Charles University in Prague

Faculty of Social Sciences
Institute of Economic Studies

BACHELOR THESIS

Portfolio selection based on hierarchical structure of its components

Author: **Robert Ševinský**

Supervisor: **PhDr. Ladislav Krištoufek, Ph.D.**

Academic Year: **2014/2015**

Declaration of Authorship

The author hereby declares that he compiled this thesis independently, using only the listed resources and literature.

The author grants to Charles University permission to reproduce and to distribute copies of this thesis document in whole or in part.

Prague, May 14, 2015

Signature

Acknowledgments

The author is grateful especially to PhDr. Ladislav Krištoufek, Ph.D., who inspired him to investigate properties of network based portfolios and for his valuable comments and suggestions.

Abstract

This thesis investigate empirical performance of three portfolio selection and covariance matrix models. The goal is to find a strategy that outperform equally weighted portfolio in the long run and survives even in times of financial distress. Two models based on Markowitz approach absolutely failed in this context, however the last approach based on network analysis indeed outperform the market even after risk adjustment of returns. Moreover this model have sparse transaction matrix throughout time, therefore exhibit excellent properties even in the presence of transaction costs. Results for network based portfolio were obtained from running a back test on 160 member companies of S&P 500 index for 6'000 trading days.

JEL Classification	G11, G32, C10
Keywords	Portfolio selection, Minimum spanning tree, Transaction costs, Covariance matrix

Author's e-mail	`r.sevinsky@gmail.com`
Supervisor's e-mail	`kristoufek@ies-prague.org`

Abstrakt

Tato bakalářská práce zkoumá historické výsledky tří přístupů k výběru portfolia a modelování kovariančních matic. Cílem této práce je nalézt takovou strategii pro výběr portfolia, která svými výsledky překoná rovnoměrně vážené portfolio v dlouhém období a vykazuje určitou stabilitu i během krize. Dva z modelů, které jsou založené na Markowitzově přístupu k optimalizaci portfolia nedosahují těchto vlastností a jsou silně podprůměrné, zatímco portfolio založené na network analýze překonává to rovnoměrně vážené ve smyslu rizikově očištěných výnosů. Navíc tento model je velice úsporný i z hlediska transakčních nákladů a jeho výsledky jsou testovány na 160 akciích z indexu S&P 500, pro 6'000 obchodních dní.

Klasifikace JEL	G11, G32, C10
Klíčová slova	Výběr portfolia, Minimální kostra, Transakční náklady, Kovarianční matice

E-mail autora	r.sevinsky@gmail.com
E-mail vedoucího práce	kristoufek@ies-prague.org

Contents

List of Tables

List of Figures

Chapter 1

Introduction

This bachelor thesis compare performance of three portfolio selection models, namely Markowitz model, Markowitz model with penalization for transaction costs and Minimum spanning tree based portfolio. Moreover all these models are calculated using three different approaches to stock returns covariance matrix modeling. Throughout this work, performance of all models is compared based on their stock picking abilities from set of 50 member companies of S&P 500 index, where the models are reweighed daily for 2250 trading days. Moreover the MST model is compared to performance of equally weighted portfolio for 160 stocks and 6000 trading days with and without adjustments for transaction costs.

The objective of this thesis is to find a portfolio selection model that outperforms equally weighted portfolio, have limited exposure to risk of individual stocks, is feasible with reasonably low amount of transactions and requires holding of only small set of stocks. Method with such properties is highly relevant for managers of small investment funds, or high net worth individuals, since it is well diversified and requires only limited amount of wealth, while the transactions does not eat up all the investments.

The thesis is structured as follows: Chapter 2 describe general covariance matrix estimators and gives theoretical justification of their use for the purpose of estimating population logarithmic stock returns covariance matrix. In chapter 3 the construction and logic behind each portfolio selection model is described. Chapter 4 presents the empirical findings for each model and compare all the methods based on their historical performance. Finally chapter 5 summarizes our findings.

Chapter 2

Volatility and mean modeling

As Markowitz states in his paper (1) the process of selecting a portfolio can be divided into two stages, the first starts with examination of historical behavior of stock prices and ends with making forecasts of their future behavior based on historical experience. The second stage starts where the first one ends taking predictions from the first stage and ends with a choice of optimal portfolio.

This chapter is devoted to the first stage of finding feasible predictors of relevant parameters used in the second stage of portfolio selection. Considering relevant parameters there are two major notions in the modern concept of portfolio optimization and these are risk and return. The concept of risk and return modeling are topics covered in this chapter.

Last comment concerns the terms and approaches used in this chapter. The models tend to be vague relying a lot on intuition rather than exact scientific based approach. Unfortunately there is no way around this since prediction of future behavior of such a complex system as stock market have to rely on assumptions whose validity sometimes cannot be verified even ex-post. Therefore the most important feature of these models will be their historical performance in the context of realized portfolio returns behavior, rather than their theoretical validity, which cannot be assessed.

2.1 Volatility prediction

Optimal risk measure in the context of portfolio optimization is the one that most precisely capture information about interdependence between returns on assets included in the portfolio. However optimality is rather philosophical notion and this paper is more focused on finding suboptimal, but feasible solution with most appealing properties for the filtering procedure.

Present section is devoted to investigation of the most popular approach to risk modeling in the form of covariance matrices and the goal is to obtain the most precise predictor for one step ahead portfolio returns volatility forecast. Three models for volatility prediction will be defined and their properties discussed.

