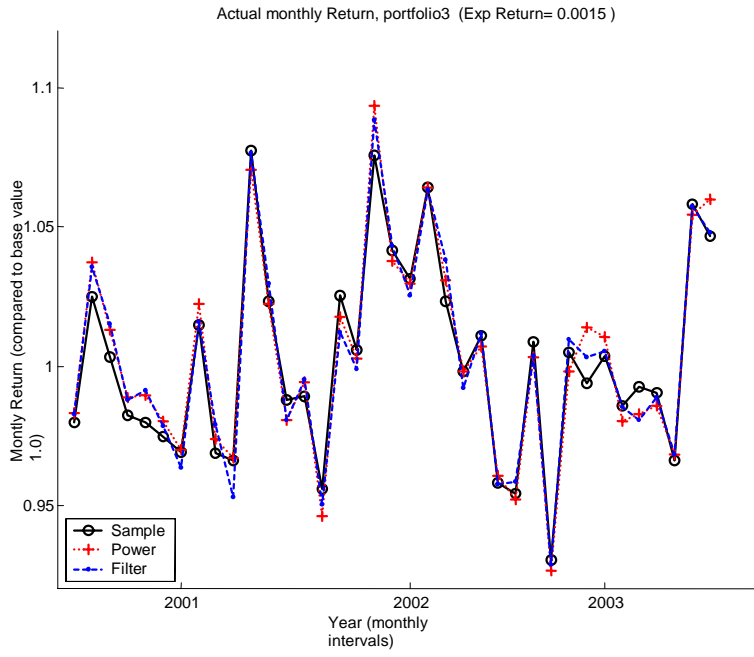


A.1.2 Portfolios 2 and 3 constraints imposed



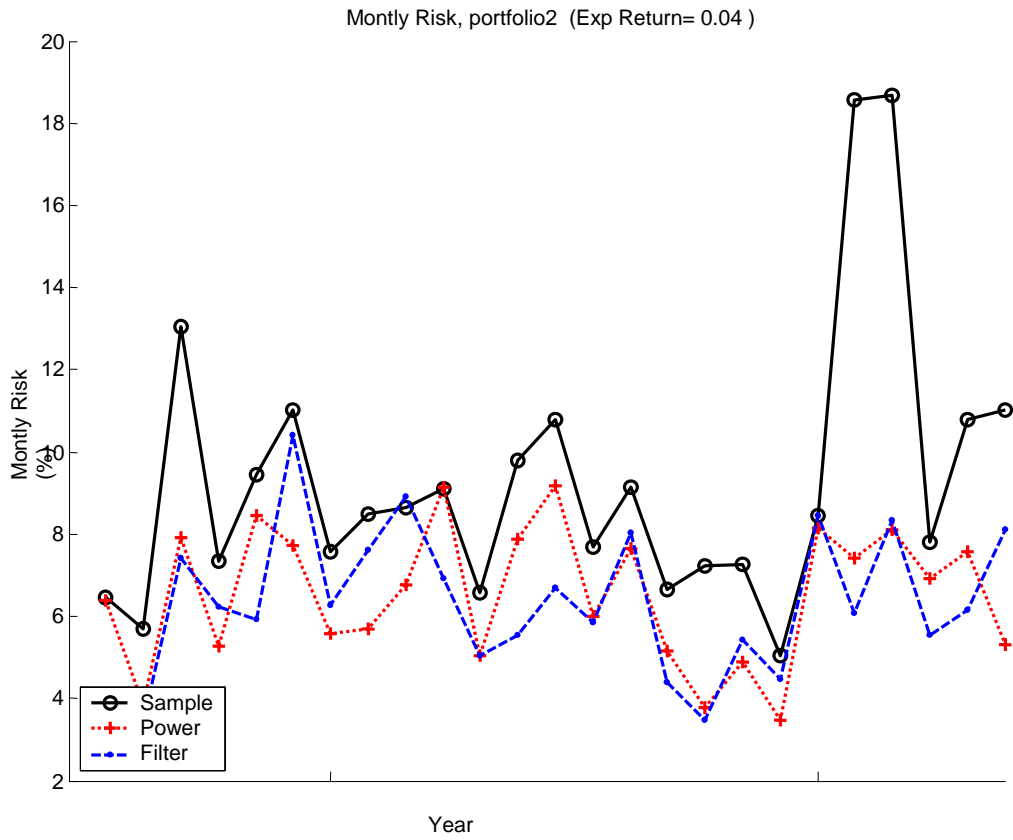


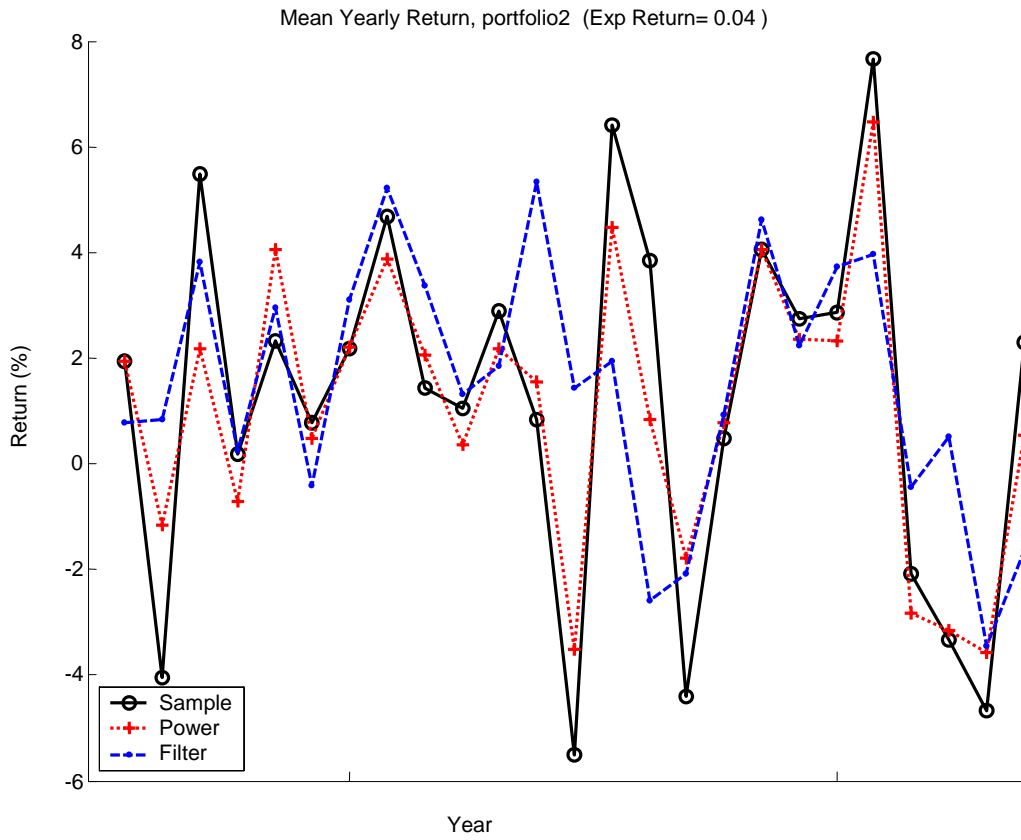


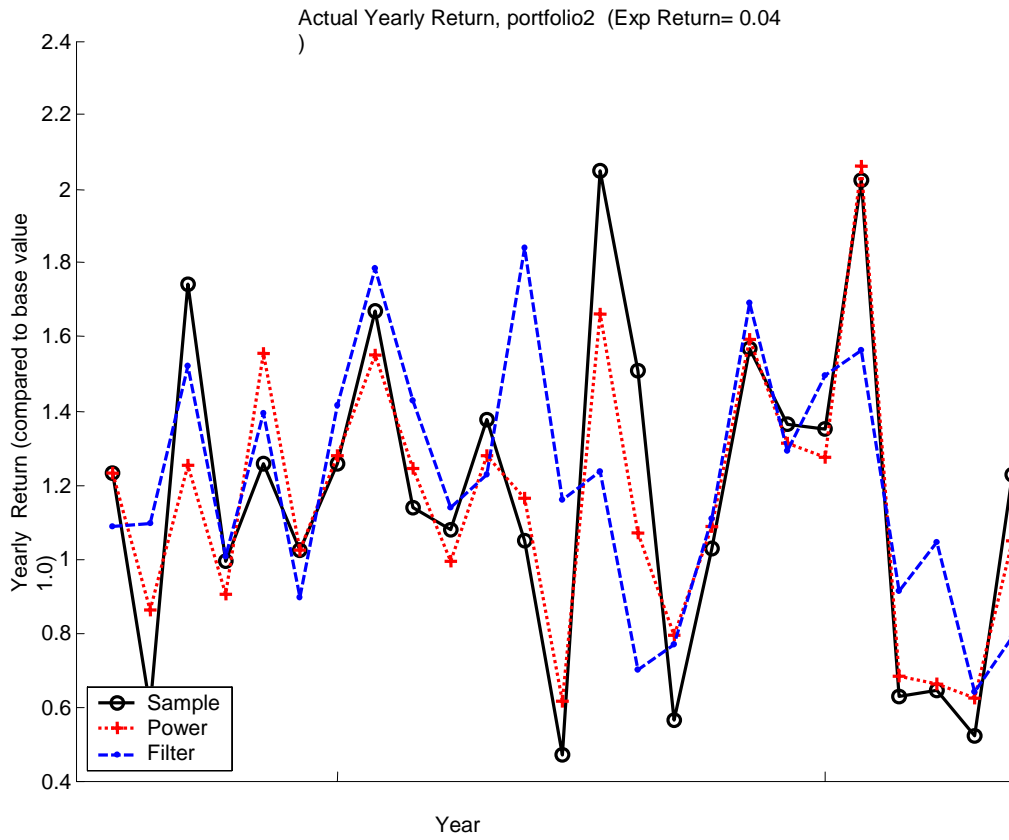


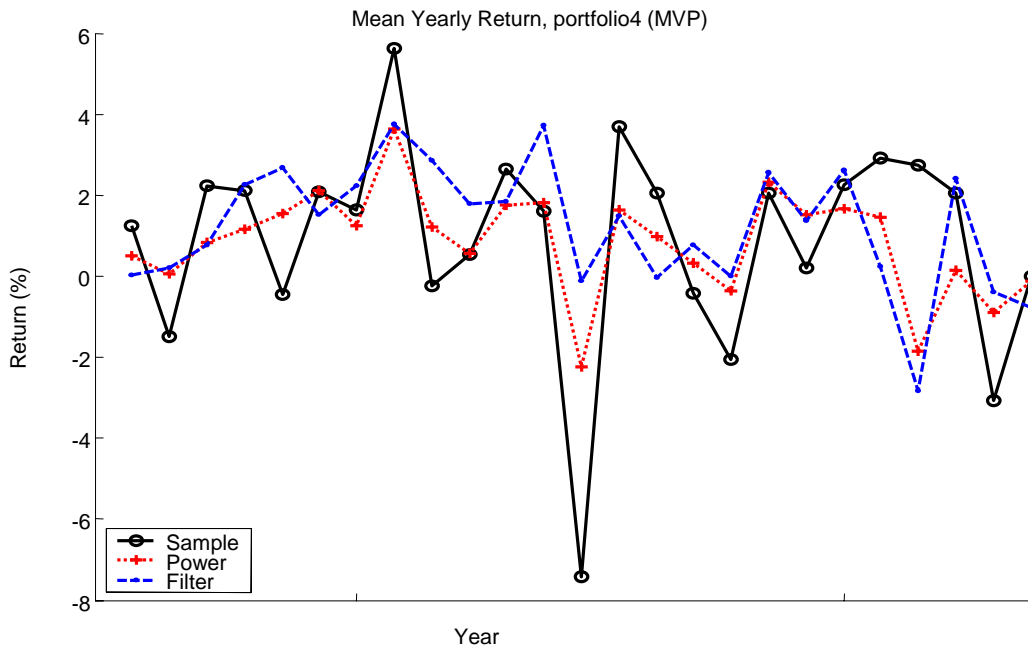
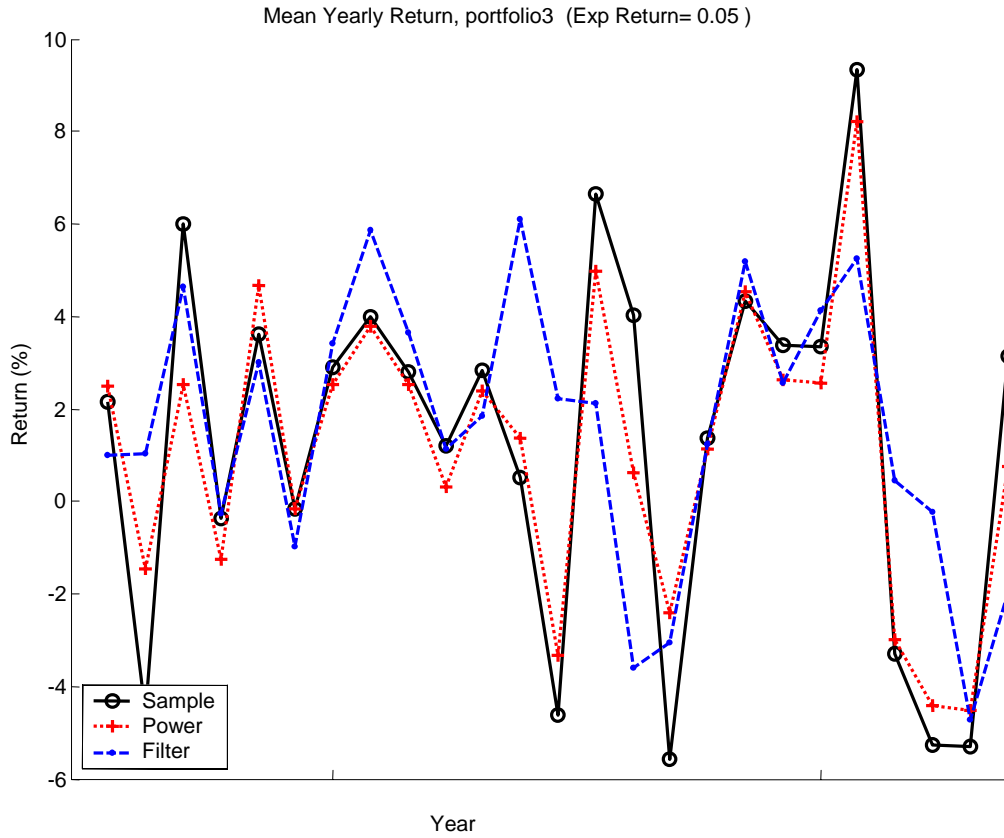
A.2 Portfolios Created using U.S. Industry Portfolios

A.2.1 Portfolios Number 2 and 3 no Constraints Imposed

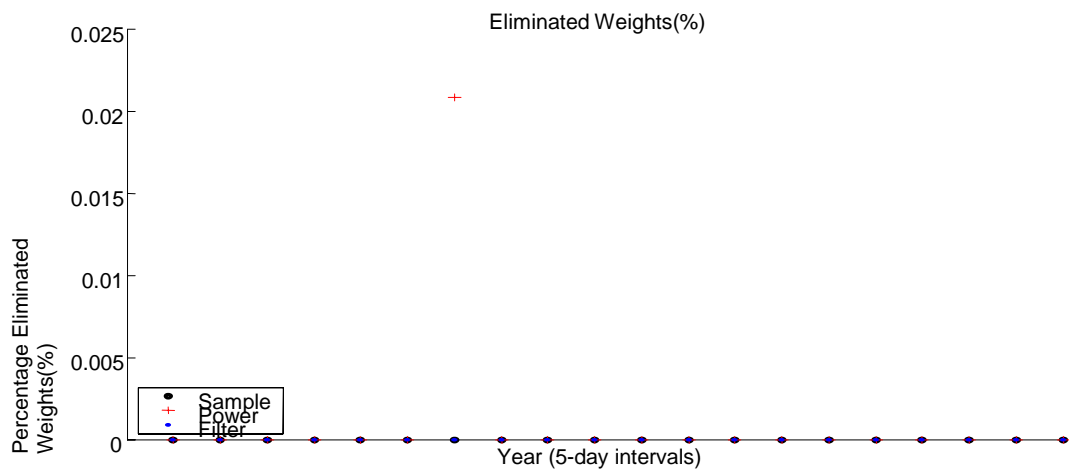








A.2.2 Weights excluded (no constraints)



A.3 Companies of the Portfolios

A.3.1 The 197 Swedish Companies

Information for the 197 Swedish Companies

In the following sub chapter the characteristics for the 197 Swedish companies are listed. The name, number and the existing branch should be the most interesting values, but the number of holes in the original data is also of some interest. Observe that the number of data points is the original maximum numbers of data points. In the thesis the number of data points is as described in Chapter 4 and Chapter 5.

Branch description for the 197 Swedish Companies

Table A.3.1 Branches according to Affärsvärlden¹¹⁵

Commodity	1
Industry companies	2
Consumer goods	3
Health	4
Finance	5
IT-companies	6
Telecommunication	7
Media & Entertainment	8
Services	9

The branch indices seen above have been chosen by Affärsvärlden¹¹⁶. It is important to note that these are chosen from other factors than what we normally consider for a branch in the correlation matrix. In the correlation matrix we expect the different branches to react similarly in the correlation matrix; in “real” branches this does not always happen though since companies in the same branch often have at least some differences in stock movements.

¹¹⁵ Affärsvärlden, 2003-08-30, www.affarsvarlden.se

¹¹⁶ Ibid.