2.1.1 Log returns time series behavior, short comment

First of all it is important to justify use of logarithmic returns rather than ordinary rate of return for portfolio optimization. Even though the function to be maximized in second stage of portfolio selection process (see section 3.1) is reasonable to use only with ordinary returns, the logarithmic returns are preferred because of authors belief that they can be predicted with higher precision, therefore being better input into optimization model, while the objective function is "approximately sensible".[1] Secondly since stock markets are far too complex, no model for logarithmic stock returns time series behavior will be presented. The predictors of volatility and expected returns will be rather based on combination of intuition and a "misuse" of results from mathematical statistics.[2] Some assumptions about logarithmic stock returns time series, however, were already made with the decision to use rolling window when estimating moments of logarithmic stock returns vector. One assumption is that time series are not covariance stationary,[3] but their behavior is of such nature, that they can still be predicted with sufficient precision from historical data.[4]

[1]The goal is to maximize total returns of portfolio over the whole investment horizon, but this is not accomplished by maximizing weighted sum of logarithmic returns at each reweighing period (The objective function would have to be adjusted), however it at should be approximately correct for small returns. This comes from relationship between logarithmic and ordinary returns.

[2]The word misuse is used here to address the issue, that the statistics which will be used as estimators in subsequent sections are very good under assumption of stationarity, normality and infiniteness of sample as discussed later. However in this paper none of these assumptions are made and these statistics are used only because it is assumed that they may be good for modeling of stock behavior

[3]Otherwise the longest possible history would be used to predict their future behavior

[4]Otherwise there would be no sense in constructing estimators of their future development

2.1.2 Sample covariance with equal weights

Usual approach for modeling interdependence between asset returns is by sample covariance matrix. This statistic is widely used due to its appealing properties under certain assumptions. Let $\mathbf{Y} \in M(N \times \Delta t)$[5] denote matrix of logarithmic returns, then statistics $\mathbf{m} \in M(N \times 1)$ and $\mathbf{S} \in M(N \times N)$ defined as bellow are sample mean vector and covariance matrix of logarithmic stock returns respectively:

$$\mathbf{m} = \frac{1}{\Delta t}\mathbf{Y} * \mathbf{1}'$$
$$\mathbf{S} = \frac{1}{\Delta t}\mathbf{Y}\left(\mathbf{I} - \frac{1}{\Delta t}\mathbf{1}\mathbf{1}'\right)\mathbf{Y}'$$

Where $\mathbf{I} \in M(\Delta t \times \Delta t)$ is na identity matrix and $\mathbf{1} \in M(\Delta t \times 1)$ is a column vector of ones.

The rank of sample covariance matrix $\mathbf{S}$ has one important feature that is: $Rank(\mathbf{S}) \leq min\{N, \Delta t - 1\}$.[6] Simply put this property mean that when length of rolling window (minus 1) is lower than number of assets for which covariance matrix is estimated, the matrix will suffer from rank deficiency. As noted in (4) intuitive interpretation of this fact is that the data does not contain enough information to estimate covariance matrix with sufficient precision.

Unfortunately this is not the only deficiency of sample covariance matrix. As described in (7) other inefficiencies are:

- It is not robust to outliers
- It might be spurious if logarithmic returns on assets considered are dependent on common factor. This is indeed an issue when stock returns are considered and few extreme cases are described in (10) and several popular business articles such is (7).[7]
- With logarithmic returns it might happen that population variances are undefined since the probability of company liquidation without any resid-

[5] Index denoting time of the last observation of rolling window is omitted since the calculation of sample covariance matrix and mean are same for each time period. The length of rolling window Δt will always be 251 in this paper.

[6] The proof is given in Appendix A Theorem 1

[7] All of these articles are describing spurious correlations rather than covariances, however for the purpose of this text the term spurious covariance between two variables will describe a situation when these variables suffer from spurious correlation, which makes the two terms interchangeable.

ual claims left to shareholders might be positive (i.e. Positive probability of stock price 0).

- It may not be the best measure of risk for fat tailed distributions. It is of course perfect under assumption of multivariate normal distribution which is fully determined by its mean and covariance. However this assumption, even though often present in portfolio optimization literature, is not supported by empirical evidence that rather tends to support the proposition of heavy tailed and skewed stock returns probability density function as noted in (8).
- It ignores the order of observations from which it is computed giving equal weights to each observation in the sample. This may be a problem for such a dynamic system as stock market where covariance stationarity can be hardly assumed and where the most relevant informations might be contained in the most recent observations.

Some of these inefficiencies are addressed in subsequent covariance models, nevertheless even these are still far from perfect when used for modeling of stock return covariances.

There are however several reasons for usage of sample covariance as a measure of risk. The most obvious one is that it have few appealing properties in the process of portfolio optimization. As will be discussed in chapter 3 the restriction on portfolio sample variance in the form:

$$\mathbf{w}'\mathbf{S}\mathbf{w} \leq \sigma_{max} \tag{2.1}$$

Where $\mathbf{w} \in M(N \times 1)$ are weights assigned to each asset in portfolio, is convex and quadratic constraint. This is an important property, since as described in (9) chapter 4.4, constraint on expected variance of portfolio returns together with other linear constraints on portfolio weights and portfolio expected mean return as an objective functions forms a second order cone program (SOCP). The most important is that constraint 2.1 meets the property of convexity, since according to (9) chapter 1 there exists efficient and reliable algorithms that can solve even very large[8] convex problem.

Additionally under certain assumptions that will be briefly discussed in

[8]There is however not given any threshold defining what is considered to be "very large" convex problem.

section 2.1.4 it can be proven that it is an consistent estimator of population covariance matrix.

2.1.3 Exponentially weighted covariance

One way around some problems connected to usage of sample covariance matrix with equal weights for predicting future joint movements of logarithmic stock returns is to find a statistic that have more appealing empirical properties. As is discussed in (10) one such statistic is exponentially smoothed weighted correlation.[9]

The argument behind this approach is that it better captures dynamics of stock market, since it puts higher weights[10] to more recent observations. This argument seems reasonable under assumption that population covariance matrix of logarithmic returns is not stable in time and it is believed that more recent information about behavior of logarithmic returns captures more important information about their future development than do the older somehow obsolete observations.

There are multiple ways how to define a rule by which the logarithmic returns can be weighted based on their order, but there exist no rule by which optimal weights could be found. This section is describing one of approaches that is called exponential smoothing. The weights are based on work of Pozzi, di Matteo and Aste (10), and are assigned to each observation of logarithmic returns from the rolling window concerned based on the rule:

$$v_l = v_0 * exp\left(\frac{l - \Delta t}{\theta}\right), \forall l \in \{1, ..., \Delta t\} \tag{2.2}$$

Where Δt is length of rolling window, and θ denote a parameter whose optimal value will be assigned based on several measures. The weights have to satisfy a condition from which value of v_0 can be uniquely expressed as a function of θ:

$$\sum_{l=1}^{\Delta t} v_l = 1 \Rightarrow v_0 = \frac{1 - exp\left(-\frac{1}{\theta}\right)}{1 - exp\left(-\frac{1}{\theta}\Delta t\right)}$$

From this condition and equation 2.2 it is obvious that for $\theta \to \infty$ the weights

[9] Indeed the work is dedicated to correlation matrices rather than covariance matrices, however since correlation is just covariance of variables that are scaled by their standard deviation the results apply also for covariance matrices.

[10] Therefore higher importance

are constant and $v_l = \frac{1}{\Delta t}, \forall l \in \{1, ..., \Delta t\}$.[11] This is an important observation, because it makes equally weighted covariance a special case of exponentially smoothed covariance.

In the text weights will be usually expressed in vector form as $\mathbf{v} \in M(\Delta t \times 1)$, where $\mathbf{v}_{1,1} = v_{\Delta t}, ..., \mathbf{v}_{\Delta t,1} = v_1$. The exponentially weighted mean $\mathbf{m}^{exp}$ and covariance matrix $\mathbf{S}^{exp}$ are then computed from the following formula:

$$\mathbf{m}^{exp} = \mathbf{v}'\mathbf{Y}'$$
$$\mathbf{S}^{exp} = (\mathbf{Y}' - \overline{\mathbf{Y}})'[(\mathbf{Y}' - \overline{\mathbf{Y}}) \circ \mathbf{W}]$$

Where,

$$\overline{\mathbf{Y}} = \begin{pmatrix} \mathbf{v}'\mathbf{Y}' \\ \mathbf{v}'\mathbf{Y}' \\ \vdots \\ \mathbf{v}'\mathbf{Y}' \end{pmatrix} \quad and \quad \mathbf{W} = \begin{pmatrix} \mathbf{v}, \mathbf{v}, \ldots, \mathbf{v} \end{pmatrix}$$

$\overline{\mathbf{Y}} \in M(\Delta t \times N)$, $\mathbf{W} \in M(\Delta t \times N)$ are matrices and operator " $\circ$ " denote Hadamard product. Here it is necessary to add that the most recent observations of logarithmic returns are in the first row of matrix $\mathbf{Y}'$ going down to the last row, where the oldest observations from the rolling window are.

As stated at the beginning of this section there is a theoretical justification of this approach, but more importantly there is also a lot of supporting empirical research described in (10) that gives evidence for its superiority over equally weighted correlations.

Pozzi, di Matteo and Aste present several measures of correlatin matrix stability in their work (10) to show which correlation matrix and therefore covariance matrix have more appealing properties. It was found that for 300 NYSE highly capitalized stocks between years 2001 and 2003[12] for the rolling window of length $\Delta t = 251$ the weighted correlation have following properties:

- It suffers from rank deficiency as is the case for sample correlation matrix with equal weights.
- The condition number is at reasonable levels for $\theta \in [251/7, \inf)$, meaning that for these levels of θ covariance matrix is numerically stable.
- The matrix rank is maximal for $\theta \in [251/7, \inf)$.

[11] The limit can be easily computed using basic limits theorems as the limit of composite function theorem and the arithmetic limits laws defined in (5) chapter 3

[12] For more information about the data see (10)

- The average eigenvalues can be considered high for $\theta \in [251/7, \inf)$, meaning that there is little noise and less linear dependency between matrix rows and columns (After reduction of matrix into row echelon form). The description of principal component analysis and interpretation of eigenvalues is given in (11).

- The minimal average standard deviation of correlation coefficients taken over 498 dynamic correlation matrices attain its minimum in interval $\theta \in [251/3, \inf)$. Here it is of course disputable whether such a property is something desirable, maybe individual correlations in population correlation matrices are more spread.

- The autocorrelation function of average correlation coefficient with linear detrending $\overline{\rho}_t$[13] is damping faster with more displacements for lower values of θ. Unfortunately Pozzi, Matteo and Aste do not investigate behaviour of $\overline{\rho}_t$ any further, therefore there is no evidence on whether $\overline{\rho}_t$ is autoregressive process or stochastic trend with or without a drift [14] (or possibly whether it follows any other time series generating process). Only message they try to address is that the lower θ which means higher weights for more recent observations the less effect of remote past over the present. They judge that this is a required property.

- It was found that correlations $\overline{\rho}_t$ react faster to market outliers (i.e. getting to stable pre-crisis values at faster pace) than equally weighted correlations.[15]

Based on the observed properties of weighted covariances for different values of parameter θ Pozzi, Matteo and Aste claims that a value of $\theta = 251/3$ is a reasonable choice for defining weighting vector. This value indeed seems to belong among reasonable choices and will be used for purposes of portfolio optimization. The caution when using word reasonable rather than optimal is in order here indeed, since any of the measures for evaluation of properties of weighted covariance matrix presented above can hardly be used to find an optimal value of parameter θ. The most these measures can do is to give an

[13]The average correlation coefficient is calculated as $\overline{\rho}_t = \mathbf{1}'\mathbf{S}_t^{exp}\mathbf{1}/N^2$

[14]This would be very important observation since it would make any further comments invalid, because autocorrelation would be spurious and linear detrending would not be helpful in this context. This and other issues connected to time series are described in (3) chapter 18

[15]This observation was made for $\rho = 251/3$

intuitive background for what is still an acceptable value of this parameter. So when the premise that 'more important informations are captured in more recent observations' is considered to be correct, these measures can give at least some guide to in what interval should we seek the value of θ that allows us to put different weights to different observations while keeping reasonable stability of the weighted correlation matrices.