Data for the 197 Swedish companies

Number	1	2	3	4	5
Namn	Svenska Cell Karlshamns AB Holmen AB s Höganäs AB SSAB Svensk				
Nbr Days	4147	1532	4148	2328	3522
Holes	10	6	155	0	37
Holes percent	0,0024108	0,0039139	0,037358	0	0,010502
Avg return	0,00024685	0,000288549	0,000115566	-7,7509E-05	9,7595E-06
std	0,017966154	0,020754639	0,021257324	0,022244869	0,022421634
Branch	1	1	1	1	1
Number	6	7	8	9	10
Namn	Stora Enso ORottneros AB Norsk Hydro KLIPPAN AB, Boliden AB, B				
Nbr Days	1139	2955	4148	2178	1055
Holes	0	193	1265	132	0
Holes percent	0	0,065291	0,30489	0,060578	0
Avg return	-0,00010642	0,000297777	0,00019916	0,000284698	-0,00261211
std	0,025318742	0,025528338	0,028171158	0,03318697	0,037141849
Branch	1	1	1	1	1
Number	11	12	13	14	15
Namn	Rörvik TimbeNorth AtlanticScanMining ATrelleborg ABSCANIA AB s				
Nbr Days	1519	1520	1036	4148	1823
Holes	205	77	242	97	0
Holes percent	0,13487	0,050625	0,23337	0,023379	0
Avg return	-0,0002628	0,000284255	-0,00082104	0,000125189	-0,00024852
std	0,039083296	0,045949834	0,072634378	0,017118647	0,017936667
Branch	1	1	1	2	2
Number	16	17	18	19	20
Namn	NIBE Industri Peab AB ser. Närkes ElektrHexagon AB Seco Tools A				
Nbr Days	1525	3555	4147	3685	3500
Holes	32	467	1245	990	841
Holes percent	0,02097	0,13133	0,30014	0,26858	0,24022
Avg return	0,00095678	0,000583803	0,000440214	0,000199024	8,98459E-05
std	0,018386586	0,019057032	0,019806512	0,020545583	0,020752907
Branch	2	2	2	2	2
Number	21	22	23	24	25
Namn	Cardo AB, CAVolvo, AB serNCC AB ser. SAAB AB ser Skanska AB s				
Nbr Days	2099	4148	3715	1274	4148
Holes	5	5	546	0	23
Holes percent	0,002381	0,0012051	0,14693	0	0,0055435
Avg return	-0,00025622	-0,00044145	-0,00078682	0,000461372	-0,00056857
std	0,020798054	0,020821469	0,02135941	0,021433155	0,021814305
Branch	2	2	2	2	2
Number	26	27	28	29	30
Namn	Sapa AB, SAFFagerhult, ABGunnebo AB, SKF, AB ser. Munters AB, M				
Nbr Days	1543	1548	2399	4148	1435
Holes	4	185	31	3	3
Holes percent	0,0025907	0,11943	0,012917	0,00072307	0,0020891
Avg return	0,000102924	0,000313716	0,000744147	0,000347016	0,000652606
std	0,0219967	0,022173203	0,022190749	0,023022606	0,023564801
Branch	2	2	2	2	2

Number	31	32	33	34	35
Namn	Karolin Mach Midway Hold Haldex AB, H Ångpanneför B&N Nordsjöf				
Nbr Days	1323	3315	3623	4148	3032
Holes	49	695	948	704	126
Holes percent	0,037009	0,20959	0,26159	0,16968	0,041543
Avg return	-0,0005833	-0,00038226	-0,00025776	-0,00017657	-0,00049449
std	0,024040637	0,024384286	0,024738418	0,024924408	0,025823502
Branch	2	2	2	2	2
Number	36	37	38	39	40
Namn	OEM Internat Gorthon Line Malmbergs E Finnveden A Broström AB				
Nbr Days	3559	1531	1089	4148	1275
Holes	1017	176	343	239	17
Holes percent	0,28567	0,11488	0,31468	0,057604	0,013323
Avg return	-0,00028599	-0,00057372	-0,00014646	-0,00116238	0,001151757
std	0,025913402	0,028014131	0,028167934	0,028530722	0,029391345
Branch	2	2	2	2	2
Number	41	42	43	44	45
Namn	ASSA ABLOY Profil Grupper Xpon Card Gr Beijer AB, G & HL Display AB				
Nbr Days	2176	1522	2399	3664	2399
Holes	3	313	304	981	112
Holes percent	0,001378	0,20552	0,12667	0,26767	0,046667
Avg return	-9,8779E-05	-2,1902E-06	0,000234972	-0,00042917	-0,00048261
std	0,029557546	0,029965966	0,030127602	0,031800286	0,031805583
Branch	2	2	2	2	2
Number	46	47	48	49	50
Namn	Strålfors AB s SWECO AB s Elanders AB Nefab AB ser Consilium AB				
Nbr Days	4148	1208	3641	1785	2302
Holes	727	192	882	207	217
Holes percent	0,17522	0,15881	0,24217	0,1159	0,094225
Avg return	-0,00050032	0,000670952	-0,00106782	0,000335316	-7,9805E-05
std	0,031837282	0,032717817	0,03340731	0,033629005	0,033732187
Branch	2	2	2	2	2
Number	51	52	53	54	55
Namn	Beijer Alma A Westergyllen Concordia Mäntellecta AB Elektronik Gru				
Nbr Days	3929	3490	3542	3306	3552
Holes	874	1182	297	926	1038
Holes percent	0,22239	0,33858	0,083827	0,28001	0,29215
Avg return	-0,00064804	-0,00137903	0,000149303	-0,00111239	-0,00026787
std	0,033920477	0,034509809	0,036483779	0,037182086	0,037927636
Branch	2	2	2	2	2
Number	56	57	58	59	60
Namn	Bergman & B Segerström & All Cards Ser Semcon AB, S T T Systems				
Nbr Days	4148	1590	1026	1540	1414
Holes	555	53	148	6	109
Holes percent	0,13377	0,033312	0,14411	0,0038936	0,077032
Avg return	-0,00094296	0,001076419	0,000183775	-0,00210045	-0,00087093
std	0,038140637	0,038963603	0,039189274	0,04603499	0,04671721
Branch	2	2	2	2	2

Number	61	62	63	64	65
Namn	Duroc AB serABB Ltd, ABArray AB ser.Sardus, AB, SCloetta Fazer				
Nbr Days	1016	1021	1835	1572	2289
Holes	66	0	147	17	32
Holes percent	0,064897	0	0,080065	0,010807	0,013974
Avg return	-0,00192347	-0,00199319	-0,00457372	0,000415664	0,000430027
std	0,049717778	0,050592881	0,094888843	0,017813487	0,018388881
Branch	2	2	2	3	3
Number	66	67	68	69	70
Namn	Swedish MatSvedbergs i DAutoliv Inc. SLindex, AB, LBilia AB ser. A				
Nbr Days	1794	1447	2424	2069	3703
Holes	0	121	2	3	668
Holes percent	0	0,083564	0,00082474	0,0014493	0,18035
Avg return	0,000659248	-7,9384E-05	-0,00019583	-0,00054065	0,000290483
std	0,018857194	0,020339689	0,022284819	0,023763642	0,025150818
Branch	3	3	3	3	3
Number	71	72	73	74	75
Namn	Electrolux, ABExpanda AB sNew Wave GrHennes & Ma Spendrups Br				
Nbr Days	4148	1519	1398	4148	3608
Holes	2	131	44	153	645
Holes percent	0,00048204	0,086184	0,031451	0,036876	0,17872
Avg return	-0,00018861	-0,00037956	0,00118421	-0,00020603	-0,00025718
std	0,026613583	0,026623926	0,027974319	0,029278997	0,029453707
Branch	3	3	3	3	3
Number	76	77	78	79	80
Namn	Stena Line ABsinterCast ABWedins Skor DORO AB serOpcon AB, OP				
Nbr Days	2738	2566	1515	2442	1138
Holes	374	55	225	92	205
Holes percent	0,13655	0,021426	0,14842	0,037659	0,17998
Avg return	-0,00149179	-0,00052877	-0,00236998	-0,00237332	-3,4257E-05
std	0,03382322	0,043639272	0,050451935	0,051166939	0,052843826
Branch	3	3	3	3	3
Number	81	82	83	84	85
Namn	Getinge AB sAstraZeneca Elekta AB serKaro Bio AB, Biacore Intern				
Nbr Days	2549	1074	2351	1027	1653
Holes	1	0	1	9	3
Holes percent	0,00039216	0	0,00042517	0,0087549	0,0018138
Avg return	0,000567841	4,49697E-05	0,001023149	-0,00022417	0,000829273
std	0,021461734	0,02183068	0,036576755	0,039886373	0,040981513
Branch	4	4	4	4	4
Number	86	87	88	89	90
Namn	Wilh. Soness Medivir AB seActive Biotec Ortivus AB seMaxim Pharm				
Nbr Days	1028	1665	3731	1636	1432
Holes	71	9	151	92	0
Holes percent	0,068999	0,0054022	0,040461	0,0562	0
Avg return	0,000176115	-0,00015195	-0,00163248	-0,00048137	-0,00056527
std	0,042295596	0,043216751	0,044414741	0,049692461	0,051778271
Branch	4	4	4	4	4

Number	91	92	93	94	95
Namn	BioGaia AB s Artimplant AB Dorigene, Inc. Biolin AB, BL RaySearch La				
Nbr Days	1288	1424	1663	1370	3293
Holes	19	24	6	53	993
Holes percent	0,01474	0,016842	0,0036058	0,038658	0,30146
Avg return	-0,00036949	-0,00172966	5,02017E-05	-0,00284169	-0,00166921
std	0,053010839	0,054076219	0,061824331	0,067288881	0,06852916
Branch	4	4	4	4	4
Number	96	97	98	99	100
Namn	BioPhausia ALundbergföreKungsleden ACastellum AB Törnet, Fastig				
Nbr Days	1330	3687	1068	1541	1800
Holes	24	540	0	0	2
Holes percent	0,018032	0,14642	0	0	0,0011105
Avg return	9,85518E-05	0,000576101	0,000815495	0,000578302	0,000334258
std	0,082398152	0,014186056	0,014601589	0,01588094	0,016264705
Branch	4	5	5	5	5
Number	101	102	103	104	105
Namn	Wihlborgs FaWallenstam BÖresund, InveRatos AB ser Capona AB, C				
Nbr Days	3357	2992	3688	3688	1088
Holes	788	86	684	109	15
Holes percent	0,23466	0,028734	0,18542	0,029547	0,013774
Avg return	0,000454066	0,000972536	0,000316459	0,000404026	0,000440429
std	0,016844555	0,017071211	0,018165005	0,018628903	0,018821808
Branch	5	5	5	5	5
Number	106	107	108	109	110
Namn	Drott AB ser. Pandox AB, PFöreningsSpaCustos, AB, Cheba Fastighe				
Nbr Days	1205	1521	2030	3686	2281
Holes	0	37	1	709	381
Holes percent	0	0,02431	0,00049237	0,1923	0,16696
Avg return	0,000280436	9,99393E-05	-0,00015117	-0,00021573	0,000575286
std	0,019046025	0,019066475	0,019623863	0,019793588	0,020182985
Branch	5	5	5	5	5
Number	111	112	113	114	115
Namn	Luxonen S.A.Svenska Han Investor AB sMandamus F LjungbergGru				
Nbr Days	3284	3169	4148	1277	2259
Holes	156	11	240	22	783
Holes percent	0,047489	0,00347	0,057845	0,017214	0,34646
Avg return	0,00093165	0,000317241	-0,00048115	0,000578302	0,000553622
std	0,02107692	0,021696087	0,021736839	0,02226333	0,022268802
Branch	5	5	5	5	5
Number	116	117	118	119	120
Namn	Latour, Inves Nordea AB, NGeveko, AB sSvolder AB s Skandinavisk				
Nbr Days	3688	1401	3702	2522	3540
Holes	561	0	1004	50	26
Holes percent	0,15207	0	0,27113	0,019818	0,0073426
Avg return	0,000108329	-0,00033285	-2,8346E-05	-0,00076757	-8,2554E-05
std	0,024598235	0,025301684	0,026582058	0,026792698	0,027073658
Branch	5	5	5	5	5

Number	121	122	123	124	125
Namn	Säkl AB, SAE Kinnevik, Ind Vostok Nafta, Invik & Co ABHOIST INTER				
Nbr Days	1548	2676	1318	1456	1302
Holes	333	1	0	25	1
Holes percent	0,21498	0,00037355	0	0,017159	0,00076746
Avg return	-0,00034818	-0,0002027	0,001043912	-0,00037484	0,000125963
std	0,028478826	0,029468352	0,033543232	0,034509668	0,036155843
Branch	5	5	5	5	5
Number	126	127	128	129	130
Namn	Fast Partner ÅsalusAnsvar MEDICOVER OM AB, OM, Sskandia Förs				
Nbr Days	2362	1635	1515	3055	4148
Holes	279	127	258	22	7
Holes percent	0,11807	0,077628	0,17018	0,007199	0,0016872
Avg return	-1,4746E-05	-0,00071481	-0,00044576	-0,00058568	-0,00126542
std	0,036262744	0,036337384	0,037316292	0,041410171	0,043576162
Branch	5	5	5	5	5
Number	131	132	133	134	135
Namn	Avanza AB, ARealia AB, CöSkanditek IndBure Equity ALedstiernan A				
Nbr Days	2675	1368	3676	2450	2068
Holes	78	0	761	2	250
Holes percent	0,029148	0	0,20696	0,00081599	0,12083
Avg return	-0,00118498	-0,00011717	-0,00092555	-0,00326944	-0,00181797
std	0,043702529	0,044274222	0,048212897	0,051875526	0,095997356
Branch	5	5	5	5	5
Number	136	137	138	139	140
Namn	TietoEnator AScribona AB WM-data AB Frango AB, FIBS AB ser. B				
Nbr Days	1009	2660	3663	1061	4148
Holes	0	66	933	30	399
Holes percent	0	0,024803	0,25464	0,028249	0,096168
Avg return	-0,00095855	-0,00086669	-0,00164073	-0,00088155	-0,00183188
std	0,036948543	0,037274444	0,041326572	0,042459027	0,042882172
Branch	6	6	6	6	6
Number	141	142	143	144	145
Namn	SECTRA AB sTeleca AB se Prevas AB seTicket Travel ProAct IT Gro				
Nbr Days	1096	1601	1287	1558	1017
Holes	4	1	106	10	7
Holes percent	0,0036463	0,00062422	0,082298	0,0064144	0,0068762
Avg return	0,001168634	-0,00070015	-0,00206308	-0,00210937	-0,0010181
std	0,043876385	0,044134141	0,044223772	0,044296536	0,04609347
Branch	6	6	6	6	6
Number	146	147	148	149	150
Namn	NOVOTEK ABTurnIT AB se Intentia InternTechnology NBoss Media A				
Nbr Days	1016	1315	1660	1142	1019
Holes	128	17	0	1	0
Holes percent	0,12586	0,012918	0	0,00087489	0
Avg return	-0,00041552	-0,00383422	-0,00322388	-0,000954	-0,00051738
std	0,047976947	0,048633204	0,050434929	0,051370225	0,051742625
Branch	6	6	6	6	6

Number	151	152	153	154	155
Namn	ReadSoft AB HiQ InternatioAlfaskop AB, Industrial andPricer AB ser				
Nbr Days	1021	1070	1135	1313	1693
Holes	2	0	2	0	1
Holes percent	0,0019569	0	0,0017606	0	0,00059032
Avg return	-0,00094568	-0,00063937	-0,00163765	-0,00295613	-0,00120868
std	0,053038465	0,053081716	0,055427529	0,056462502	0,056892508
Branch	6	6	6	6	6
Number	156	157	158	159	160
Namn	Protect Data Modul 1 Data MSC Konsult Know IT AB, Resco AB ser				
Nbr Days	1321	1542	1294	1090	1675
Holes	33	27	350	115	119
Holes percent	0,024962	0,017498	0,27027	0,10541	0,071002
Avg return	-0,00092809	-0,00383281	-0,00227232	-0,00295235	-0,00252854
std	0,057725239	0,057976759	0,058418404	0,058988534	0,062000678
Branch	6	6	6	6	6
Number	161	162	163	164	165
Namn	Telelogic AB, Softronic AB RKS AB ser. Fingerprint C Jeeves Inform				
Nbr Days	1093	1155	1046	1272	1063
Holes	0	116	170	19	294
Holes percent	0	0,10035	0,16237	0,014925	0,27632
Avg return	-0,00039752	-0,00225254	-0,00259492	-0,0003601	-0,00068195
std	0,062804402	0,065019295	0,065342637	0,068204374	0,068556673
Branch	6	6	6	6	6
Number	166	167	168	169	170
Namn	Trio AB, TRIQULT Mobile C Framfab AB, Icon MedialabNOCOM AB s				
Nbr Days	1148	1392	1020	1166	1137
Holes	12	502	0	49	66
Holes percent	0,010444	0,36037	0	0,041988	0,057996
Avg return	-0,00221225	-0,00144204	-0,00407618	-0,00354198	-0,00128407
std	0,068969753	0,071905915	0,077299391	0,082899078	0,08463238
Branch	6	6	6	6	6
Number	171	172	173	174	175
Namn	Labs2 Group Adera AB ser Effnet Group Enea Data ABMandator AB,				
Nbr Days	1400	1029	1074	1186	1635
Holes	60	103	2	0	1
Holes percent	0,042827	0,1	0,0018605	0	0,00061125
Avg return	-0,00439895	-0,0052235	-0,00475368	-0,00473564	-0,00516478
std	0,09129297	0,10071572	0,10110027	0,120503722	0,318430023
Branch	6	6	6	6	6
Number	176	177	178	179	180
Namn	Tele2 AB ser. Europolitan V Nokia Abp, SDNolato AB se LGP Telecom				
Nbr Days	1795	2278	4147	4148	1531
Holes	0	8	480	822	9
Holes percent	0	0,0035103	0,11572	0,19812	0,0058747
Avg return	-2,2739E-05	-0,00071183	-0,00040328	-0,00094847	-0,00083099
std	0,031980411	0,033534796	0,039427519	0,042505104	0,046323256
Branch	7	7	7	7	7

Number	181	182	183	184	185
Namn	Ericsson, Tel Teligent AB, Net Insight AB, Freetel AB se, VLT AB ser. B				
Nbr Days	4148	1070	1034	3542	4144
Holes	1	16	0	626	1382
Holes percent	0,00024102	0,014939	0	0,17669	0,33341
Avg return	-0,00168686	-0,00125805	-0,00309775	-0,00261126	-0,00022237
std	0,048255596	0,059984275	0,065934265	0,104392217	0,02238109
Branch	7	7	7	7	8
Number	186	187	188	189	190
Namn	TV4 AB ser. A, Esselte AB se, Modern Time Observer AB, Digital Illusion				
Nbr Days	2320	3898	1055	1215	1335
Holes	99	218	0	7	17
Holes percent	0,042654	0,055912	0	0,0057566	0,012725
Avg return	0,00047159	-0,00068873	-0,00050261	-0,00054425	0,000761826
std	0,028360992	0,028526554	0,040106583	0,040295858	0,04665809
Branch	8	8	8	8	8
Number	191	192	193	194	195
Namn	MTV Produkt, SkiStar AB se, Securitas AB Berg & Co, A Optimail AB s				
Nbr Days	1567	2263	3016	3557	1266
Holes	286	317	68	1284	118
Holes percent	0,1824	0,14002	0,022539	0,36088	0,093133
Avg return	-0,00146115	0,000734453	-0,00041157	-0,00046676	-0,00174623
std	0,055812001	0,024771202	0,029201421	0,035298869	0,041593354
Branch	8	9	9	9	9
Number	196	197			
Namn	Poolia AB ser, Formo Services AB, FORM, SE0000458390				
Nbr Days	1020	1256			
Holes	55	261			
Holes percent	0,053869	0,20764			
Avg return	-0,00045884	-0,00294341			
std	0,044260594	0,115385869			
Branch	9	9			

A.3.2 U.S. Industry portfolios

Overview of the 48 U.S. data industry portfolios.