2.1.4 Covariance matrix with shrinkage

The shrinkage as described in (12) is a process of taking weighted average of sample covariance matrix and target matrix of same dimension in order to obtain better estimator of the true population covariance matrix.

This process among others helps to overcome the problem connected to rank deficiency, when it is expected that true covariance is actually of full rank. Additionally this estimator can be more robust to outlier events than equally weighted covariance matrix.[16] If static target matrix is chosen it will also make the estimates more stable in time.[17]

The methodology for computation of covariance matrix with shrinkage in this paper will follow the one proposed by Ledoit and Wolf in (13). This work is reaction on previously written work (4) by same authors, but is computationally simpler with better empirical performance, however with weaker theoretical background.

Theoretical justification and computation

As stated in (4) even though that under certain assumptions (specifically multivariate normal distribution of logarithmic stock returns) sample covariance matrix is maximum likelihood estimator, it does not necessarily make it the best estimator for small samples even under these assumptions, since maximum likelihood estimators exhibit desirable properties asymptotically[18] as shown in (14) chapter 12. Moreover assumption of normality is, as discussed in previous sections not reasonable. These are, among others previously presented properties of sample covariance matrix, the reasons for introducing new hopefully better estimator of population covariance matrix.

[16]This certainly depends on what target matrix and weights are chosen

[17]However as was already mentioned it cannot be said whether this is a good/bad feature.

[18]But these properties may not be met for small samples since the theory of maximum likelihood estimators is based mainly on asymptotic analysis that is valid only for infinite samples

For deeper understanding of ideas behind shrinkage it is worthwhile to briefly introduce the target covariance matrix from work (4), where Ledoit and Wolf propose to shrink sample covariance matrix to target matrix with strong structure.[19] As a target they suggest to use Sharpe's single index model, that is calculated for every window as follows:
Assume that stock returns are generated by following linear process,

$$y_{i,l} = \alpha_i + \beta_i x_t + \epsilon_{i,l}$$

where $y_{i,l}$ is logarithm return of i at position l of the rolling window for $\forall i \in \{1,..,N\}$ and $\forall t \in \{1,..,\Delta t\}$, x_t are logarithmic returns of S&P500 index, that represents overall market and $\epsilon_{i,l}$ residuals that are uncorrelated to one another and to market returns with stable variance in time. This model immediately imply covariance of stock returns:[20]

$$\mathbf{\Phi} = \sigma_x^2 \boldsymbol{\beta}\boldsymbol{\beta}' + \mathbf{\Delta} \tag{2.3}$$

Where σ_x^2 represents variance of S&P500 logarithmic returns, $\boldsymbol{\beta} \in M(N \times 1)$ is a vector of slope coefficients and $\mathbf{\Delta} \in M(N \times N)$ is diagonal matrix with diagonal elements $\delta_{i,i} = Var(\epsilon_i)$.

Even though this model have some value from theoretical perspective it is unnecessarily complicated as Ledoit and Wolf describe in their more recent work (13), where they propose to shrink sample covariance matrix towards a "constant correlation model". This constant correlation matrix at time t, $\mathbf{C}^t$ is calculated as follows:

$$\overline{\rho}_t = \frac{\mathbf{1}'\mathbf{S}_t\mathbf{1} - Trace(\mathbf{S}_t)}{N(N-1)} \tag{2.4}$$

$$\mathbf{C}^t_{i,j} = \overline{\rho}_t \sqrt{s_{i,i}s_{j,j}}, \forall i \neq j \quad and \quad \mathbf{C}^t_{i,i} = 1 * s_{i,i} \tag{2.5}$$

It is indeed very simple to compute and additionally according to (13) it have better empirical properties than model 2.3. Therefore this target matrix will be used in the rest of this work. Now the only question that remains to be answered is how to find optimal value of shrinkage coefficient.

The estimator used to estimate true covariance matrix at time l, $\boldsymbol{\Sigma}^l$ is

[19] Where strong structure means that the model have only a small number of free parameters.

[20] Proof is given in appendix A Theorem 2

defined as:

$$\mathbf{T}^l(\alpha_l, \mathbf{Y}^l) = \alpha_l \mathbf{C}^l + (1 - \alpha_l)\mathbf{S}^l$$

Where α_l is a shrinkage constant, $\mathbf{Y}^l$ is a matrix of logarithmic returns from a rolling window with length Δt and last observation at time l, $\mathbf{C}^l$ and $\mathbf{S}^l$ are target covariance matrix and sample covariance matrix at time l respectively and $\mathbf{T}^l : \mathbb{R}^{N\times\Delta t+1} \to \mathbb{R}^{N\times N}$.

The loss function is then computed as:

$$\mathcal{L}(\boldsymbol{T}^l, \boldsymbol{\Sigma}^l) = \left\|\boldsymbol{T}^l - \boldsymbol{\Sigma}^l\right\|^2 \tag{2.6}$$

Where " $\|\|$ " denote a Frobenius norm. And the risk function is defined as:

$$\mathcal{R}_{\boldsymbol{T}^l}(\boldsymbol{\Sigma}^l) = E_{\boldsymbol{\Sigma}^l}\left[\left\|\boldsymbol{T}^l - \boldsymbol{\Sigma}^l\right\|^2\right] \tag{2.7}$$

From the definition of the risk function it is obvious that the goal is to find α_l such that the risk function is minimized in a given time period. Such a goal can be easily achieved by finding first and second derivative of $\mathcal{R}_{\boldsymbol{T}^l}(\boldsymbol{\Sigma}^l)$. In (4) it is shown that second derivative is always positive and after putting $\frac{d(\mathcal{R}_{\boldsymbol{T}^l}(\boldsymbol{\Sigma}^l))}{d(\alpha)} = 0$, the "optimal value"[21] of alpha α_l^* can be expressed and is equal to:

$$\alpha_l^* = \frac{\sum_{i=1}^N \sum_{j=1}^N [Var(s_{i,j}^l) - Cov(s_{i,j}^l, c_{i,j}^l)]}{\sum_{i=1}^N \sum_{j=1}^N [Var(c_{i,j}^l - s_{i,j}^l) + (E(c_{i,j}^l) - \sigma_{i,j}^l)^2]}$$

To avoid any confusion $E(c_{i,j}^l)$ is the expected value of i-th and j-th coordinate of matrix $\mathbf{C}^l$, the same hold for variances and covariances.[22] Computational details of first two derivatives together with sample estimate $\hat{\alpha}^*$ of α^* are given in (4) chapter 2.5.

The very important property of α^* under following assumptions:

1. Stock returns are i.i.d. in time

2. Number of stocks is fixed and finite while number of observations goes to infinity

3. Stock returns have finite fourth moments

[21]By optimal it is meant that it minimize chosen risk function, not the philosophical concept of optimality

[22]i.e. its population value not the expected value of a constant after computation of the matrix from the sample.

is that as proved in (4) $\alpha^*(\Delta t) = o(1/\Delta t), \Delta t \to \infty$, where $o(x)$ denote so called little-o function, whose properties are discussed in (5) chapter 10. However it is necessary to take this result with reserve since the assumptions it is based on, specifically that logarithmic stock returns are i.i.d. goes against the very nature of using rolling window, because if the assumption was correct it would be best to use entire history of stock returns to estimate their covariance matrix. Additionally it goes against the assumptions made in section 2.1.3 of this paper, where the point that newer observations contains more information was made.

Therefore this result is presented just to show that under certain assumptions under which sample covariance matrix is consistent and target matrix possibly inconsistent the shrinkage constant will be 0. This imply that full weight will be given to consistent estimator of population covariance matrix as $\Delta t \to \infty$, meaning that the shrinkage estimator will be consistent under these assumptions.

2.2 Mean prediction

Throughout this work only 2 models for expected logarithmic returns will be used and those are sample and exponentially weighted mean as defined in sections 2.1.2 and 2.1.3 respectively. Sample mean will be used jointly with covariance prediction models described in sections 2.1.2 and 2.1.4. Exponentially weighted mean will be used together with exponentially weighted covariance matrix in the portfolio optimization model.

This distribution seems natural based on definition of covariances and ideas behind them, however as with every statistic used for predicting future, we can never be certain whether it is appropriate approach.

Chapter 3

Optimization process

This chapter discusses processes of optimal portfolio choice which is a final step in the whole process of portfolio selection after forecasts of stock returns behavior were made.[1] The foundations of this stage of portfolio selection were laid by Harry Markowitz in his work (1) that was published in 1952.

In this paper three models for optimal portfolio choice and two different approaches for filtering information obtained in the first stage of portfolio selection will be presented. The first two models are based on approach presented by Markowitz (1), where the only difference between them is that one of them include penalization for the existence of transaction costs inside the optimization process. The third model is based on methods of applied filtered network analysis described in (15), where stocks in the portfolio will be chosen based on their position [2] in the network. This is very different approach compared to one proposed by Markowitz and lies on the boundary between theory and pure data mining.

Optimal portfolios will be reweighted on a daily basis as new prediction of portfolio expected returns and covariances are made. The comparison of all models will be based on their empirical performance with and without presence of fixed transaction costs.

3.1 Markowitz approach

This section will discuss a very well known approach for portfolio choice that was firstly presented by Markowitz in his work (1) with more detailed discussion

[1] i.e. In this case forecasts of stock returns behavior are one day ahead forecasts of expected returns and covariance matrix

[2] Where position is determined by certain measures of centrality

in (2). It is relatively simple process which falls within convex optimization problems. The formulation of the portfolio optimization problem is defined in the following manner:

Maximize

$$\mathbf{w}'\boldsymbol{\mu}_e \tag{3.1}$$

Under Constraints

$$\mathbf{w}'\boldsymbol{\Sigma}_e\mathbf{w} \leq \sigma^2_{MAX} \tag{3.2}$$

$$\mathbf{1}'\mathbf{w} = 1 \tag{3.3}$$

$$\mathbf{1}'|\mathbf{w}| \leq 1.6 \tag{3.4}$$

$$\mathbf{w} \leq 0.15 * \mathbf{1} \tag{3.5}$$

$$-\mathbf{w} \leq 0.08 * \mathbf{1} \tag{3.6}$$

Where $\mathbf{w} \in M(N \times 1)$ is the vector of weights that is to be optimized. $\boldsymbol{\mu}_e \in M(N \times 1)$ and $\boldsymbol{\Sigma}_e \in M(N \times N)$ are one step ahead forecasts of logarithmic stock returns and their covariance respectively. Finally $\mathbf{1} \in M(N \times 1)$ is vector of ones and σ^2_{MAX} is upper bound on expected portfolio variance.

Very important property of this optimization problem is that, as also noted in section 2.1.2, it can be rewritten into the form of second order cone program that can be efficiently solved by interior point methods (9).[3]

The restrictions 3.3, 3.4, 3.5 and 3.6 are default for this whole paper.[4] The condition 3.3 is natural and it can be formulated in the way, that all long positions have to be financed from own resources, or by borrowing of funds in the form of short selling. Restriction 3.4 is essentially a limit on maximum proportion of wealth, that can be held in short positions.[5] Since the model sometimes tends to put too much weight onto individual stocks the restrictions 3.5 and 3.6 ensures that portfolio exposure to risk of individual company is always limited to some reasonable extent. Even though the limitations seems to be very relaxed in this case, there is a reason for it, which is that only few stocks are used in optimization process.[6] Restriction 3.2 limits the expected portfolio

[3]The proof of this statement is given in appendix

[4]With only exception for 3.4 and 3.6, when performance of no-short selling portfolio construction methods will be tested. The left side of restriction 3.4 is 1 and for 3.6 it is vector of zeros.