Nr 1 Agriculture production	Publishing-printing
0100-0199 Agric production - crops	2730-2739 Books: publishing-printing
0200-0299 Agric production - livestock	2740-2749 Misc. publishing
0700-0799 Agricultural services	2770-2771 Greeting card publishing
0910-0919 Commercial fishing	2780-2789 Book binding
2048-2048 Prepared feeds for animals	2790-2799 Service industries for print trade
2 Food Production	9 Consumer Goods
2000-2009 Food and kindred products	2047-2047 Dog and cat food
2010-2019 Meat products	2391-2392 Curtains, home furnishings
2020-2029 Dairy products	2510-2519 Household furniture
2030-2039 Canned-preserved fruits/vegs	2590-2599 Misc. furniture and fixtures
2040-2046 Grain mill products	2840-2843 Soap & other detergents
2050-2059 Bakery products	2844-2844 Perfumes cosmetics
2060-2063 Sugar/confectionery products	3160-3161 Luggage
2070-2079 Fats and oils	3170-3171 Handbags and purses
2090-2092 Misc. food preps	3172-3172 Personal leather goods, except
2095-2095 Roasted coffee	handbags
2098-2099 Misc. food preparations	3190-3199 Leather goods
3 Candy & Soda/beverage	3229-3229 Pressed and blown glass
2064-2068 Candy and other confectionery	3260-3260 Pottery and related products
2086-2086 Bottled-canned soft drinks	3262-3263 China and earthenware table
2087-2087 Flavoring syrup	articles
2096-2096 Potato chips	3269-3269 Pottery products
2097-2097 Manufactured ice	3230-3231 Glass products
4 Beer & Liquor production	3630-3639 Household appliances
2080-2080 Beverages	3750-3751 Motorcycles, bicycles and parts
2082-2082 Malt beverages	3800-3800 Misc. inst, photo goods, watches
2083-2083 Malt	3860-3861 Photographic equip (Kodak etc,
2084-2084 Wine	but also Xerox)
2085-2085 Distilled and blended liquors	3870-3873 Watches clocks and parts
5 Tobacco Producers	3910-3911 Jewelry-precious metals
2100-2199 Tobacco products	3914-3914 Silverware
6 Recreation	3915-3915 Jewelers' findings, materials
0920-0999 Fishing, hunting & trapping	3960-3962 Costume jewelry and notions
3650-3651 Household audio visual equip	3991-3991 Brooms and brushes
3652-3652 Phonographic records	3995-3995 Burial caskets
3732-3732 Boat building and repair	10 Apparel
3930-3931 Musical instruments	2300-2390 Apparel and other finished
3940-3949 Toys	products
7 Entertainment	3020-3021 Rubber and plastics footwear
7800-7829 Services - motion picture	3100-3111 Leather tanning and finishing
production and distribution	3130-3131 Boot, shoe cut stock, findings
7830-7833 Services - motion picture	3140-3149 Footwear except rubber
theatres	3150-3151 Leather gloves and mittens
7840-7841 Services - video rental	3963-3965 Fasteners, buttons, needles, pins
7900-7900 Services - amusement and	11 Healthcare
recreation	8000-8099 Services - health
7910-7911 Services - dance studios	12 Medical Equipment
7920-7929 Services - bands, entertainers	3693-3693 X-ray, electro medical app.
7930-7933 Services - bowling centers	3840-3849 Surg. & med instruments
7940-7949 Services - professional sports	3850-3851 Ophthalmic goods
7980-7980 Amusement and recreation	13 Pharmaceutical Products
services	2830-2830 Drugs
7990-7999 Services - misc. entertainment	2831-2831 Biological products
8 Printing and Publishing	2833-2833 Medicinal chemicals
2700-2709 Printing publishing and allied	2834-2834 Pharmaceutical preparations
2710-2719 Newspapers: publishing-printing	2835-2835 In vitro, in vivo diagnostics
2720-2729 Periodicals: publishing-printing	2836-2836 Biological products, except
	diagnostics

- 14 Chemicals
 2800-2809 Chemicals and allied products
 2810-2819 Industrial inorganic chemicals
 2820-2829 Plastic material & synthetic resin
 2850-2859 Paints
 2860-2869 Industrial organic chemicals
 2870-2879 Agriculture chemicals
 2890-2899 Misc. chemical products
- 15 Rubber and Plastic Products
 3031-3031 Reclaimed rubber
 3041-3041 Rubber & plastic
 3050-3053 Gaskets, hoses, etc
 3060-3069 Fabricated rubber products
 3070-3079 Misc. rubber products
 3080-3089 Misc. plastic products
 3090-3099 Misc. rubber and plastic
- 16 Textiles
 2200-2269 Textile mill products
 2270-2279 Floor covering mills
 2280-2284 Yarn and thread mills
 2290-2295 Misc. textile goods
 2297-2297 Non-woven fabrics
 2298-2298 Cordage and twine
 2299-2299 Misc. textile products
 2393-2395 Textile bags, canvas products
 2397-2399 Misc. textile products
- 17 Construction Materials
 0800-0899 Forestry
 2400-2439 Lumber and wood products
 2450-2459 Wood buildings-mobile homes
 2490-2499 Misc. wood products
 2660-2661 Building paper and board mills
 2950-2952 Paving & roofing materials
 3200-3200 Stone, clay, glass, concrete etc
 3210-3211 Flat glass
 3240-3241 Cement hydraulic
 3250-3259 Structural clay prods
 3261-3261 Vitreous china plumbing fixtures
 3264-3264 Porcelain electrical supply
 3270-3275 Concrete gypsum & plaster
 3280-3281 Cut stone and stone products
 3290-3293 Abrasive and asbestos products
 3295-3299 Non-metallic mineral products
 3420-3429 Hand tools and hardware
 3430-3433 Heating equip & plumbing fix
 3440-3441 Fabricated struct. metal products
 3442-3442 Metal doors, frames
 3446-3446 Architectural or ornamental metal work
 3448-3448 Pre-fab. metal buildings
 3449-3449 Misc. structural metal work
 3450-3451 Screw machine products
 3452-3452 Bolts, nuts screws
 3490-3499 Misc. fabricated metal products
 3996-3996 Hard surface floor cover
- 18 Construction
 1500-1511 Build construction - general
 1520-1529 Gen building contractors
 1530-1539 Operative builders
 1540-1549 Gen building contractors - non-residential
 1600-1699 Heavy Construction - not building contractors
- 1600-1699 Heavy Construction - not building contractors
 1700-1799 Construction - special contractors
- 19 Steel Works
 3300-3300 Primary metal industries
 3310-3317 Blast furnaces & steel works
 3320-3325 Iron & steel foundries
 3330-3339 Prim smelt-refin nonfer metals
 3340-3341 Secondary smelt-refin nonfer metals
 3350-3357 Rolling & drawing nonferrous metals
 3360-3369 Non-ferrous foundries and casting
 3370-3379 Steel works etc
 3390-3399 Misc. primary metal products
- 20 Fabricated Products
 3400-3400 Fabricated metal, except machinery and trans eq
 3443-3443 Fabricated plate work
 3444-3444 Sheet metal work
 3460-3469 Metal forgings and stampings
 3470-3479 Coating and engraving
- 21 Machinery
 3510-3519 Engines & turbines
 3520-3529 Farm and garden machinery
 3530-3530 Constr, mining material handling machinery
 3531-3531 Construction machinery
 3532-3532 Mining machinery, except oil field
 3533-3533 Oil field machinery
 3534-3534 Elevators
 3535-3535 Conveyors
 3536-3536 Cranes, hoists
 3538-3538 Machinery
 3540-3549 Metalworking machinery
 3550-3559 Special industry machinery
 3560-3569 General industrial machinery
 3580-3580 Refrig & service ind machines
 3581-3581 Automatic vending machines
 3582-3582 Commercial laundry and dry cleaning machines
 3585-3585 Air conditioning, heating, refrid eq
 3586-3586 Measuring and dispensing pumps
 3589-3589 Service industry machinery
 3590-3599 Misc. industrial and commercial equipment and mach
- 22 Electrical Equipment
 3600-3600 Elec mach eq & supply
 3610-3613 Elec transmission
 3620-3621 Electrical industrial appar
 3623-3629 Electrical industrial appar
 3640-3644 Electric lighting, wiring
 3645-3645 Residential lighting fixtures
 3646-3646 Commercial lighting
 3648-3649 Lighting equipment

- 3646-3646 Commercial lighting
 - 3648-3649 Lighting equipment
 - 3660-3660 Communication equip
 - 3690-3690 Miscellaneous electrical machinery and equip
 - 3691-3692 Storage batteries
 - 3699-3699 Electrical machinery and equip

- 23 Automobiles and Trucks
 - 2296-2296 Tire cord and fabric
 - 2396-2396 Auto trim
 - 3010-3011 Tires and inner tubes
 - 3537-3537 Trucks, tractors, trailers
 - 3647-3647 Vehicular lighting
 - 3694-3694 Elec eq, internal combustion engines
 - 3700-3700 Transportation equipment
 - 3710-3710 Motor vehicles and motor vehicle equip
 - 3711-3711 Motor vehicles & car bodies
 - 3713-3713 Truck & bus bodies
 - 3714-3714 Motor vehicle parts
 - 3715-3715 Truck trailers
 - 3716-3716 Motor homes
 - 3792-3792 Travel trailers and campers
 - 3790-3791 Misc. trans equip
 - 3799-3799 Misc. trans equip

- 24 Aircraft
 - 3720-3720 Aircraft & parts
 - 3721-3721 Aircraft
 - 3723-3724 Aircraft engines, engine parts
 - 3725-3725 Aircraft parts
 - 3728-3729 Aircraft parts

- 25 Shipbuilding, Railroad Equipment
 - 3730-3731 Ship building and repair
 - 3740-3743 Railroad Equipment

- 22 Electrical Equipment
 - 3600-3600 Elec mach eq & supply
 - 3610-3613 Elec transmission
 - 3620-3621 Electrical industrial appar
 - 3623-3629 Electrical industrial appar
 - 3640-3644 Electric lighting, wiring
 - 3645-3645 Residential lighting fixtures
 - 3646-3646 Commercial lighting
 - 3648-3649 Lighting equipment
 - 3660-3660 Communication equip
 - 3690-3690 Miscellaneous electrical machinery and equip
 - 3691-3692 Storage batteries
 - 3699-3699 Electrical machinery and equip

- 23 Automobiles and Trucks
 - 2296-2296 Tire cord and fabric
 - 2396-2396 Auto trim
 - 3010-3011 Tires and inner tubes
 - 3537-3537 Trucks, tractors, trailers
 - 3647-3647 Vehicular lighting
 - 3694-3694 Elec eq, internal combustion engines
 - 3700-3700 Transportation equipment
 - 3710-3710 Motor vehicles and motor vehicle equip
 - 3711-3711 Motor vehicles & car bodies
 - 3713-3713 Truck & bus bodies
 - 3714-3714 Motor vehicle parts
 - 3715-3715 Truck trailers
 - 3716-3716 Motor homes

- 3792-3792 Travel trailers and campers
 - 3790-3791 Misc. trans equip
- 24 Aircraft
 - 3720-3720 Aircraft & parts
 - 3721-3721 Aircraft
 - 3723-3724 Aircraft engines, engine parts
 - 3725-3725 Aircraft parts
 - 3728-3729 Aircraft parts

- 25 Shipbuilding, Railroad Equipment
 - 3730-3731 Ship building and repair
 - 3740-3743 Railroad Equipment

- 26 Defense
 - 3760-3769 Guided missiles and space vehicles
 - 3795-3795 Tanks and tank components
 - 3480-3489 Ordnance & accessories

- 27 Precious Metals
 - 1040-1049 Gold & silver ores

- 28 Non-Metallic and Industrial Metal Mining
 - 1000-1009 Metal mining
 - 1010-1019 Iron ores
 - 1020-1029 Copper ores
 - 1030-1039 Lead and zinc ores
 - 1050-1059 Bauxite and other aluminum ores
 - 1060-1069 Ferroalloy ores
 - 1070-1079 Mining
 - 1080-1089 Mining services
 - 1090-1099 Misc. metal ores
 - 1100-1119 Anthracite mining
 - 1400-1499 Mining non-metallic minerals

- 29 Coal
 - 1200-1299 Bituminous coal

- 24 Aircraft
 - 3720-3720 Aircraft & parts
 - 3721-3721 Aircraft
 - 3723-3724 Aircraft engines, engine parts
 - 3725-3725 Aircraft parts
 - 3728-3729 Aircraft parts

- 25 Shipbuilding, Railroad Equipment
 - 3730-3731 Ship building and repair
 - 3740-3743 Railroad Equipment

- 26 Defense
 - 3760-3769 Guided missiles and space vehicles
 - 3795-3795 Tanks and tank components
 - 3480-3489 Ordnance & accessories

- 27 Precious Metals
 - 1040-1049 Gold & silver ores

- 28 Non-Metallic and Industrial Metal Mining
 - 1000-1009 Metal mining
 - 1010-1019 Iron ores
 - 1020-1029 Copper ores
 - 1030-1039 Lead and zinc ores
 - 1050-1059 Bauxite and other aluminum ores
 - 1060-1069 Ferroalloy ores
 - 1070-1079 Mining
 - 1080-1089 Mining services
 - 1090-1099 Misc. metal ores
 - 1100-1119 Anthracite mining
 - 1400-1499 Mining and non-metallic minerals

- 29 Coal
 - 1200-1299 Bituminous coal

- 30 Petroleum and Natural Gas
1300-1300 Oil and gas extraction
1310-1319 Crude petroleum & natural gas
1320-1329 Natural gas liquids
1330-1339 Petroleum and natural gas
1370-1379 Petroleum and natural gas
1380-1380 Oil and gas field services
1381-1381 Drilling oil & gas wells
1382-1382 Oil-gas field exploration
1389-1389 Oil and gas field services
2900-2912 Petroleum refining
2990-2999 Misc. petroleum products
- 31 Utilities
4900-4900 Electric, gas, sanitary services
4910-4911 Electric services
4920-4922 Natural gas transmission
4923-4923 Natural gas transmission-distr
4924-4925 Natural gas distribution
4930-4931 Electric and other services
combined
4932-4932 Gas and other services combined
4939-4939 Combination utilities
4940-4942 Water supply
- 32 Communication
4800-4800 Communications
4810-4813 Telephone communications
4820-4822 Telegraph and other message
communication
4830-4839 Radio-TV Broadcasters
4840-4841 Cable and other pay TV services
4880-4889 Communications
4890-4890 Communication services
(Comsat)
4891-4891 Cable TV operators
4892-4892 Telephone interconnect
4899-4899 Communication services
- 33 Personal Services
7020-7021 Rooming and boarding houses
7030-7033 Camps and recreational vehicle
parks
7200-7200 Services - personal
7210-7212 Services - laundry, cleaners
7214-7214 Services - diaper service
7215-7216 Services - coin-op cleaners, dry
cleaners
7217-7217 Services - carpet, upholstery
cleaning
7219-7219 Services - laundry, cleaners
7220-7221 Services - photo studios, portrait
7230-7231 Services - beauty shops
7240-7241 Services - barber shops
7250-7251 Services - shoe repair
7260-7269 Services - funeral
7270-7290 Services - misc.
7270-7290 Services - misc.
7291-7291 Services - tax return
7292-7299 Services - misc.
7395-7395 Services - photofinishing labs
(School pictures)
7500-7500 Services - auto repair, services
7520-7529 Services - automobile parking
7530-7539 Services - auto repair shops
7540-7549 Services - auto services, except
repair (car washes)
7600-7600 Services - Misc. repair services
7620-7620 Services - Electrical repair shops
7622-7622 Services - Radio and TV repair
shops
7623-7623 Services - Refridg and air
conditioner repair
8710-8713 Services - engineering,
- 7629-7629 Services - Electrical repair shops
7630-7631 Services - Watch, clock and
jewelry repair
7640-7641 Services - Reupholster furniture
repair
7690-7699 Services - Misc. repair shops
8100-8199 Services - legal
8200-8299 Services - educational
8300-8399 Services - social services
8400-8499 Services - museums, galleries,
botanic gardens
8600-8699 Services - membership
organizations
8800-8899 Services - private households
- 34 Business Services
2750-2759 Commercial printing
3993-3993 Signs, advertising specialty
7218-7218 Services - industrial launderers
7300-7300 Services - business services
7310-7319 Services - advertising
7320-7329 Services - credit reporting
agencies, collection services
7330-7339 Services - mailing, reproduction,
commercial art
7340-7342 Services - services to dwellings,
other buildings
7349-7349 Services - cleaning and building
maintenance
7350-7351 Services - misc. equip rental and
leasing
7352-7352 Services - medical equip rental
7353-7353 Services - heavy construction
equip rental
7359-7359 Services - equip rental and
leasing
7360-7369 Services - personnel supply
services
7370-7372 Services - computer
programming and data processing
7374-7374 Services - computer processing,
7375-7375 Services - information retrieval
services
7376-7376 Services - computer facilities
management service
7377-7377 Services - computer rental and
leasing
7378-7378 Services - computer maintenance
and repair
7379-7379 Services - computer related
services
7380-7380 Services - misc. business
services
7381-7382 Services - security
7383-7383 Services - news syndicates
7384-7384 Services - photofinishing labs
7385-7385 Services - telephone
interconnections
7389-7390 Services - misc. business s
7391-7391 Services - R&D labs
7392-7392 Services - management
consulting & P.R.
7393-7393 Services - detective and
protective (ADT)
7394-7394 Services - equipment rental &
leasing
7396-7396 Services - trading stamp services
7397-7397 Services - commercial testing
labs
7399-7399 Services - business services
7510-7519 Services - truck, auto, trailer
rental and leasing
8700-8700 Services - engineering,
accounting, research, management
3220-3221 Glass containers
3410-3412 Metal cans and containers

accounting, surveying
 8720-8721 Services - accounting, auditing,
 bookkeeping
 8730-8734 Services - research, development,
 testing labs
 8740-8748 Services - management, public
 relations, consulting
 8900-8910 Services - misc.
 8911-8911 Services - engineering &
 architect
 8920-8999 Services - misc.

35 Comps Computers
 3570-3579 Office computers
 3680-3680 Computers
 3681-3681 Computers - mini
 3682-3682 Computers - mainframe
 3683-3683 Computers - terminals
 3684-3684 Computers - disk & tape drives
 3685-3685 Computers - optical scanners
 3686-3686 Computers - graphics
 3687-3687 Computers - office automation
 systems
 3688-3688 Computers - peripherals
 3689-3689 Computers - equipment
 3695-3695 Magnetic and optical recording
 media
 7373-7373 Computer integrated systems
 design

36 Chips Electronic Equipment
 3622-3622 Industrial controls
 3661-3661 Telephone and telegraph
 apparatus
 3662-3662 Communications equipment
 3663-3663 Radio TV comm. equip &
 apparatus
 3664-3664 Search, navigation, guidance
 systems
 3665-3665 Training equipment & simulators
 3666-3666 Alarm & signaling products
 3669-3669 Communication equipment
 3670-3679 Electronic components
 3810-3810 Search, detection, navigation,
 guidance
 3812-3812 Search, detection, navigation,
 guidance

37 Lab Eq Measuring and Control Equipment
 3811-3811 Engr. lab and research equipment
 3820-3820 Measuring and controlling
 equipment
 3821-3821 Lab apparatus and furniture
 3822-3822 Automatic controls
 3823-3823 Industrial measurement instrum.
 3824-3824 Totalizing fluid meters
 3825-3825 Elec meas & test instrum.
 3826-3826 Lab analytical instruments
 3827-3827 Optical instrum. and lenses
 3829-3829 Meas and control devices
 3830-3839 Optical instrum and lenses

38 Paper Business Supplies
 2520-2549 Office furniture and fixtures
 2600-2639 Paper and allied products
 2670-2699 Paper and allied products
 2760-2761 Manifold business forms
 3950-3955 Pens pencils and office supplies

39 Boxes Shipping Containers
 2440-2449 Wood containers
 2640-2659 Paperboard containers, boxes,
 drums, tubs

mining equipment
 5083-5083 Wholesale - farm and garden

40 Trans Transportation
 4000-4013 Railroads-line haul
 4040-4049 Railway express service
 4100-4100 Transit and passenger trans
 4110-4119 Local passenger trans
 4120-4121 Taxicabs
 4130-4131 Intercity bus trans (Greyhound)
 4140-4142 Bus charter
 4150-4151 School buses
 4170-4173 Motor vehicle terminals, service
 facilities
 4190-4199 Misc. transit and passenger
 transportation
 4200-4200 Motor freight trans,
 warehousing
 4210-4219 Trucking
 4220-4229 Warehousing and storage
 4230-4231 Terminal facilities - motor
 freight
 4240-4249 Transportation
 4400-4499 Water transport
 4500-4599 Air transportation
 4600-4699 Pipelines, except natural gas
 4700-4700 Transportation services
 4710-4712 Freight forwarding
 4720-4729 Travel agencies, etc
 4730-4739 Arrange trans - freight and
 cargo
 4740-4749 Rental of railroad cars
 4780-4780 Misc. services incidental to
 trans
 4782-4782 Inspection and weighing
 services
 4783-4783 Packing and crating
 4784-4784 Fixed facilities for vehicles, not
 elsewhere classified
 4785-4785 Motor vehicle inspection
 4789-4789 Transportation services

41 Wholesale
 5000-5000 Wholesale - durable goods
 5010-5015 Wholesale - autos and parts
 5020-5023 Wholesale - furniture and home
 furnishings
 5030-5039 Wholesale - lumber and
 construction materials
 5040-5042 Wholesale - professional and
 commercial equipment and supplies
 5043-5043 Wholesale - photographic
 equipment
 5044-5044 Wholesale - office equipment
 5045-5045 Wholesale - computers
 5046-5046 Wholesale - commercial equip
 5047-5047 Wholesale - medical, dental
 equip
 5048-5048 Wholesale - ophthalmic goods
 5049-5049 Wholesale - professional equip
 and supplies
 5050-5059 Wholesale - metals and
 minerals
 5060-5060 Wholesale - electrical goods
 5063-5063 Wholesale - electrical apparatus
 and equipment
 5064-5064 Wholesale - electrical appliance
 TV and radio
 5065-5065 Wholesale - electronic parts
 5070-5078 Wholesale - hardware,
 plumbing, heating equip
 5080-5080 Wholesale - machinery and
 equipment
 5081-5081 Wholesale - machinery and
 equipment
 5082-5082 Wholesale - construction and
 equipment stores
 5710-5719 Retail - home furnishings stores

machinery
 5084-5084 Wholesale - industrial machinery and equipment
 5085-5085 Wholesale - industrial supplies
 5086-5087 Wholesale - machinery and equipment (?)
 5088-5088 Wholesale - trans eq except motor vehicles
 5090-5090 Wholesale - misc. durable goods
 5091-5092 Wholesale - sporting goods, toys
 5093-5093 Wholesale - scrap and waste materials
 5094-5094 Wholesale - jewelry and watches
 5099-5099 Wholesale - durable goods
 5100-5100 Wholesale - nondurable goods
 5110-5113 Wholesale - paper and paper products
 5120-5122 Wholesale - drugs & proprietary
 5130-5139 Wholesale - apparel
 5140-5149 Wholesale - groceries & related prods
 5150-5159 Wholesale - farm products
 5160-5169 Wholesale - chemicals & allied prods
 5170-5172 Wholesale - petroleum and petrol prods
 5180-5182 Wholesale - beer, wine
 5190-5199 Wholesale - non-durable goods
 42 Retail
 5200-5200 Retail - bldg material, hardware, garden
 5210-5219 Retail - lumber & other building mat
 5220-5229 Retail
 5230-5231 Retail - paint, glass, wallpaper
 5250-5251 Retail - hardware stores
 5260-5261 Retail - nurseries, lawn, garden stores
 5270-5271 Retail - mobile home dealers
 5300-5300 Retail - general merchandise stores
 5310-5311 Retail - department stores
 5320-5320 Retail - general merchandise stores
 5330-5331 Retail - variety stores
 5334-5334 Retail - catalog showroom
 5340-5349 Retail
 5390-5399 Retail - Misc. general merchandise stores
 5400-5400 Retail - food stores
 5410-5411 Retail - grocery stores
 5412-5412 Retail - convenience stores
 5420-5429 Retail - meat, fish.
 5430-5439 Retail - fruit and vegetable markets
 5440-5449 Retail - candy, nut, confectionary stores
 5450-5459 Retail - dairy product stores
 5460-5469 Retail - bakeries
 5490-5499 Retail - miscellaneous food stores
 5500-5500 Retail - auto dealers and gas stations
 5510-5529 Retail - auto dealers
 5530-5539 Retail - auto and home supply stores
 5540-5549 Retail - gasoline service stations
 5550-5559 Retail - boat dealers
 5560-5569 Retail - recreational vehicle dealers
 5570-5579 Retail - motorcycle dealers
 5590-5599 Retail - automotive dealers
 5600-5699 Retail - apparel & acces.
 5700-5700 Retail - home furniture and
 6120-6129 S&Ls
 6130-6139 Agricultural credit institutions
 5720-5722 Retail - household appliance stores
 5730-5733 Retail - radio, TV and consumer electronic stores
 5734-5734 Retail - computer and computer software stores
 5735-5735 Retail - record and tape stores
 5736-5736 Retail - musical instrument stores
 5750-5799 Retail
 5900-5900 Retail - misc.
 5910-5912 Retail - drug & proprietary stores
 5920-5929 Retail - liquor stores
 5930-5932 Retail - used merchandise stores
 5940-5940 Retail - misc.
 5941-5941 Retail - sporting goods stores, bike shops
 5942-5942 Retail - book stores
 5943-5943 Retail - stationery stores
 5944-5944 Retail - jewelry stores
 5945-5945 Retail - hobby, toy and game shops
 5946-5946 Retail - camera and photo shop
 5947-5947 Retail - gift, novelty
 5948-5948 Retail - luggage
 5949-5949 Retail - sewing & needlework stores
 5950-5959 Retail
 5960-5969 Retail - non-store retailers (catalogs, etc)
 5970-5979 Retail
 5980-5989 Retail - fuel & ice stores (Penn Central Co)
 5990-5990 Retail - retail stores
 5992-5992 Retail - florists
 5993-5993 Retail - tobacco stores
 5994-5994 Retail - news dealers
 5995-5995 Retail - computer stores
 5999-5999 Retail stores
 43 Meals Restaurants, Hotels, Motels
 5800-5819 Retail - eating places
 5820-5829 Restaurants, hotels, motels
 5890-5899 Eating and drinking places
 7000-7000 Hotels, other lodging places
 7010-7019 Hotels motels
 7040-7049 Membership hotels and lodging
 7213-7213 Services - linen
 44 Banks Banking
 6000-6000 Depository institutions
 6010-6019 Federal reserve banks
 6020-6020 Commercial banks
 6021-6021 National commercial banks
 6022-6022 State banks - Fed Res. System
 6023-6024 State banks - not Fed Res. System
 6025-6025 National banks - Fed Res. System
 6026-6026 National banks - not Fed Res. System
 6027-6027 National banks, not FDIC
 6028-6029 Banks
 6030-6036 Savings institutions
 6040-6059 Banks (?)
 6060-6062 Credit unions
 6080-6082 Foreign banks
 6090-6099 Functions related to deposit banking
 6100-6100 No depository credit institutions
 6110-6111 Federal credit agencies
 6112-6113 FNMA

6140-6149 Personal credit institutions
 (Beneficial)
 6150-6159 Business credit institutions
 6160-6169 Mortgage bankers
 6170-6179 Finance lessors
 6190-6199 Financial services
 45 Insurance
 6300-6300 Insurance
 6310-6319 Life insurance
 6320-6329 Accident and health insurance
 6330-6331 Fire, marine, property-casualty
 ins
 6350-6351 Surety insurance
 6360-6361 Title insurance
 6370-6379 Pension, health, welfare funds
 6390-6399 Insurance carriers
 6400-6411 Insurance agents

 46 REEst Real Estate
 6500-6500 Real estate
 6510-6510 Real estate operators
 6512-6512 Operators - non-resident
 buildings
 6513-6513 Operators - apartment buildings
 6514-6514 Operators - other than apartment
 6515-6515 Operators - residential mobile
 home
 6517-6519 Lessors of real property
 6520-6529 Real estate
 6530-6531 Real estate agents and managers
 6532-6532 Real estate dealers
 6540-6541 Title abstract offices
 6550-6553 Real estate developers
 6590-6599 Real estate
 6610-6611 Combined real estate, insurance,
 etc
 47 Fin Trading
 6200-6299 Security and commodity brokers
 6700-6700 Holding, other investment offices
 6710-6719 Holding offices
 6720-6722 Investment offices
 6723-6723 Management investment, closed-
 end
 6724-6724 Unit investment trusts
 6725-6725 Face-amount certificate offices
 6730-6733 Trusts
 6740-6779 Investment offices
 6790-6791 Miscellaneous investing
 6792-6792 Oil royalty traders
 6793-6793 Commodity traders
 6794-6794 Patent owners & lessors
 6795-6795 Mineral royalty traders
 6798-6798 REIT
 6799-6799 Investors, NEC

 48 Other Miscellaneous (also includes other SIC
 codes not in a specific industry)
 3999-3999 Misc manufacturing industries
 4950-4959 Sanitary services
 4960-4961 Steam, air conditioning supplies
 4970-4971 Irrigation systems
 4990-4991 Cogeneration - SM power
 producer
 9900-9999 Not classifiable

A.4 Noh's Model

Noh's model mathematical description

$F(t)$ ¹¹⁷ is a function over the time series; the time average of the function is defined as:

$$\langle F(t) \rangle_T = \frac{1}{T} \sum_{t=1}^T F(t)$$

The time average of $F(t)$ is dependent of the length T of the time series. When $T \rightarrow \infty$ we get the following results:

$$\langle M_k(t) \rangle_\infty = 0 \qquad \langle M_k^2(t) \rangle_\infty = 0$$

To achieve the formulas above we used the following formulas (observe that $\delta_{kl} = 1$ when $k = l$ and $\delta_{kl} = 0$ when $k \neq l$):

$$\begin{aligned} \langle \eta_{b(k)}(t) \eta_{b(l)}(t) \rangle_\infty &= \delta_{b(k)b(l)} & \langle \varepsilon_k(t) \varepsilon_l(t) \rangle_\infty &= \delta_{kl} \\ \langle \eta_{b(k)}(t) \rangle_\infty &= 0 & \langle \varepsilon_k(t) \rangle_\infty &= 0 \\ \langle \eta_{b(k)}(t) \varepsilon_l(t) \rangle_\infty &= 0 & & \end{aligned}$$

C_{kl} , the correlation coefficient between the two companies l and b is the average over the product of the two normalized time series¹¹⁸,

$$C_{kl}(T) = \frac{1}{T} \sum_{t=1}^T M_k(t) M_l(t) = \langle M_k(t) M_l(t) \rangle_T$$

We can view the numbers $M_k(t)$ as entries in a $K \times T$ matrix M . For the $K \times K$ correlation matrix we then have¹¹⁹,

$$C(T) = \frac{1}{T} M M^\dagger = \langle M_k(t) M_l(t) \rangle_T$$

These averages depend on the length T of the time series, we therefore add T to the correlation matrix, within our model above one finds that for infinitely long time series¹²⁰, $T \rightarrow \infty$,

$$C_{kl}(\infty) = \lim_{T \rightarrow \infty} C_{kl}(T) = \frac{1}{1 + p_{b(k)}} (p_{b(k)} \delta_{b(k)b(l)} + \delta_{kl})$$

¹¹⁷ Thomas Guhr & Bernd Kälber, 28 Jun 2002, arXiv:cond-mat/0206577 v1 p. 4

¹¹⁸ Ibid.

¹¹⁹ Ibid.

¹²⁰ M. Marsili, cond-mat/0003241 referred to in Thomas Guhr & Bernd Kälber, 28 Jun 2002, arXiv:cond-mat/0206577 v1 p. 3

Using the formula above the matrix $C(\infty)$ now consists of B square blocks on the diagonal of dimension $\kappa_b \times \kappa_b$. The off-diagonal entries are $\frac{p_b}{1+p_b}$ for branch b , and

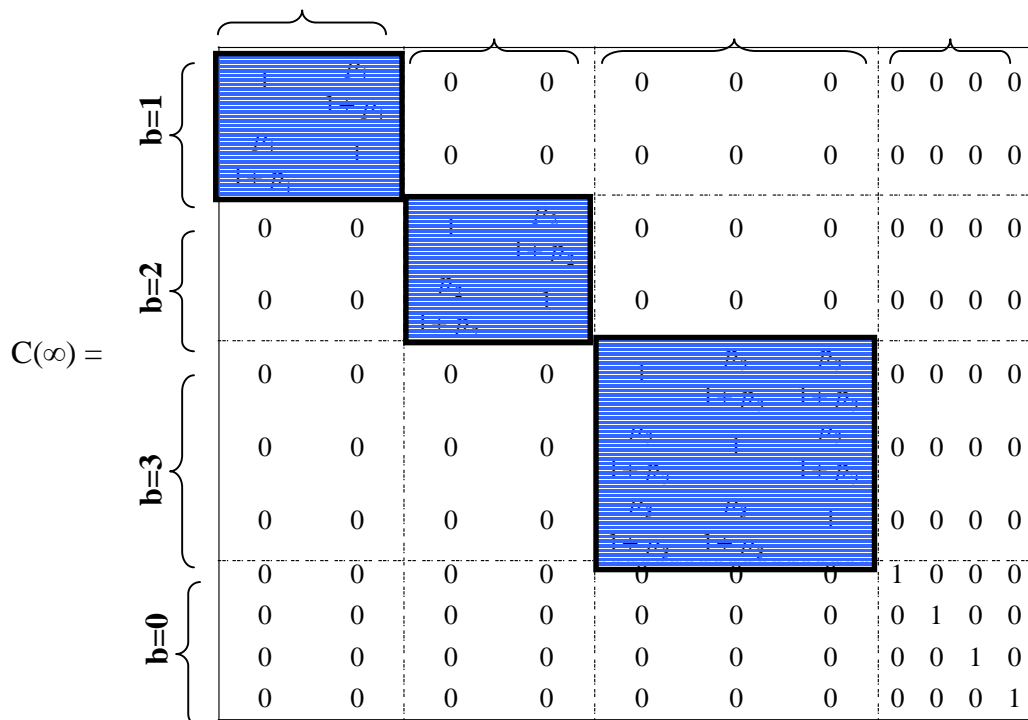
$\kappa \times \kappa$ unit matrix for companies which are in no branch. While the diagonal entries are all unity, all other entries are zero. The entries which are zero are correlations between:

- companies in *no* branch
- companies belonging to *different* branches
- one company that is in a branch when the other is *not*.