[5]Alternatively it is a restriction on maximum leverage. The value 1.6 was chosen arbitrarily as it is perceived by the author as a "sensible value"

[6]All the computations in data section are performed on a small sample of stocks, because back testing would last far too long for higher dimension problems

variance. The reason for not giving any specific value is that in order to make some reasonable back testing when the restriction 3.2 is almost always active it is necessary to make it flexible. Under the optimization algorithm 6 different levels of volatility will be allowed, and when the lowest predicted volatility will not be feasible in given period, the program will perform the optimization with another level of variance constraint. Even though this measure does not seem ideal, it is necessary, since the predicted covariance matrix is very dynamic in time, especially during financial crisis, that totally disturbed all three models presented in chapter 2. These facts are depicted on graph B.1, where the variance of equally weighted portfolio is depicted.[7] The variance attain its minimum around value of 0.05×10^{-3} and the maximum of 1.4×10^{-3} and 10^{-3} for weighted and sample covariance respectively. The maximum is therefore 28 and 20 multiple of its minimum value, which is absolutely horrific example of instability.

3.2 Portfolio optimization in the presence of transaction costs

Purpose of this section is to incorporate transaction costs into optimization process. It address the problem with Markowitz approach where a large number of very small transactions is realized at each reweighing period. In real world, where each transaction have its costs that are usually not proportional to the transaction value, but are rather fixed[8] posses a huge problem.

Therefore it seems reasonable to adjust Markowitz method for this inefficiency by penalizing for each transaction in such a way, that will lead to sparse transaction vector in each time period, while preserving structure of the Markowitz method.

3.2.1 Optimization problem

Thanks to special characteristics of the problem, namely incorporation of transaction cost into optimization process the problem have to be defined in a bit different manner than Markowitz approach. The problem can be defined as

[7] This gives very much the same information as average covariance in a matrix, just differently scaled.

[8] According to (17) this cost is $ 4.95 flat per trade, at the cheapest brokerage firm in US

follows:

Maximize

$$\mathbf{x}_t' \boldsymbol{\mu}_e \tag{3.7}$$

Under Constraints

$$(\boldsymbol{x}_t + \mathbf{w}_{t-1})' \boldsymbol{\Sigma}_e (\boldsymbol{x}_t + \mathbf{w}_{t-1}) \leq \sigma^2_{MAX} \tag{3.8}$$

$$\mathbf{1}' \mathbf{x}_t + \phi(x_t) \leq 0 \tag{3.9}$$

$$\mathbf{1}' |\boldsymbol{x}_t + \mathbf{w}_{t-1}| \leq 1.6 \tag{3.10}$$

$$\boldsymbol{x}_t + \mathbf{w}_{t-1} \leq 0.15 * \mathbf{1} \tag{3.11}$$

$$-(\boldsymbol{x}_t + \mathbf{w}_{t-1}) \leq 0.08 * \mathbf{1} \tag{3.12}$$

$$|\boldsymbol{x}_t| \leq 0.15 * \mathbf{1} \tag{3.13}$$

Where $\mathbf{x}_t \in M(N \times 1)$ is vector of trades at time t, $\mathbf{w}_{t-1} \in M(N \times 1)$ is vector of portfolio weights from the last period, where each weight is multiplied by return for the period, in order to obtain feasible solution.[9] Symbol $\phi(x_t)$ stands for transaction costs function at time t, that will be further discussed in next subsection.

All the conditions are very similar to those in previous section, however two comments are in place. First of all the objective function 3.7 can be understood as maximization of expected returns that comes from transactions realized in a given time period, and is equivalent to maximizing $(\boldsymbol{x}_t + \mathbf{w}_{t-1})' \boldsymbol{\mu}_e$. More importantly restriction 3.9, so called self financing constraint, bounds transaction vector and its interpretation is, that the value of stocks sold at time t have to cover both transaction costs and value of stocks bought at time t. This constraint therefore penalize for each transaction, since the more transactions occurs the less wealth[10] will be invested in given time period, and therefore lower wealth is expected at the end of period.

Constraint 3.9 have one important implication for functionality of the model which is that at each period when transaction occurs the portfolio weights are less than one. This is a problem since constraints are defined in a manner that is suitable for weights that sum to one at each period. There are fortunately several ways how to overcome this specific issue, one of them is described in (16), where authors propose to define the constraints as proportion of the end

[9] At the end of each period vector $\mathbf{w}_{t-1}$ is normalized to 1, just for computational comfort. It does not affect the solution

[10] Where wealth in this case is understood as sum of absolute values of weights invested in each asset

of period wealth, for example constraint 3.11 would be formulated as:

$$\boldsymbol{x}_t + \mathbf{w}_{t-1} \leq 0.15 * \mathbf{1}'(\boldsymbol{x}_t + \mathbf{w}_{t-1})\mathbf{1}$$

Other constraints would follow the same logic.

A bit different approach is used throughout this paper. Weights are always normalized to one after each period. This approach however requires to make some changes in formulation of several constraints. Constraint 3.10, 3.11 and 3.12 will be adjusted after each period, so that there is no need to, for example sell small proportion of stock that reached upper bound in the previous time period and is after reweighing slightly above this threshold in the next period. Such transactions would go against the very nature of introducing fixed transaction costs in the optimization process at the first place.