When the formula presented earlier,

$$C_{kl}(\infty) = \lim_{T \rightarrow \infty} C_{kl}(T) = \frac{1}{1+p_{b(k)}} (p_{b(k)} \delta_{b(k)b(l)} + \delta_{kl})$$

, is used an example of the correlation matrix could look like in the picture below.



An example of the correlation matrix created by Noh's method when $T=\infty$.

As mentioned earlier it has been shown that true correlations (for real stock data) can be buried under noise. The correlation matrix shown in the earlier picture that is described by the formula below does not contain noise since $T \rightarrow \infty$.

$$C_{kl}(\infty) = \lim_{T \rightarrow \infty} C_{kl}(T) = \frac{1}{1+p_{b(k)}} (p_{b(k)} \delta_{b(k)b(l)} + \delta_{kl})$$

Such a T value would not be possible in a real market; there will always be finite intervals between the updates of stock data. Often when large portfolios are analysed day data or hour data is used. If hour data is studied for i.e. 3 years we have $T \approx 20000$, a T value that is far from infinite. When situations with $T \ll \infty$ are studied it is clear that $\langle \varepsilon_k(t)\varepsilon_l(t) \rangle_T$, $\langle \eta_{b(k)}(t)\eta_{b(l)}(t) \rangle_T$ and $\langle \eta_{b(k)}(t)\varepsilon_l(t) \rangle_T$ no longer give zero or unity (as shown below), but instead some finite number¹²¹.

$$\begin{aligned} \langle \varepsilon_k(t)\varepsilon_l(t) \rangle_T &\neq \delta_{kl} & \langle \eta_{b(k)}(t)\eta_{b(l)}(t) \rangle_T &\neq \delta_{b(k)b(l)} \\ \langle \eta_{b(k)}(t)\varepsilon_l(t) \rangle_T &\neq 0 \end{aligned}$$

The finite numbers from the averages above describes the noise dressing¹²² of the financial correlation matrices essential to Noh's model. The always present finiteness of the time series introduces a random offset to every correlation coefficient; the true correlation coefficients are effectively buried under noise. The correlation matrix is then:

$$C = C_{true} + C_{noise}$$

We use standard statistics [24]¹²³ to attain the average of two time series $\langle \alpha_k(t)\alpha_l(t) \rangle_T$ to explain the introduction of the random elements. We have two time series of uncorrelated random numbers $\alpha_k(t)$ and $\alpha_l(t)$, $t = 1, \dots, T$, with standard normal distributions and zero mean value. $\langle \alpha_k(t)\alpha_l(t) \rangle_T$ is the average and a random number following a Gauss distribution (mean value is unity and the variance is $2/T$ if $k=l$. If $k \neq l$ it follows a Gauss distribution centered at mean value zero with variance $1/T$. This average can then be written as¹²⁴:

$$\langle \alpha_k(t)\alpha_l(t) \rangle_T = \delta_{kl} + \sqrt{\frac{1 + \delta_{kl}}{T}} a_{kl}$$

a_{kl} are all uncorrelated random numbers independent of T and with a normal distribution and zero mean value. For the correlation coefficients this yields the expression¹²⁵:

¹²¹ Thomas Guhr & Bernd Kälber, 28 Jun 2002, arXiv:cond-mat/0206577 v1 p. 5

¹²² Ibid.

¹²³ U. Krengel, 1991, Einführung in die Wahrscheinlichkeitstheorie und Statistik (Viewweg, Branunschweig) referred to by Thomas Guhr & Bernd Kälber, 28 Jun 2002, arXiv:cond-mat/0206577 v1 p. 5

¹²⁴ Thomas Guhr & Bernd Kälber, 28 Jun 2002, arXiv:cond-mat/0206577 v1 p. 5

¹²⁵ Ibid.

$$\begin{aligned}
C_{kl}(T) = & \sqrt{\frac{p_{b(k)}}{1+p_{b(k)}}} \sqrt{\frac{p_{b(l)}}{1+p_{b(l)}}} \left(\delta_{b(k)b(l)} + \sqrt{\frac{1+\delta_{b(k)b(l)}}{T}} a_{b(k)b(l)} \right) + \\
& + \frac{1}{\sqrt{1+p_{b(k)}}} \frac{1}{\sqrt{1+p_{b(l)}}} \left(\delta_{kl} + \sqrt{\frac{1+\delta_{kl}}{T}} a_{kl} \right) + \\
& \sqrt{\frac{p_{b(k)}}{1+p_{b(k)}}} \frac{1}{\sqrt{1+p_{b(l)}}} \frac{1}{\sqrt{T}} a_{b(k)l} + \sqrt{\frac{p_{b(l)}}{1+p_{b(l)}}} \frac{1}{\sqrt{1+p_{b(k)}}} \frac{1}{\sqrt{T}} a_{kb(l)}
\end{aligned}$$

In the correlation matrix above the first two terms stem from the averages $\langle \eta_{b(k)}(t) \eta_{b(l)}(t) \rangle_T$ and $\langle \varepsilon_k(t) \varepsilon_l(t) \rangle_T$. The last two terms stem from the average $\langle \eta_{b(k)}(t) \varepsilon_l(t) \rangle_T$. The equation above has four different sets of random numbers, $a_{b(k)b(l)}$, $a_{b(k)l}$ and $a_{kb(l)}$, to avoid overly complex notations no more indices are introduced to specify the difference. If $T \rightarrow \infty$ the equation above yields:

$$C_{kl}(\infty) = \lim_{T \rightarrow \infty} C_{kl}(T) = \frac{1}{1+p_{b(k)}} (p_{b(k)} \delta_{b(k)b(l)} + \delta_{kl})$$

A.5 Principal Component Analysis - PCA

PCA involves a mathematical procedure that transform a number of (possible) correlated variables into a smaller number of uncorrelated variables called Principal Components (PC). PCA is widely used in areas such as mathematics, physics, image processing and financial applications to extract the most important uncorrelated sources of information in multivariate systems. The aim of PCA is to reduce the dimensionality from a number of variables to something much less while preserving the variance-covariance structure¹²⁶ This means that one tries to explain the variance-covariance structure through a few linear combinations of the original variables. The idea is that the first few Principal Components account for a large proportion of the variability (80-90%) and thereby achieved the objective of dimension reduction. This means that one replaces the original number of variables by the first and few Principal Components (instead of using all Principal components)¹²⁷ The main advantages of a PCA approach are that the computational efficiency that results from lack of correlation between the Principal Components and the dimension reduction referring to using only a few number of the Principal Components¹²⁸

The input data used in PCA analysis must be stationary; this means that non-stationary data as stock prices or yields should be transformed into e.g. returns before the PCA is performed.¹²⁹ The returns used must also undergo a normalization process before the PCA is executed, this since the very first principal component otherwise will be dominated by the input variable with the largest degree of volatility¹³⁰. The normalisation process is simply

$$M_k = \frac{G_k - \langle G_k \rangle}{\sqrt{\langle G_k^2 \rangle - \langle G_k \rangle^2}} \text{ (Giving a mean of zero and the variance of one)}$$

By starting with a number of p random variables denoted X_1, X_2, \dots, X_n where one for the moment makes no assumption of multivariate normality, then consider the p linear combinations

$$\begin{aligned} Y_1 &= l_{11}X_1 + l_{21}X_2 + \dots + l_{p1}X_p \\ Y_2 &= l_{12}X_1 + l_{22}X_2 + \dots + l_{p2}X_p \\ &\vdots \\ &\vdots \\ Y_p &= l_{1p}X_1 + l_{2p}X_2 + \dots + l_{pp}X_p \end{aligned}$$

This can be written in vector form as $Y = L_i^T X$ with $i = 1, 2, \dots, p$

¹²⁶ Dodgson, Jeffrey. 2001

¹²⁷ Ibid.

¹²⁸ Johnson & Wichern, 2003

¹²⁹ Alexander Carol, 2002

¹³⁰ Ibid.

where

$$L_i = \begin{bmatrix} l_{1i} \\ l_{2i} \\ \cdot \\ \cdot \\ l_{pi} \end{bmatrix} \quad \text{and} \quad X = \begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ x_n \end{bmatrix}$$

If Σ denotes the variance-covariance matrix of X then

$$\text{Var}(Y_i) = L_i^T \Sigma L_i \quad \text{and} \quad \text{Cov}(Y_i, Y_k) = L_i^T \Sigma L_k$$

One requires uncorrelated linear combinations Y_i which maximize $\text{Var}(Y_i)$. Also one will impose the restriction that the coefficient vectors have unit length (that is $L_i^T L_i = 1$)

Then the result would be (Principal Component is abbreviated PC below)

$$PC_1 = L_1^T X \quad \text{that maximizes } \text{Var}(L_1^T X) \text{ subject to } L_1^T L_1 = 1$$

$$PC_2 = L_2^T X \quad \text{that maximizes } \text{Var}(L_2^T X) \text{ subject to } L_2^T L_2 = 1 \text{ and } \text{Cov}(L_1^T X, L_2^T X) = 0$$

·
·

$$PC_i = L_i^T X \quad \text{that maximizes } \text{Var}(L_i^T X) \text{ subject to } L_i^T L_i = 1 \text{ and } \text{Cov}(L_i^T X, L_k^T X) = 0 \text{ where } k < i$$

Principal component analysis (PCA) is intrinsically connected to eigenvalue and eigenvector theory as follows

Let $(\lambda_1, e_1), (\lambda_2, e_2), \dots, (\lambda_p, e_p)$ be the eigenvalue vector pairs of the matrix Σ

$$\text{where} \quad \lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$$

Then the i :th principal component is given by

$$Y_i = e_i^T X = e_{1i} X_1 + e_{2i} X_2 + \dots + e_{pi} X_p$$

$$\text{Given } \text{Var}(Y_i) = \lambda_i \text{ and } \text{Cov}(Y_i, Y_j) = 0.$$

Since the variance of each Principal Component is equal to the corresponding eigenvalue, and the total system variability is just the sum of all eigenvalues, one have

$$\text{proportion of variance} = \frac{\lambda_k}{\lambda_1 + \lambda_2 + \dots + \lambda_p}$$

that is explained by PC_k

PC_1 (the first PC) is the linear combination that accounts for the most variation.

PC_2 (the second PC) is the linear combination that accounts for most of the remaining variation, while being uncorrelated to PC_1 .

Each subsequent PC accounts for as much as possible of the remaining variation while being uncorrelated with previously determined PCs. The coefficients of each PC vector measure the importance of each variable. It is often possible to interpret the first few PCs in terms of an average/overall effect, or a contrast between groups of variables.

A.6 Program Code

A.6.1 Program Code Overview

To make it easier for the reader to study the details of the calculations and the calculations processes, and furthermore to present an opportunity to reproduce, replicate and expand the results in this thesis the Matlab 6.5 code is presented below. The code is fairly well documented; the sign % in Matlab 6.5 code identifies comments made by the authors.

Note: It is not always possible to just run the programs like they are; some settings in the beginning should generally be changed (see code for details). You also need stock data for the compilations. Stock data is not provided but the programs are suited for use with any stock data with similar characteristics as those found in the thesis.

The programs provided are:

Main_analysis.m

This is the main program used in the thesis for creating the portfolios, creating the new matrices with the filtering methods and plotting all the results.

MeanColumn.m

This program works with the column values. This program is not fully functional since it only plays a smaller part of the analysis. The main base should be a good starting point for more advanced studies in this area.

Data.m

This is the program used to replace the missing stock data values for the 197 Swedish companies.

A.6.2 Main_Analysis.m

```

close all
do_plot=0; do_calc=1;

if do_calc==1 %***** DO CALCULATIONS *****

OpenGL neverselect %Will use software OpenGL if choosen once. Will be much faster
clear all
stockdata_file=3; only_mvp=1; do_plot=0; fast_mvp=1;

switch stockdata_file
case 1
  'StockData_Final_LN_1009'
  NbrComp=20; % 197 all
  DataRaw = xlsread('StockData_Final_LN_1009');
  SizeXY=size(DataRaw); NbrDaysTotInData=SizeXY(1)-20; % Size of the data file
  DataAll=DataRaw(21:end,2:NbrComp+1); % y,x: Top Line:Names of Corps, First
line excluded:Dates,
  %-----
  NbrDaysInPeriod=5; %The number of days in one period/step.
  NbrPeriodsInTot=60; % 200! %The total number of periods for all the time
  NbrPeriodsInGather=50; %The number of periods in the Gather time. %and
use this data to create the future portfolios
  NbrPointsInYear=250;
  NbrDaysInMonth=20;
  GetRet=[0.002,0.0025,0.003]; AssetMin=-1; %Shortselling is permitted
  % GetRet=[0.001,0.00125,0.0015]; AssetMin=0; %Shortselling is NOT permitted
case 2
  '22 Companies'
  NbrComp=22;
  DataRaw = xlsread('gamla'); % GAML A
  DataAll=DataRaw(:,:); %y,x: Top Line:Names of Corps, First line
excluded:Dates,
  SizeXY=size(DataRaw); NbrDaysTotInData=SizeXY(1); % Size of the data file
  %-----
  NbrDaysInPeriod=12; %The number of days in one period/step.
  NbrPeriodsInTot=34; % OBS PERIODS! % 408 days %The total number of
periods for all the time
  NbrPeriodsInGather=10; %The number of periods in the Gather time. %and
use this data to create the future portfolios
  NbrPointsInYear=12;

  NbrDaysInMonth=12; % Actually months in in year for this example
  AssetMin=-1;
  GetRet=[0.01,0.02,0.03];

case 3

% '48 INDUSTRIES' % % *****INDUSTRIES*****
% NbrComp=48;
% DataRaw = xlsread('48Industry8shorter');
% DataAll=log(DataRaw(:,2:NbrComp+1)/100+1); %y,x: Top Line:Names of Corps,
First line excluded:Dates,
% SizeXY=size(DataRaw); NbrDaysTotInData=SizeXY(1); % Size of the data file
% %-----
% NbrDaysInPeriod=12; % months in period %The number of days(data
points actually) in one period/step.

```

```

% NbrPeriodsInTot=22; %OBS PERIODS=YEARS! %The total number of
periods for all the time
% NbrPeriodsInGather=2; %10 normally %The number of periods in the Gather
time. %and use this data to create the future portfolios
% NbrPointsInYear=12; %Nbr Month
% NbrDaysInMonth=12; %Actually months in in year for this example
%
% AssetMin=-1; GetRet=[0.03,0.04,0.05];
%

% '48 INDUSTRIES' % % *****INDUSTRIES*****
% NbrComp=48;
% DataRaw = xlsread('48Industry5shorter');
% DataAll=log(DataRaw(:,2:NbrComp+1)/100+1); %y,x: Top Line:Names of Corps,
First line excluded:Dates,
% SizeXY=size(DataRaw); NbrDaysTotInData=SizeXY(1); % Size of the data file
% %-----
% NbrDaysInPeriod=12; %months in period %The number of days(data
points actually) in one period/step.
% NbrPeriodsInTot=25; %OBS PERIODS=YEARS! %The total number of
periods for all the time
% NbrPeriodsInGather=10; %10 normally %The number of periods in the Gather
time. %and use this data to create the future portfolios
% NbrPointsInYear=12; %Nbr Month
% NbrDaysInMonth=12; %Actually months in in year for this example
% %AssetMin=-1; GetRet=[0.05,0.075,0.1];
% AssetMin=0; GetRet=[0.03,0.04,0.05];

'48 INDUSTRIES' % % *****INDUSTRIES*****
NbrComp=48;
DataRaw = xlsread('48Industry');
DataAll=log(DataRaw(3:end,2:NbrComp+1)/100+1); %y,x: Top Line:Names of Corps,
First line excluded:Dates,
SizeXY=size(DataRaw); NbrDaysTotInData=SizeXY(1)-2; % Size of the data file
%-----
NbrDaysInPeriod=12; %months in period %The number of days(data
points actually) in one period/step.
NbrPeriodsInTot=30; %OBS PERIODS=YEARS! %The total number of periods
for all the time
NbrPeriodsInGather=8; %10 normally %The number of periods in the Gather time.
%and use this data to create the future portfolios
NbrPointsInYear=12; %Nbr Month
NbrDaysInMonth=12; %Actually months in in year for this example

AssetMin=-1; GetRet=[0.03,0.04,0.05];

case 4
'43 INDUSTRIES - Long '
'Beginning at 1931-01 -> 2002-12'
NbrComp=43;
DataRaw = xlsread('43Industry');
DataAll=log(DataRaw(:,2:NbrComp+1)/100+1); %y,x: Top Line:Names of Corps,
First line excluded:Dates,
SizeXY=size(DataRaw); NbrDaysTotInData=SizeXY(1); % Size of the data file
%-----
NbrDaysInPeriod=12; %The number of days(data points actually) in one
period/step.
NbrPeriodsInTot=72; %OBS PERIODS! %The total number of periods for all
the time