Constraints will be adjusted by following rule: After each period take the constant used for normalization of weighing vector at the end of period i.e.:

$$c_1 = 1 \quad and \quad c_t = min\left\{\frac{1}{\mathbf{1}'(\boldsymbol{x}_t + \mathbf{w}_{t-1})}; \frac{1}{1 - N * \beta}\right\}, \ \forall t \in 2, ..., T$$

And return vector for the period:

$$\mathbf{Ret}_{i,t} = P_{i,t}/P_{i,t-1}, \ \forall i \in \{1, ..., N\}$$

Where $P_{i,t}$ is price of stock i at time t. And reformulate the problem 3.2.1 by multiplying the right side of constraints 3.11 and 3.12 at time J by the rule: For $J \in (j * 200; (j+1) * 200], \ j = 0, ..., \lfloor T/200 \rfloor$ "Hadamard multiply" the constraints by $\prod_{i=j*200}^{J} \circ c_i * \mathbf{Ret}_i$.[11] This solution is just very easy way to overcome the problem and have several obvious imperfections, for example each stock have different upper bound based on its previous performance. The reason it is used in this work is that transaction costs will be small enough given the back testing period to cause any major disruption of the model and it does its job by limiting the amount of transactions to reasonably low level while not disturbing the model "very much".

[11] Where $\circ$ denote a Hadamard product of vectors

3.2.2 Transaction costs

This section is strongly inspired by work of Lobo, Fazel and Boyd (16).[12] Transaction cost in this work will be only of fixed nature and will be equivalent for each stock as well as for long and short positions.[13] The transaction cost are defined by following formulas:

$$\phi(x_t) = \sum_{i=1}^{N} \phi_i(x_{i,t}) \tag{3.14}$$

Where,

$$\phi_i(x_{i,t}) = \begin{cases} 0, & |x_{i,t}| = 0 \\ \beta, & |x_{i,t}| \neq 0 \end{cases} \tag{3.15}$$

Where $\phi(x_t)$ are total transaction costs at time t, $\phi_i(x_{i,t})$ are transaction cost paid at time t for stock i, and β[14] denote the fixed costs that are the same for each stock. The problem connected with this constraint is that it is not convex and finding perfect solution would be infeasible for higher dimension problems, since 2^N optimization problems would have to be solved in order to obtain optimal solution as discussed in (16). However Lobo, Fazel and Boyd propose efficient heuristic which can be used to obtain approximate solution by solving optimization problem when formulating this constraint as a convex constraint.

The heuristic as described in (16) and as it will be used in this work can be described by the following algorithm:

1. First of all replace constraint 3.9 with convex constraint of the form:

$$\mathbf{1}'\mathbf{x}_t + \phi^{c.e.}(x_t) \leq 0 \tag{3.16}$$

Where,

$$\phi_i^{c.e.}(x_t) = \frac{\beta_i}{u} * x_{i,t} \tag{3.17}$$

The symbol u denotes upper bound on the maximal [15] value of transaction for each asset which is strictly restricted by restriction 3.13 to be 0.2.

[12]The only main difference is that the optimization is done for several time periods and certain technical problems connected to this fact occurs

[13]No additional costs connected to holding short position will be modeled

[14]Even though there is no time index for β, it will change proportionately to wealth after each period (The reason is that when more wealth is invested the lower proportion of it are cost of individual transactions that are fixed and flat at \$ 4.)

[15]maximum from the perspective of absolute value maximality

Other symbols are explained in previous sections and the dimension of each vector is clear from context.

The function $\phi^{c.e.}(x_t)$ was not chosen arbitrarily, but it is the largest convex function that is lower or equal to $\phi(x_t)$ as shown in (16).

2. For k=0 solve the optimization problem 3.2.1 where constrain 3.13 is replaced by 3.17 and get the solution $\mathbf{x}_t^0$.

3. For k=k+1 take the optimal solution $\mathbf{x}_t^{*k-1}$ from previous and define new transaction costs function:

$$\phi_i^k(x_t) = \frac{\beta_i}{|x_{i,t}^{*k-1}| + \delta} * |x_{i,t}^k|$$

   Put this into constraint 3.13 as a transaction cost function and solve the problem 3.2.1 to obtain vector $\mathbf{x}_t^{*k}$. Where $\delta = 10^{-7}$ is a threshold, that helps us to decide whether or not should a given asset be traded. This is further explained in (16).

4. If one of the following conditions:

$$k >= 7 \ \vee \ \left\|\mathbf{x}_t^{*k} - \mathbf{x}_t^{*k-1}\right\| <= 10^{-4}$$

   is met, stop the process and set $\mathbf{x}_t^{*k}$ as the optimal vector of transactions at time t. Otherwise repeat the process once again.

This algorithm gives only approximate, but based on simulations performed by Lobo, Fazel and Boyd, quite precise solution.

Before closing this section it is worth to mention that β does not necessarily be equal to proportion of transaction costs on total wealth, but it can be used only in a way to penalize for each transaction. For example if transaction costs were \$ 4 flat per transaction and on average the position would be held for 10 days it would be reasonable to penalize each transaction by lower amount in the model, obtain optimal weight for these lower costs and then redefine them in order to obtain feasible weights. This process is however very complicated from technical perspective and given empirical performance of this and Markowitz model, compared to equally weighted portfolio and MST portfolio discussed in next section, it is better to concentrate the effort for further development of MST or generally to network based portfolio models.

3.3 Minimum spanning tree portfolio

The last model is based both on theory and data mining. This section will describe the theoretical background behind this model of an outstanding empirical performance.

Construction of Minimum spanning tree (MST) portfolio optimization model is very simple. The only input information for each period is predicted covariance matrix while the prediction of mean will be absolutely ignored. But before describing the model some basic notions from measure and graph theory need to be introduced.

This section is mostly influenced by thoughts presented in (15), (19) and (20).