```



```

    NbrPeriodsInGather=10; % 10 normally                                % The number of periods in the
Gather time. %and use this data to create the future portfolios
    NbrPointsInYear=12;
    NbrPointsInYear=12;
    NbrDaysInMonth=12; % Actually months in in year for this example
    AssetMin=-1;   GetRet=[0.022,0.06,0.1];

case 5
'48 COMPANIES - 3465 days'
NbrComp=48; % 48 all
DataRaw = xlsread('StockData_Final_LN_48_3465');
SizeXY=size(DataRaw); NbrDaysTotInData=SizeXY(1)-20;    % Size of the data file
DataAll=DataRaw(21:end,2:NbrComp+1);                    % y,x: Top Line:Names of Corps, First
line excluded:Dates,
%-----
NbrDaysInPeriod=5;                                     % The number of days in one period/step.
NbrPeriodsInTot=693; % All is 693!                     % The total number of periods for all the time
NbrPeriodsInGather=50;                                 % The number of periods in the Gather time.
%and use this data to create the future portfolios
NbrPointsInYear=250;
NbrDaysInMonth=20; % days per month
AssetMin=-1;
GetRet=[0.002,0.0025,0.003];
%GetRet=[0.001,0.00125,0.0015];

case 6
'28 companies at a time with different STD'
NbrComp=28; % 7*28=196 all
StdInterval=7;
DataRaw = xlsread('StockData_Final_LN_1009_std');
SizeXY=size(DataRaw); NbrDaysTotInData=SizeXY(1)-20;    % Size of the data file
DataAll=DataRaw(21:end,(2+NbrComp*(StdInterval-1)):NbrComp*(StdInterval)+1);
% y,x: Top Line:Names of Corps, First line excluded:Dates,
%-----
NbrDaysInPeriod=5;                                     % The number of days in one period/step.
NbrPeriodsInTot=200; % 200!                             % The total number of periods for all the time
NbrPeriodsInGather=50;                                 % The number of periods in the Gather time.
%and use this data to create the future portfolios
NbrPointsInYear=250;
NbrDaysInMonth=20; % Actually months in in year for this example
AssetMin=-1;
GetRet=[0.002,0.0025,0.003];
%GetRet=[0.001,0.00125,0.0015];

otherwise
    error('Data dont exist');
end
NbrComp=NbrComp;
%*****
NbrPeriodsInRun=NbrPeriodsInTot-NbrPeriodsInGather;    % Number of steps, with the gather
period excluded. This is RUN period
NbrDaysInRun=NbrDaysInPeriod*NbrPeriodsInRun;
NbrDaysInGather=NbrPeriodsInGather*NbrDaysInPeriod;    % The total number of *days* in the
Gather time

NbrMonthsInRun=(NbrDaysInRun/NbrDaysInMonth);
'START-----'

'----- M Begins----'
DataM=zeros(NbrDaysTotInData,NbrComp);

```

```

Mmean=mean(DataAll);Mstd=std(DataAll);

for day=1:NbrDaysTotInData
    DataM(day,:)=(DataAll(day,:)-Mmean)./Mstd;
end
'----- M End ---'

'CSample and ExpReturn begins'
ReturnProc=exp(DataAll)-1;
DecayFactor=0.99;
for Period=1:1:NbrPeriodsInRun;
    CSample(:,,Period)=corrcoef(DataM((((Period-1)*NbrDaysInPeriod+1):((Period-
1)*NbrDaysInPeriod+NbrDaysInGather)),:));
    RetSeries=DataAll((((Period-1)*NbrDaysInPeriod+1):((Period-
1)*NbrDaysInPeriod+NbrDaysInGather)),:);

    ExpReturnGatherLog(Period,:)=mean(RetSeries); %Procentual differences for expected return
    % [ExpExpoReturnGatherLog(Period,:), ExpExpoCovariance(:,,Period)] = ewstats(RetSeries,
DecayFactor);

end
'CSample and ExpReturn ends'
%ExpReturnGatherLog=ExpExpoReturnGatherLog;

%----Exponential correlation matrices and returns could be inserted HERE

DataInRun=DataAll((1+NbrDaysInGather):(NbrDaysInRun+NbrDaysInGather),:); % If not all data
is used. Data all will be a lot shorter than DataAll

'POWER MAPPING Beginning-----'
q_pow=1.5;
CPower=sign(CSample).*abs(CSample).^q_pow; %Much faster than looping over the elements 2s
vs 30s
'POWER MAPPING End-----'

[EigenvectorSample,DSample]=eig(CSample(:,,NbrPeriodsInRun));
EigenvalueSample=diag(DSample);
[temp,DPower]=eig(CPower(:,,NbrPeriodsInRun));          EigenvaluePower=diag(DPower);

'RMT-FILTERING Beginning-----'
CFilter=zeros(NbrComp,NbrComp,NbrPeriodsInRun);
EigenvalueMatrixSample=zeros(NbrComp,NbrPeriodsInRun);% bug nbrsteps are lower here
%
for Period=1:1:NbrPeriodsInRun
    [EigenvectorSample,DSample]=eig(CSample(:,,Period));    EigenvalueSample=diag(DSample);
    EigenvalueMatrixSample(:,Period)=EigenvalueSample;

    %%-----
    Q=NbrDaysInGather/NbrComp;    AlphaMax=1+1/Q+2*sqrt(1/Q);
    FilterNbr=NbrComp;
    while (AlphaMax<=EigenvalueSample(FilterNbr))

CFilter(:,,Period)=CFilter(:,,Period)+EigenvalueSample(FilterNbr)*EigenvectorSample(:,FilterNbr
)*EigenvectorSample(:,FilterNbr)'; %Decide nbr of eigenvalues
        FilterNbr=FilterNbr-1;
    end

% CFilter(:,,Period)=EigenvalueSample(FilterNbr)*EigenvectorSample(:,FilterNbr)*EigenvectorSa

```

```

mple(:,FilterNbr)+EigenvalueSample(FilterNbr-1)*EigenvectorSample(:,FilterNbr-
1)*EigenvectorSample(:,FilterNbr-1); %use two eigenvalues
    CFilter(:,Period)=CFilter(:,Period)-diag(diag(CFilter(:,Period)))+eye(NbrComp); %adds a
diagonal. Shall ALWAYS be there

end
'RMT-FILTERING Ends-----'
%Period=1; CSampleOne=CSample(:,Period); CPowerOne=CPower(:,Period);
CFilterOne=CFilter(:,Period);
% figure(10); pcolor(CSampleOne); shading flat; colormap('gray');

% 'PM-FILTERING Beginning-----'
% q_pow=1.5;
% CFilter=sign(CFilter).*abs(CFilter).^q_pow; %Much faster than looping over the elements 2s vs
30s
% 'PM-Filtering End-----'
% % %CPower=rand(NbrComp,NbrComp,NbrPeriodsInRun);

% % 'IMAGE FILTERING begins-----'
% CImageBefore=uint16((CSample+1)*32768);
% CImageAfter=CImageBefore;
% CImageAfter=imadjust(CImageBefore,[0.05 1],[,]);
% CFilter=double(CImageAfter)/32768-1;
% % 'IMAGE FILTERING ends-----'

'Manual Markowitz begins -----'
OneRow=ones(1,NbrComp);OneColumn=ones(NbrComp,1);

PortNbr=4;
PortRiskEstSample=zeros(PortNbr,NbrPeriodsInRun);PortReturnEstSample=zeros(PortNbr,NbrPeri
odsInRun);PortWtsSample=zeros(PortNbr,NbrComp,NbrPeriodsInRun);
PortRiskEstPower=zeros(PortNbr,NbrPeriodsInRun);PortReturnEstPower=zeros(PortNbr,NbrPeriod
sInRun);PortWtsPower=zeros(PortNbr,NbrComp,NbrPeriodsInRun);
PortRiskEstFilter=zeros(PortNbr,NbrPeriodsInRun);PortReturnEstFilter=zeros(PortNbr,NbrPeriodsI
nRun);PortWtsFilter=zeros(PortNbr,NbrComp,NbrPeriodsInRun);
MvpWtsSample=zeros(PortNbr,NbrComp,NbrPeriodsInRun);MvpWtsPower=zeros(PortNbr,NbrCo
mp,NbrPeriodsInRun);MvpWtsFilter=zeros(PortNbr,NbrComp,NbrPeriodsInRun);

OneRow=ones(1,NbrComp);OneColumn=ones(NbrComp,1);

PVal=1.0; AssetMax=1.0;
ConSet = portcons('Port Value', PVal, NbrComp,'AssetLims',AssetMin,AssetMax, NbrComp);

for Period=1:1:NbrPeriodsInRun % ---- Weights are calculated with Markowitz theory Sample-----
---
    DataGatherForRun=DataAll((((Period-1)*NbrDaysInPeriod+1):((Period-
1)*NbrDaysInPeriod+NbrDaysInGather)),:);
    V=diag(std(DataGatherForRun));
    CovMatrixSample=V*CSample(:,Period)*V; CovMatrixPower=V*CPower(:,Period)*V;
CovMatrixFilter=V*CFilter(:,Period)*V;

['Markowitz optimizing. Period=',num2str(mean(Period))]

    if fast_mvpm==1
        'ShortSelling is permitted'
        CovMatrixSampleInv=inv(CovMatrixSample);
PortWtsSample(PortNbr,:,Period)=(CovMatrixSampleInv*OneColumn/(OneRow*CovMatrixSampl
eInv*OneColumn));

```

```

    CovMatrixPowerInv=inv(CovMatrixPower);
    PortWtsPower(PortNbr,.,Period)=(CovMatrixPowerInv*OneColumn/(OneRow*CovMatrixPowerInv*OneColumn));
    CovMatrixFilerInv=inv(CovMatrixFilter);
    PortWtsFilter(PortNbr,.,Period)=(CovMatrixFilerInv*OneColumn/(OneRow*CovMatrixFilerInv*OneColumn));
    else
        'ShortSelling is NOT permitted'
        [RiskMvp, ReturnMvp, WeightMvp] = portopt(ExpReturnGatherLog(Period,1:NbrComp),
    CovMatrixSample(1:NbrComp,1:NbrComp), 2, [], ConSet);
        PortRiskEstSample(PortNbr,Period)=RiskMvp(1);
        PortReturnEstSample(PortNbr,Period)=ReturnMvp(1);
        PortWtsSample(PortNbr,.,Period)=WeightMvp(1,:);

        [RiskMvp, ReturnMvp, WeightMvp] = portopt(ExpReturnGatherLog(Period,1:NbrComp),
    CovMatrixPower(1:NbrComp,1:NbrComp), 2, [], ConSet);
        PortRiskEstPower(PortNbr,Period)=RiskMvp(1);
        PortReturnEstPower(PortNbr,Period)=ReturnMvp(1);
        PortWtsPower(PortNbr,.,Period)=WeightMvp(1,:);

        [RiskMvp, ReturnMvp, WeightMvp] = portopt(ExpReturnGatherLog(Period,1:NbrComp),
    CovMatrixFilter(1:NbrComp,1:NbrComp), 2, [], ConSet);
        PortRiskEstFilter(PortNbr,Period)=RiskMvp(1);
        PortReturnEstFilter(PortNbr,Period)=ReturnMvp(1);
        PortWtsFilter(PortNbr,.,Period)=WeightMvp(1,:);
    end

    %-- with different expected return -----
    if only_mvp==0 %Othwerwise only MVP will be calculated
        try [PortRiskEstSample(1:PortNbr-1,Period), PortReturnEstSample(1:PortNbr-1,Period),
    PortWtsSample(1:PortNbr-1,.,Period)] = portopt(ExpReturnGatherLog(Period,1:NbrComp),
    CovMatrixSample(1:NbrComp,1:NbrComp), [],GetRet,ConSet);
        catch
            'Sample Fel'
            lasterr % Displays the last error
        end

        try [PortRiskEstPower(1:PortNbr-1,Period), PortReturnEstPower(1:PortNbr-1,Period),
    PortWtsPower(1:PortNbr-1,.,Period)] = portopt(ExpReturnGatherLog(Period,1:NbrComp),
    CovMatrixPower(1:NbrComp,1:NbrComp), [],GetRet,ConSet);
        catch
            'Power Fel'
            lasterr
        end

        try [PortRiskEstFilter(1:PortNbr-1,Period), PortReturnEstFilter(1:PortNbr-1,Period),
    PortWtsFilter(1:PortNbr-1,.,Period)] = portopt(ExpReturnGatherLog(Period,1:NbrComp),
    CovMatrixFilter(1:NbrComp,1:NbrComp), [],GetRet,ConSet);
        catch
            'Filter fel'
            lasterr
        end
    end
    end
    % PLOTTING*****
    % Period=1;
    % portopt(ExpReturnGatherLog(Period,1:NbrComp),
    CovMatrixSample(1:NbrComp,1:NbrComp), 10,[],ConSet); title('Sample Last Period');
    % portopt(ExpReturnGatherLog(Period,1:NbrComp), CovMatrixPower(1:NbrComp,1:NbrComp),
    10,[],ConSet); title('Power Period');
    % portopt(ExpReturnGatherLog(Period,1:NbrComp), CovMatrixFilter(1:NbrComp,1:NbrComp),
    10,[],ConSet); title('Filter Period');

```

```

end

'Eliminating too small weights BEGINS-----'
PortWtsSampleIntact=PortWtsSample; PortWtsPowerIntact=PortWtsPower;
PortWtsFilterIntact=PortWtsFilter;

PortWtsSampleElim=zeros(1,NbrPeriodsInRun);PortWtsPowerElim=zeros(1,NbrPeriodsInRun);Por
tWtsFilterElim=zeros(1,NbrPeriodsInRun);
for x=1:1:PortNbr
    for y=1:1:NbrComp;
        for Period=1:1:NbrPeriodsInRun
            if (abs(PortWtsSample(x,y,Period))<=0.0001)
                PortWtsSample(x,y,Period)=0;
                PortWtsSampleElim(Period)=PortWtsSampleElim(Period)+1;
            end
            if (abs(PortWtsPower(x,y,Period))<=0.0001)
                PortWtsPower(x,y,Period)=0;
                PortWtsPowerElim(Period)=PortWtsPowerElim(Period)+1;
            end
            if (abs(PortWtsFilter(x,y,Period))<=0.0001)
                PortWtsFilter(x,y,Period)=0;
                PortWtsFilterElim(Period)=PortWtsFilterElim(Period)+1;
            end
        end
    end
end

'Eliminating too small weights Ends-----'

PortReturnEstPowerMinusSample=(PortReturnEstPower-PortReturnEstSample);
PortReturnEstFilterMinusSample=(PortReturnEstFilter-PortReturnEstSample);

PortRiskEstPowerMinusSample=(PortRiskEstPower-PortRiskEstSample);
PortRiskEstFilterMinusSample=(PortRiskEstFilter-PortRiskEstSample);

end %do_calc
'Manual Markowitz ends -----'

clear Total*;
'----- Risk And Return Begins'

TotalActualRiskSampleMonth=zeros(NbrMonthsInRun,PortNbr);
TotalActualRiskPowerMonth=zeros(NbrMonthsInRun,PortNbr);
TotalActualRiskFilterMonth=zeros(NbrMonthsInRun,PortNbr);
TotalReturnActualSampleMonth=ones(NbrMonthsInRun,PortNbr);
TotalReturnActualPowerMonth=ones(NbrMonthsInRun,PortNbr);
TotalReturnActualFilterMonth=ones(NbrMonthsInRun,PortNbr);
TotalReturnMeanSampleMonth=zeros(NbrMonthsInRun,PortNbr);
TotalReturnMeanPowerMonth=zeros(NbrMonthsInRun,PortNbr);
TotalReturnMeanFilterMonth=zeros(NbrMonthsInRun,PortNbr);

NbrPeriodsInGather %Writes gather time
for SelPort=1:PortNbr

    if SelPort==4
        ['***** Portfolio ',num2str(SelPort),' (MVP) *****']
    else
        ['***** Portfolio ',num2str(SelPort),' (Exp Return= ',num2str(GetRet(SelPort)),')
*****']
    end
end

```

```

end

SumRiskSample=zeros(NbrDaysInRun,1); SumRiskPower=zeros(NbrDaysInRun,1);
SumRiskFilter=zeros(NbrDaysInRun,1);

for Period=1:1:NbrPeriodsInRun %3, nbr years in run
    for Day=(1+(Period-1)*NbrDaysInPeriod):1:(Period*NbrDaysInPeriod)
        SumRiskSample(Day,:)=sum(DataInRun(Day,:).*PortWtsSample(SelPort,:,Period));
        SumRiskPower(Day,:)=sum(DataInRun(Day,:).*PortWtsPower(SelPort,:,Period));
        SumRiskFilter(Day,:)=sum(DataInRun(Day,:).*PortWtsFilter(SelPort,:,Period));
    end
end

%sqrt(NbrPointsInYear)
TotalActualRiskSample=100*std(SumRiskSample); %% 100 for percentage YEARLY RISK
TotalActualRiskPower=100*std(SumRiskPower); %% 100 for percentage
TotalActualRiskFilter=100*std(SumRiskFilter); %% 100 for percentage
['TotalActualRiskSample= ',num2str(TotalActualRiskSample),'% (daily) ,Yearly=
',num2str(sqrt(NbrPointsInYear)*TotalActualRiskSample),'%']
['TotalActualRiskPower= ',num2str(TotalActualRiskPower),'% (daily) ,Yearly=
',num2str(sqrt(NbrPointsInYear)*TotalActualRiskPower),'% , (Vs Sample=
',num2str(100*(TotalActualRiskPower-TotalActualRiskSample)/TotalActualRiskSample),'%')]
['TotalActualRiskFilter= ',num2str(TotalActualRiskFilter),'% (daily) ,Yearly=
',num2str(sqrt(NbrPointsInYear)*TotalActualRiskFilter),'% , (Vs Sample=
',num2str(100*(TotalActualRiskFilter-TotalActualRiskSample)/TotalActualRiskSample),'%')]

TotalReturnMeanSample=100*mean(SumRiskSample); ['TotalReturnMeanSample=
',num2str(TotalReturnMeanSample),'% (daily return)']
TotalReturnMeanPower=100*mean(SumRiskPower); ['TotalReturnMeanPower=
',num2str(TotalReturnMeanPower),'% (daily return)']
TotalReturnMeanFilter=100*mean(SumRiskFilter); ['TotalReturnMeanFilter=
',num2str(TotalReturnMeanFilter),'% (daily return)']

TotalReturnMeanDivRiskSample=TotalReturnMeanSample/TotalActualRiskSample;
TotalReturnMeanDivRiskPower=TotalReturnMeanPower/TotalActualRiskPower;
TotalReturnMeanDivRiskFilter=TotalReturnMeanFilter/TotalActualRiskFilter;

['TotalReturnMeanDivRiskSample= ',num2str(TotalReturnMeanDivRiskSample),' (daily)']
['TotalReturnMeanDivRiskPower= ',num2str(TotalReturnMeanDivRiskPower),' (daily) (Vs
Sample= ',num2str(100*(TotalReturnMeanDivRiskPower-
TotalReturnMeanDivRiskSample)/TotalReturnMeanDivRiskSample),'%')]
['TotalReturnMeanDivRiskFilter= ',num2str(TotalReturnMeanDivRiskFilter),' (daily) (Vs
Sample= ',num2str(100*(TotalReturnMeanDivRiskFilter-
TotalReturnMeanDivRiskSample)/TotalReturnMeanDivRiskSample),'%')]

SumOneSample=SumRiskSample+1; SumOnePower=SumRiskPower+1;
SumOneFilter=SumRiskFilter+1;

for month=1:1:NbrMonthsInRun %May also mean number year in Run.
    day_start=(month-1)*NbrDaysInMonth+1;
    day_end=NbrDaysInMonth+(month-1)*NbrDaysInMonth;

TotalActualRiskSampleMonth(month,SelPort)=100*std(SumRiskSample(day_start:day_end,:));
%Daily
TotalActualRiskPowerMonth(month,SelPort)=100*std(SumRiskPower(day_start:day_end,:));
TotalActualRiskFilterMonth(month,SelPort)=100*std(SumRiskFilter(day_start:day_end,:));

```

```

TotalReturnMeanSampleMonth(month,SelPort)=100*mean(SumRiskSample(day_start:day_end,:));
%Daily!

TotalReturnMeanPowerMonth(month,SelPort)=100*mean(SumRiskPower(day_start:day_end,:));

TotalReturnMeanFilterMonth(month,SelPort)=100*mean(SumRiskFilter(day_start:day_end,:));

    for Period=day_start:day_end        % Monthly actual returns.

TotalReturnActualSampleMonth(month,SelPort)=TotalReturnActualSampleMonth(month,SelPort)*
SumOneSample(Period); %For a MONTH

TotalReturnActualPowerMonth(month,SelPort)=TotalReturnActualPowerMonth(month,SelPort)*Su
mOnePower(Period);

TotalReturnActualFilterMonth(month,SelPort)=TotalReturnActualFilterMonth(month,SelPort)*Sum
OneFilter(Period);
    end
end

    TotalReturnActualSample=1;        TotalReturnActualPower=1;    TotalReturnActualFilter=1;

    for Period=1:1:NbrDaysInRun
        TotalReturnActualSample=TotalReturnActualSample*SumOneSample(Period); % WHOLE
PERIOD
        TotalReturnActualPower=TotalReturnActualPower*SumOnePower(Period);
        TotalReturnActualFilter=TotalReturnActualFilter*SumOneFilter(Period);
    end

    TotalReturnActualSample=(TotalReturnActualSample); ['TotalReturnActualSample=
',num2str(TotalReturnActualSample),' times start value (whole period)']
    TotalReturnActualPower=(TotalReturnActualPower); ['TotalReturnActualPower=
',num2str(TotalReturnActualPower),' times start value (whole period)']
    TotalReturnActualFilter=(TotalReturnActualFilter); ['TotalReturnActualFilter=
',num2str(TotalReturnActualFilter),' times start value (whole period)']

TotalReturnActualDivRiskSample=TotalReturnActualSample/(sqrt(NbrDaysInRun)*TotalActualRis
kSample); % Whole Period!

TotalReturnActualDivRiskPower=TotalReturnActualPower/(sqrt(NbrDaysInRun)*TotalActualRisk
Power); %Nbrdaysinrun is STEPS in Run!

TotalReturnActualDivRiskFilter=TotalReturnActualFilter/(sqrt(NbrDaysInRun)*TotalActualRiskFil
ter);

    ['TotalReturnActualDivRiskSample= ',num2str(TotalReturnActualDivRiskSample),' (whole
period)']
    ['TotalReturnActualDivRiskPower= ',num2str(TotalReturnActualDivRiskPower),' (whole period)
(Vs Sample= ',num2str(100*(TotalReturnActualDivRiskPower-
TotalReturnActualDivRiskSample)/TotalReturnActualDivRiskSample),'% )']
    ['TotalReturnActualDivRiskFilter= ',num2str(TotalReturnActualDivRiskFilter),' (whole period)
(Vs Sample= ',num2str(100*(TotalReturnActualDivRiskFilter-
TotalReturnActualDivRiskSample)/TotalReturnActualDivRiskSample),'% )']

end

CSampleMeanAll=zeros(1,NbrPeriodsInRun); CSampleMeanAll(1,:)=mean(mean(CSample));
CPowerMeanAll=zeros(1,NbrPeriodsInRun); CPowerMeanAll(1,:)=mean(mean(CPower));

```

```

CFilterMeanAll=zeros(1,NbrPeriodsInRun); CFilterMeanAll(1,:)=mean(mean(CFilter));
%*****
lw=2;
if do_plot==1 %----- PLOTTING -----

sum(PortWtsSample(2,:,:));%% Plotting of holes
figure(221);
hold on
plot(PortWtsSampleElim/NbrComp,'ko','MarkerSize',4,'LineWidth',lw);
plot(PortWtsPowerElim/NbrComp,'r+'); plot(PortWtsFilterElim/NbrComp,'b. ');
title('Eliminated Weights(%)' ); xlabel('Year (5-day intervals)'); ylabel('Percentage Eliminated
Weights(%)' );
set(gca,'XTick',[25 75 125]); set(gca,'XTickLabel',{'2001','2002','2003'} ) ;
legend('Sample','Power','Filter',3);
hold off

%-----MVP -----
if fast_mvp==0;
    %MVP est risk/return not possible with our formulals, must use Matlabs,
    %and not the fast mpv.
    figure(111); hold on
    plot(PortReturnEstPowerMinusSample(4,:)*100,'r','LineWidth',lw); title('Difference in Est.
Return (MVP) for (Sample - Power) and (Sample-Power)'); ylabel('Difference'); xlabel('Year (5-day
intervals)');
    plot(PortReturnEstFilterMinusSample(4,:)*100,'b--','LineWidth',lw);
    set(gca,'XTick',[25 75 125]); set(gca,'XTickLabel',{'2001','2002','2003'} ) ;
    legend('Power','Filter',2);
    hold off

    figure(112); hold on
    plot(PortRiskEstPowerMinusSample(4,:)*100,'r','LineWidth',lw); title('Difference in Est. Risk
(MVP) for (Sample - Power) and (Sample-Filter)'); ylabel('Difference'); xlabel('Year (5-day
intervals)');
    plot(PortRiskEstFilterMinusSample(4,:)*100,'b--','LineWidth',lw);
    set(gca,'XTick',[25 75 125]); set(gca,'XTickLabel',{'2001','2002','2003'} ) ;
    legend('Power','Filter',2);
    hold off
end

figure(1001); 'For Portfolio1'
hold on
plot(PortRiskEstSample(1,:),'k-', 'LineWidth',lw)
plot(PortRiskEstPower(1,:),'r', 'LineWidth',lw)
plot(PortRiskEstFilter(1,:),'b--', 'LineWidth',lw)
title('Estimated risk for different methods'); ylabel('Risk '); xlabel('Year');
set(gca,'XTick',[25 75 125]); set(gca,'XTickLabel',{'2001','2002','2003'});
hold off

figure(121); hold on
plot(PortRiskEstPowerMinusSample(1,:),'-', 'LineWidth',lw);
plot(PortRiskEstPowerMinusSample(2,:),'--', 'LineWidth',lw);
plot(PortRiskEstPowerMinusSample(3,:),'.', 'LineWidth',lw);
title('Difference in Estimated Risk for (Power - Sample)'); ylabel('Risk Diff '); xlabel('Year (monthly
intervals)');
set(gca,'XTick',[25 75 125]); set(gca,'XTickLabel',{'2001','2002','2003'});
legend(['Port1 (Exp Return= ',num2str(GetRet(1)),')'],['Port2 (Exp Return= ',num2str(GetRet(2)),
')'],['Port3(Exp Return= ',num2str(GetRet(3)),')'],3);
hold off

figure(122); hold on

```



```

plot(PortRiskEstFilterMinusSample(1,:),'-', 'LineWidth',lw);
plot(PortRiskEstFilterMinusSample(2,:),'--', 'LineWidth',lw);
plot(PortRiskEstFilterMinusSample(3,:),'.', 'LineWidth',lw);
title('Difference in Estimated Risk for (Filter - Sample)'); ylabel('Risk Diff '); xlabel('Year (monthly intervals)');
set(gca,'XTick',[25 75 125]); set(gca,'XTickLabel',{'2001','2002','2003'});
legend(['Port1 (Exp Return= ',num2str(GetRet(1)),')'],['Port2 (Exp Return= ',num2str(GetRet(2)),')'],['Port3 (Exp Return= ',num2str(GetRet(3)),')'],3);
hold off

```

```

figure(131); hold on % Plotting Weights for the 197 companies
plot(PortWtsSample(1,:,1),'ko','LineWidth',lw); plot(PortWtsPower(1,:,1),'r+', 'LineWidth',lw);
plot(PortWtsFilter(1,:,1),'.', 'LineWidth',lw);
title('Weights for the first portfolio with diff. methods'); ylabel('Weight '); xlabel('Year (monthly intervals)');
xlabel('Company');
legend('Sample','Power','Filter',3);
hold off

```

```

for SelPort=1:PortNbr
    figure(850+SelPort); hold on
    plot(TotalActualRiskSampleMonth(:,SelPort),'ko-', 'LineWidth',lw);
    plot(TotalActualRiskPowerMonth(:,SelPort),'r+', 'LineWidth',lw)
    plot(TotalActualRiskFilterMonth(:,SelPort),'b.-', 'LineWidth',lw)
    ylabel('Daily Risk (%)'); xlabel('Year (monthly intervals)');
    if SelPort==4
        title(['Daily Risk, portfolio',num2str(SelPort),' (MVP)']);
    else;
        title(['Daily Risk, portfolio',num2str(SelPort),' (Exp Return= ',num2str(GetRet(SelPort)),')']); end;
    set(gca,'XTick',[7 20 31]); set(gca,'XTickLabel',{'2001','2002','2003'});
    legend('Sample','Power','Filter',3); hold off;

```

```

% axis([0 35 0 35]);
figure(860+SelPort); hold on
plot(TotalReturnMeanSampleMonth(:,SelPort),'ko-', 'LineWidth',lw);
plot(TotalReturnMeanPowerMonth(:,SelPort),'r+', 'LineWidth',lw)
plot(TotalReturnMeanFilterMonth(:,SelPort),'b.-', 'LineWidth',lw)
ylabel('Return (%)'); xlabel('Year (monthly intervals)');
if SelPort==4
    title(['Mean daily Return per month, portfolio',num2str(SelPort),' (MVP)']);
else;
    title(['Mean daily Return per month, portfolio',num2str(SelPort),' (Exp Return= ',num2str(GetRet(SelPort)),')']); end;
set(gca,'XTick',[7 20 31]); set(gca,'XTickLabel',{'2001','2002','2003'});
legend('Sample','Power','Filter',3); hold off;

```

```

% axis([0 35 0 35]);
figure(880+SelPort); hold on
plot(TotalReturnActualSampleMonth(:,SelPort),'ko-', 'LineWidth',lw);
plot(TotalReturnActualPowerMonth(:,SelPort),'r+', 'LineWidth',lw)
plot(TotalReturnActualFilterMonth(:,SelPort),'b.-', 'LineWidth',lw)
ylabel('Montly Return (compared to base value 1.0)'); xlabel('Year (monthly intervals)');
if SelPort==4
    title(['Actual monthly Return, portfolio',num2str(SelPort),' (MVP)']);
else;
    title(['Actual monthly Return, portfolio',num2str(SelPort),' (Exp Return= ',num2str(GetRet(SelPort)),')']); end;
set(gca,'XTick',[7 20 31]); set(gca,'XTickLabel',{'2001','2002','2003'});
legend('Sample','Power','Filter',3); hold off ;

```

```

figure(890+SelPort); hold on

```

```
plot(TotalReturnActualSampleMonth(:,SelPort)./(sqrt(NbrDaysInMonth)*TotalActualRiskSampleMonth(:,SelPort)), 'ko-', 'LineWidth', lw)
```

```
plot(TotalReturnActualPowerMonth(:,SelPort)./(sqrt(NbrDaysInMonth)*TotalActualRiskPowerMonth(:,SelPort)), 'r+:', 'LineWidth', lw)
```

```
plot(TotalReturnActualFilterMonth(:,SelPort)./(sqrt(NbrDaysInMonth)*TotalActualRiskFilterMonth(:,SelPort)), 'b--', 'LineWidth', lw)
```

```
    ylabel('Monthly ActualReturn/ActualRisk'); xlabel('Year (monthly intervals)');
    if SelPort==4
        title(['Monthly ActualReturn/ActualRisk', num2str(SelPort), ' (MVP)']);
    else
        title(['Monthly ActualReturn/ActualRisk ', num2str(SelPort), ' (Exp Return=
', num2str(GetRet(SelPort)), ' )']); end;
    set(gca, 'XTick', [7 20 31]); set(gca, 'XTickLabel', {'2001', '2002', '2003'});
    legend('Sample', 'Power', 'Filter', 3); hold off;
end
```

```
% Correlation matrices
```

```
figure(881); pcolor(CSample(:, :, 1)); shading flat; colormap('gray'); title('CSample for first year');
figure(882); pcolor(CPower(:, :, 1)); shading flat; colormap('gray'); title('CPower for first year');
figure(883); pcolor(CFilter(:, :, 1)); shading flat; colormap('gray'); title('CFilter for first year');
```

```
figure(991);
hist(EigenvalueSample, 100); title('Eigenvalue spectrum for Sample (last portfolio)'); % Eigenvalue
spectra
ylabel('Density'); xlabel('Eigenvalue');
figure(992);
hist(EigenvaluePower, 100); title('Eigenvalue spectrum for Power (last portfolio)');
ylabel('Density'); xlabel('Eigenvalue');
```

```
% ----- MEAN value of C -----
```

```
figure(77777);
hold on
plot(CSampleMeanAll, 'k-', 'MarkerSize', 4, 'LineWidth', lw); plot(CPowerMeanAll, 'r-', 'LineWidth', lw);
plot(CFilterMeanAll, 'b--', 'LineWidth', lw);
set(gca, 'XTick', [25 75 125]); set(gca, 'XTickLabel', {'2001', '2002', '2003'});
legend('Sample', 'Power', 'Filter', 3); title('Mean value of C for Sample, Power and Filter');
ylabel('Mean value'); xlabel('Year (5-day intervals)');
hold off
```

```
%- Plotting Several C for Each Year
% Will be compared to the mean values
```

```
empty_tick=[10000]; empty_tick_label={'_'};
nbr=5; figure(7777);
```

```
if stockdata_file==1
    i=1; PeriodLen=30; x_len=5; y_len=3;
    for px=1:x_len
        subplot(y_len, x_len, i);
        pcolor(CSample(:, :, i*PeriodLen)); shading flat; colormap('gray');
        set(gca, 'XTick', empty_tick); set(gca, 'XTickLabel', empty_tick_label);
        set(gca, 'YTick', empty_tick); set(gca, 'YTickLabel', empty_tick_label);

        if i==1; ylabel('Sample'); ; end;
        if i==3; title('CSample, CPower and CFilter'); end;

        subplot(y_len, x_len, i+x_len);
        pcolor(CPower(:, :, i*PeriodLen)); shading flat; colormap('gray');
```

```

set(gca,'XTick',empty_tick); set(gca,'XTickLabel',empty_tick_label);
set(gca,'YTick',empty_tick); set(gca,'YTickLabel',empty_tick_label);
if i==1; ylabel('Power'); end;

subplot(y_len,x_len,i+x_len*2);
pcolor(CFilter(:,i*PeriodLen)); shading flat; colormap('gray');
set(gca,'XTick',empty_tick); set(gca,'XTickLabel',empty_tick_label);
set(gca,'YTick',empty_tick); set(gca,'YTickLabel',empty_tick_label);
if i==1; ylabel('Filter'); end;
xlabel(['Period ', num2str(i*PeriodLen)]);

i=i+1;
end
end

if stockdata_file==3
i=1; PeriodLen=4; x_len=5; y_len=3;
for px=1:x_len
subplot(y_len,x_len,i);
pcolor(CSample(:,i*PeriodLen)); shading flat; colormap('gray');
set(gca,'XTick',empty_tick); set(gca,'XTickLabel',empty_tick_label);
set(gca,'YTick',empty_tick); set(gca,'YTickLabel',empty_tick_label);

if i==1; ylabel('Sample'); ; end;
if i==3; title('CSample, CPower and CFilter'); end;

subplot(y_len,x_len,i+x_len);
pcolor(CPower(:,i*PeriodLen)); shading flat; colormap('gray');
set(gca,'XTick',empty_tick); set(gca,'XTickLabel',empty_tick_label);
set(gca,'YTick',empty_tick); set(gca,'YTickLabel',empty_tick_label);
if i==1; ylabel('Power'); end;

subplot(y_len,x_len,i+x_len*2);
pcolor(CFilter(:,i*PeriodLen)); shading flat; colormap('gray');
set(gca,'XTick',empty_tick); set(gca,'XTickLabel',empty_tick_label);
set(gca,'YTick',empty_tick); set(gca,'YTickLabel',empty_tick_label);
if i==1; ylabel('Filter'); end;
xlabel(['Period ', num2str(i*PeriodLen)]);

i=i+1;
end
end

% nbr=5; figure(7777);
% i=1; PeriodLen=10; x_len=5; y_len=3;
% for px=1:x_len
% for py=1:y_len
% subplot(y_len,x_len,i);
% pcolor(CSample(:,i*PeriodLen)); shading flat; colormap('gray');
% set(gca,'XTick',empty_tick); set(gca,'XTickLabel',empty_tick_label);
% xlabel([num2str(2000+i)]);
% i=i+1;
% end
% end
% end

end
%-----PLOTTING ENDS -----

```

A.6.3 MeanColumn.m

```

OpenGL neverselect %Will use software OpenGL if choosen once. Will be much faster
clear all

do_plot=0;

NbrDaysInPeriod=1; %The number of days in one period/step.
NbrDaysInMonth=20;
NbrPeriodsInTot=3465; %3465 %The total number of periods for all the time
NbrPeriodsInGather=750; %The number of periods in the Gather time. %and
use this data to create the future portfolios
NbrPeriodsInRun=NbrPeriodsInTot-NbrPeriodsInGather; %Number of steps, with the gather
period excluded. This is RUN period
NbrDaysInRun=NbrDaysInPeriod*NbrPeriodsInRun

NbrMonthsInRun=double(int8(NbrDaysInRun/NbrDaysInMonth));
NbrDaysInGather=NbrPeriodsInGather*NbrDaysInPeriod; %The total number of *days* in the
Gather time

'START-----'

DataRow = xlsread('StockData_Final_LN_48_3465');
SizeXY=size(DataRow); NbrComp=SizeXY(2)-1; % Size of the data file
DataAll=DataRow(21:end,2:NbrComp+1); % y,x: Top Line:Names of Corps, First
line excluded:Dates,
ReturnProc=exp(DataAll)-1;
ReturnProcSmall=ReturnProc(1:50,1:3);
'----- M Begins---'
DataM=DataAll;
Mmean=mean(DataAll); Mstd=std(DataAll);
for day=1:NbrPeriodsInTot
    DataM(day,:)=(DataAll(day,:)-Mmean)./Mstd;
end
'----- M End ---'

'CSample Beginning-----'
CSample=zeros(NbrComp,NbrComp,NbrPeriodsInRun); %All the eval data for the different
periods in run period
ReturnProcMean=zeros(NbrPeriodsInRun,NbrComp);

%%NbrPeriodsInRun
for Period=1:1:NbrPeriodsInRun;
    CSample(:,Period)=corrcoef(DataM((((Period-1)*NbrDaysInPeriod+1):((Period-
1)*NbrDaysInPeriod+NbrDaysInGather)),:));
    ReturnProcMean(Period,:)=mean(ReturnProc((((Period-1)*NbrDaysInPeriod+1):((Period-
1)*NbrDaysInPeriod+NbrDaysInGather)),:));
    Period
end
CSampleSmall=CSample(:,1:4);
ReturnProcMeanSmall=ReturnProcMean(1:20,1:3);
'CSample End-----'

'POWER MAPPING Beginning-----'
q_pow=1.5;
CPower=sign(CSample).*abs(CSample).^q_pow; %Much faster than looping over the elements 2s
vs 30s
CPowerSmall=CPower(:,1:4);
'POWER MAPPING End-----'

```

```

'Creating different temp correlation matrices, mean, abs, neg, pos'
NbrTest=4;
tempSample=zeros(1,NbrComp,NbrPeriodsInRun,NbrTest);
tempPower=zeros(1,NbrComp,NbrPeriodsInRun,NbrTest);
CSampleMean=zeros(NbrPeriodsInRun,NbrComp,NbrTest);
CPowerMean=zeros(NbrPeriodsInRun,NbrComp,NbrTest);

tempSample(:,:,1)=mean(CSample);                               tempPower(:,:,1)=mean(CPower);
tempSample(:,:,2)=mean(abs(CSample));
tempPower(:,:,2)=mean(abs(CPower));
tempSample(:,:,3)=mean(double(uint32(CSample*pow2(32)))/pow2(32));
tempPower(:,:,3)=mean(double(uint32(CPower*pow2(32)))/pow2(32));
tempSample(:,:,4)=mean(-double(uint32(-CSample*pow2(32)))/pow2(32));
tempPower(:,:,4)=mean(-double(uint32(-CPower*pow2(32)))/pow2(32));

'Mean Values Beginning-----'
CSampleMean=zeros(NbrPeriodsInRun,NbrComp,NbrTest);
CPowerMean=zeros(NbrPeriodsInRun,NbrComp,NbrTest);

for itest=1:NbrTest
    for day=1:1:NbrPeriodsInRun
        CSampleMean(day,:,itest)=tempSample(1,:,day,itest)';
        CPowerMean(day,:,itest)=tempPower(1,:,day,itest)';
        CSampleMean(day,:,itest)=tempSample(1,:,day,itest)';
        CPowerMean(day,:,itest)=tempPower(1,:,day,itest)';
        CSampleMean(day,:,itest)=tempSample(1,:,day,itest)';
        CPowerMean(day,:,itest)=tempPower(1,:,day,itest)';
        CSampleMean(day,:,itest)=tempSample(1,:,day,itest)';
        CPowerMean(day,:,itest)=tempPower(1,:,day,itest)';
    end
end

CMeanDiff=(CSampleMean-CPowerMean); %Difference in Mean for Sample and Power
CMeanDiffProc=CMeanDiff./abs(CSampleMean);
CMeanDiffSmall=CMeanDiff(1:30,1:4,:); CMeanDiffProcSmall=CMeanDiffProc(1:30,1:4,:);

        CSampleMeanSmall=CSampleMean(1:10,1:4,:);
        CPowerMeanSmall=CPowerMean(1:10,1:4,:);
'Mean Values Ends-----'

%-- Correlations between the mean value of correlation matrix columns and
%-- the procentual returns

CorrCSampleMeanToReturn=zeros(NbrMonthsInRun,NbrComp,NbrTest);
CorrCPowerMeanToReturn=zeros(NbrMonthsInRun,NbrComp,NbrTest);
% '---CorrCSampleMeanToReturn'
% for itest=1:1
%     itest
%     for Comp=1:1:NbrComp
%         Comp
%         for Month=1:1:NbrMonthsInRun           %Note, some months don't have 20, very rare
%             though.
%                 day=Month*(NbrDaysInMonth-1)+1;
%                 temp1=corrcoef(CSampleMean(day:(day+NbrDaysInMonth-1),Comp,NbrTest),ReturnProcMean(day:(day+NbrDaysInMonth-1),Comp));
%                 temp2=corrcoef(CPowerMean(day:(day+NbrDaysInMonth-1),Comp,NbrTest),ReturnProcMean(day:(day+NbrDaysInMonth-1),Comp));
%                 CorrCSampleMeanToReturn(Month,Comp,NbrTest)=temp1(1,2);
%                 CorrCPowerMeanToReturn(Month,Comp,NbrTest)=temp2(1,2);
%         end

```

```

% end
% end

% if do_plot==1 %-----
% figure(661)
% plot(CorrCSampleMeanToReturn(1:end,12));
% title(' for for comp 12 (Volvo)');
%
% figure(662)
% plot(CorrCPowerMeanToReturn(1:end,12));
% title(' for for comp 12 (Volvo)');
%
% figure(992)
% nbr=8
% for comp=1:1:(nbr)
% subplot(nbr,1,comp)
% plot(CorrCSampleMeanToReturn(1:end,comp));
% end

%-----

year_tick=[85 335 585 836 1088 1341 1592 1843 2092 2342 2594 2845 3095 3345]-
NbrPeriodsInGather;
year_tick_label={'90','91','92','93','94','95','96','97','98','99','00','01','02','03'};
empty_tick=[10000]; empty_tick_label={'_'};

%- Plotting One C for 1990-----
figure(7770)
pcolor(CSample(:,84)); shading flat; colormap('gray');
set(gca,'XTick',empty_tick); set(gca,'XTickLabel',empty_tick_label);
xlabel(['Year ',num2str(1990)]); ylabel('Companies');

%- Plotting Several C for Each Year
% Will be compared to the mean values
nbr=NbrComp-1
figure(7777)
i=1; PeriodLen=250; x_len=5;y_len=2
for px=1:x_len
    for py=1:y_len
        subplot(y_len,x_len,i);
        pcolor(CSample(:,1:nbr+1,i*PeriodLen+84)); shading flat; colormap('gray');
        set(gca,'XTick',empty_tick); set(gca,'XTickLabel',empty_tick_label);
        xlabel([num2str(1990+i)]);
        i=i+1;
    end
end

itest=1;
%----- Plotting CSampleColumnMean for several companies -----
figure(8888);
nbr=5; step=0;
for comp=(step+1):(nbr+step)
    subplot(nbr,1,comp-step); plot(CSampleMean(:,comp,itest));
    set(gca,'XTick',year_tick); set(gca,'XTickLabel',year_tick_label);
    if comp==1; title('CSampleColumnMean'); end
    ylabel(['Comp=',num2str(comp)]);
    axis([0 NbrPeriodsInRun 0 0.4])
end

% itest=4;

```

```

% %----- Plotting CSampleColumnMean NEGATIVE for several companies -----
% figure(8888);
% nbr=5; step=0;
% for comp=(step+1):(nbr+step)
%   subplot(nbr,1,comp-step);      plot(CSampleMean(:,comp,itest));
%   set(gca,'XTick',year_tick);    set(gca,'XTickLabel',year_tick_label);
%   if comp==1; title('CSampleColumnMean');    end
%   ylabel(['Comp=',num2str(comp)]);
%   axis([0 NbrPeriodsInRun -0.1 0.3])
% end

comp=3; nbr=4;
figure(9999);
for i=1:2:nbr
    subplot(nbr,1,i);      plot(ReturnProcMean(:,comp));
    if i==1; title('Mean Return');    end
    set(gca,'XTick',empty_tick);    set(gca,'XTickLabel',empty_tick_label);
    ylabel(['Comp=',num2str(comp)])

    subplot(nbr,1,i+1);    plot(CSampleMean(:,comp,itest));
    if i==1; title('CSampleColumnMean');    end
    if i==nbr-1;    set(gca,'XTick',year_tick);    set(gca,'XTickLabel',year_tick_label);
    else
        set(gca,'XTick',empty_tick);    set(gca,'XTickLabel',empty_tick_label);
    end;

    comp=comp+1;
end

subplot(3,1,3)
plot(CorrCSampleMeanToReturn(:,1,itest));
ylabel('Correlation');
set(gca,'XTick',[0 136])
set(gca,'XTickLabel',{'1990','2003'})
title('Correlation Mean Return Procentual Diff/CSampleColumMean');

%% Att plotta börskursen överst, sedan medelvärdet på kolumnerna, göra det
%% för flera företag och sedan dra pilar ner som visar var det händer
%% saker. möjligen också med korrelationen på samma

% namnen ska läggas in i vektor
figure(661)
nbr=4
for comp=1:1:(nbr)
    subplot(nbr,1,comp)
    plot(CSampleMean(:,comp));
    if (comp==1)
        title('CSampleColumnMean');
    end

    if comp==3
        ylabel('CSampleColumnMean');
    end
    if (comp==nbr)
        set(gca,'XTick',[85 335 585 836 1088 1341 1592 1843 2092 2342 2594 2845 3095 3345])
        set(gca,'XTickLabel',{'1990','91','92','93','94','95','96','97','98','99','00','01','02','03'})
        xlabel('Year');
    else
end

```

```

        set(gca,'XTick',[10000])
        set(gca,'XTickLabel',{' '})
    end
    % ylabel('Col. Mean');
end

'Mean Values End-----'

ccc1=CSample(:,,1);
% ccc2=CPower(:,,1);
% ccc3=ccc2-ccc1;
%
figure(2); pcolor(ccc1); shading flat; colormap('gray');
% figure(3); pcolor(ccc2); shading flat; colormap('gray');
% figure(4); pcolor(ccc3); shading flat; colormap('gray');

%'Creating Movie-----'
% for day = 1:1000
%   pcolor(CSample(:,,day));
%   shading flat;
%   colormap('gray');
%   %M(day) = getframe;
%   xlabel((1989*250+(144+day))/250)
%   M(day) = getframe(gcf);
% end
%
% title(['q=',num2str(q_pow(ii-1))]);
%
end % -----Plotting ENDS

```


A.6.4 Data.m

```
clear all
hole_len=0; hole_start=0; hole_end=0; hole_std=0;
std_len=50; %the evaluation length of std backwards
%----
%Data=rand(30,4);
DataRow = xlsread('StockData_S_OO_With_Holes');
SizeXY=size(DataRow);SizeY=SizeXY(1,1);
NbrComp=SizeXY(1,2)-1;
hole_nbr=zeros(1,NbrComp);
Data=DataRow(1:SizeY,2:NbrComp+1); SizeY=SizeY-10;      %The first column have the dates
in number format in DataRow

                                % The size should be correct
StartStd=zeros(1,NbrComp); StdData=zeros(std_len,NbrComp);
%%NbrComp=2;
CompSizeY=zeros(1,NbrComp);
%
=====
=====
%if there are holes in the beginning, it uses the first std_len (say 30)
%numbers to compute the std, until 30 numbers exists
for Comp=1:1:NbrComp
    iy=1;  nbr_nbrs=0;

    while ((iy<std_len*5) && (nbr_nbrs<std_len))

        if (isnan(Data(iy,Comp))~=1)
            nbr_nbrs=nbr_nbrs+1;
            StdData(nbr_nbrs,Comp)=Data(iy,Comp);
        end
        StartStd(Comp)=std(StdData(:,Comp));
        iy=iy+1;
    end
end

% If the first number is lacking the first existing will be copied to it.
for Comp=1:1:NbrComp
    Data(1,Comp)=StdData(1,Comp);
end

% Calculates the lengths of the series of every company
for Comp=1:1:NbrComp
    CompSizeY(Comp)=SizeY;
    iy=SizeY;
    finished=0;
    while ((iy>1) && (finished==0))

        if (isnan(Data(iy,Comp))==1)
            CompSizeY(Comp)=CompSizeY(Comp)-1;
        else
            finished=1;
        end
        iy=iy-1;
    end
end
end
```

```

%
=====
% Searches for holes and fills them with random numbers with std of close
% numbers.
%***Should stop if the holes are too long with a company if hole is too
% long. this means that it doesn't matter if data is out for a company, it
% will stop.
DataFinal=Data;
for Comp=1:1:NbrComp
    for (iy=1:1:CompSizeY(Comp))
        if (isnan(DataFinal(iy,Comp))==1)
            hole_start=iy;
            while (isnan(DataFinal(iy,Comp))&&(iy<(CompSizeY(Comp)))) % Figure out the size of
the hole
                hole_nbr(Comp)=hole_nbr(Comp)+1;
                iy=iy+1;

            end
            hole_end=iy-1;
            hole_len=hole_end-hole_start+1;

            % -----fill in the gaps in the hole
            if (iy>=(std_len*2))
                hole_std=std(DataFinal(hole_start-std_len-1:hole_start-1,Comp));
            else
                hole_std=StartStd(Comp);
                % 'Using Start Std'
            end
            hole_direct=DataFinal(hole_end+1,Comp)-DataFinal(hole_start-1,Comp);

            for hi=1:1:hole_len
                DataFinal(hole_start-1+hi,Comp)=DataFinal(hole_start-
1,Comp)+hole_direct*hi/(hole_len+2)-hole_std/2+hole_std*rand;
                %%DataBackup(hole_start-1+hi,Comp)=Data(hole_start-1+hi,Comp);
            end

        end
    end
end

DataFlipped=flipud(DataFinal);
% csvwrite('Data_filled.dat',DataFinal);

hole_nbr
hole_nbr_mean=mean(hole_nbr)
hole_nbr_perc=hole_nbr./CompSizeY;
hole_nbr_perc_mean=mean(hole_nbr_perc);

hole_nbr_perc_sort=sort(hole_nbr_perc);

hole_nbr_sort=sort(hole_nbr);

%----- Plotting of data and examples -----

figure(1)
hist(hole_nbr_perc,50)
title('Number of holes in original stock data');
xlabel('Number of holes');
ylabel('Number of occurrences');

```

```

figure(2)
hist(hole_nbr,50)
title('Percentages of holes in original stock data');
xlabel('Number of holes');
ylabel('Number of occurrences');

Statistics=zeros(2,NbrComp);
Statistics(1,:)=hole_nbr;
Statistics(2,:)=hole_nbr_perc;

% - Curve Plotting filled in -----
figure(3)
ExampleHoles=Data(:,60);
ExampleFilled=DataFinal(:,60)-50;
hold on
plot(ExampleHoles(1:500),'LineWidth',1)
plot(ExampleFilled(1:500),'LineWidth',1)
hold off
title('Stockdata for TV4 with holes (top) and without holes (bottom)');
xlabel('Time(days)');
ylabel('Stock price (Swedish Kr)');

% - Curve Zooming Plotting filled in -----
figure(5)
ExampleFilled=DataFinal(:,60);
hold on
plot(ExampleHoles(50:80),'LineWidth',3)
plot(ExampleFilled(50:80),'LineWidth',1)
hold off
title('Closeup for the filling procedure');
xlabel('Time (days)');
ylabel('Stock price (Swedish Kr)');

% figure(4)
% plot(ExampleFilled(1:500),'LineWidth',1)
% title('Stockdata for TV4 with holes filled');
% xlabel('Time(days)');
% ylabel('Stock price (Swedish Kr)');

```