3.3.1 Basic Terminology

Throughout this work the MST will be defined as a spanning tree of edge weighted,undirected, connected graph, that have minimal distance among other spanning trees of the particular graph. To make the definition more transparent it is in place to define previously mentioned mathematical terms:

Definition 3.1. Let V be a finite set, and $E(V) = \{\{u,v\}|u,v \in V, u \neq v\}$, then a pair $G = (V,E)$, $E \subseteq E(V)$ is called a graph on V. The elements of V are called vertices and those of E are called edges of graph G.[16] Additionally G is called connected if $\forall u \in V,\ \exists v \in V,\ v \neq u : \{u,v\} \in E$.

Definition 3.2. The graph G^α is said to be edge weighted, when G^α is a graph G together with a weight function $\alpha:\ E_G \to \mathbb{R}$ on its edges.

Definition 3.3. Let $e_i = \{u_i, u_{i+1}\} \in E$, then the sequence $W = e_1e_2...e_k$ is called a path of length k from u_1 to u_{k+1}, if $u_i \neq u_j,\ \forall i \neq j$.

Finally thanks to theorem 2.4 in (18) the tree can be defined as follows:

Definition 3.4. Let T be a graph, then T is called a tree iff any two vertices of T are connected by a unique path. Additionally let $\mathbb{H}_G$ denote a set of all subgraphs[17] of graph G, that are trees.

The Minimum spanning tree of an edge weighted,undirected and connected

[16] The definition is taken from (18)
[17] The term subgraph is defined in (18)

graph G^α is such $T^* \in \mathbb{H}_{G^\alpha}$[18] for which the following condition holds:

$$\sum_{e \in E_{T^*}} \alpha(e) \leq \sum_{e \in E_T} \alpha(e), \ \forall T \in \mathbb{H}_{G^\alpha}$$

I.e. the overall "weighted length of a tree" is minimal for weighted graph G^α.

In order to find optimal portfolio one more term have to be defined and it is the measure of centrality that will be used for picking assets to portfolio. The very simplest centrality measure will be used throughout this work and it will be degree of vertex:

Definition 3.5. Let $G = (V, E)$ be a graph. For each $u \in V$ define set $N_u = \{v \in V | \{u, v\} \in E\}$. Then the degree of vertex u from graph G is defined as $d_G(u) = |N_u|$, where operator "||" computes number of elements in a set.

3.3.2 Construction of MST portfolio

Let V_t^{Assets} denote a set of assets at time t from which optimal portfolio is supposed to be chosen.[19] Then V_t^{Assets} is a set of vertices and $E_t = \{\{u, v\} | u, v \in V_t^{Assets}, u \neq v\}$ is a set of edges. Together with metric α_t as defined bellow they form a weighted,undirected, connected graph $G_t^{\alpha_t} = (V_t^{Assets}, E_t)$, for which the MST will be found.

The weighting function $\alpha_t : V_t^{Assets} \times V_t^{Assets} \to [0, \sqrt{2}]$ is defined as in (19):

$$\forall u, v \in V_t^{Assets} : \quad \alpha_t(u, v) = \sqrt{2 * (1 - \rho_{u,v}^t)}$$

Where $\rho_{u,v}^t$ is one step ahead forecast of correlation coefficient between assets u and v at time t calculated from any model presented in chapter 2.[20]

The weighing matrix $\boldsymbol{\alpha}^t \in M(N \times N)$, that assign weights to each edge is defined as:

$$\boldsymbol{\alpha}_{u,v}^t = \alpha_t(u, v)$$

The intuitive interpretation behind this distance function is that it clusters closely related assets[21] since the lower distance is assigned to pairs of stocks with higher correlations. The rule for picking stocks will be such that peripheral

[18] Where $\mathbb{H}_G$ is clearly nonempty, since G is connected

[19] In this work it will hold that $V_j^{Assets} = V_i^{Assets}, \ \forall i, j \in \{1, ..., T\}$ (i.e. The set of assets is the same for each time period)

[20] More about derivation of weighting matrix and computation of correlations is given in appendix A.3

[21] Closely relate means assets with high joint correlation coefficient

stock (i.e. those with low centrality measure) will be chosen to take part in portfolio, since I believe that such stocks will have higher diversification potential as they stands out of clusters. As shown in (18) weighting function α_t have also an appealing mathematical property, since it is a metric on V_t^{Assets}.[22] Furthermore as will be shown in empirical section the MST constructed from this measure tends to be quite stable in peripheries, which have the advantage, that there are very low transaction cost related to this method of portfolio selection.

After defining the weights over all edges, the MST can be easily found by following algorithm:

- Find the [N+2*(N-1)]-th lowest value of matrix $\boldsymbol{\alpha}_t$,then put all higher elements equal to zero and denote the new instrumental matrix as $\boldsymbol{J}^t$.
- The pairs of assets with positive inputs in $\boldsymbol{J}^t$ forms all the edges of minimum spanning tree. I.e. the MST of graph $G_t^{\alpha_t}$, is T_t^* defined as:

$$T_t^* = (V_t^{Assets}, E_t^*)$$

Where,

$$E_t^* = \{\{u, v\} \in E_t | \boldsymbol{J}_{u,v}^t > 0\}$$

Finally the last step of portfolio selection is to pick the stocks to invest in. The rule used in this work is following:

1. Put all positive elements of $\boldsymbol{J}^t$ equal to one and denote the new matrix as $\boldsymbol{D}^t$.
2. For each stock compute its degree in MST as: $d_{T_t^*}(u) = \sum_{j=1}^{N} \boldsymbol{D}_{u,j}^t$, and construct a vector $\boldsymbol{d}_u^{T_t^*} = d_{T_t^*}(u)$, where degree of each stock is on a given coordinate in each time period (i.e. the order of stocks in this vector does matter).[23]
3. Find the $\lfloor N/c \rfloor$-th lowest value of vector $\boldsymbol{d}^{T_t^*}$ and denote it as λ, where $c \in [1, N]$ is a constant that specify how much stocks will be included in portfolio.

[22] Which means that ordered pair (V_t^{Assets}, α_t) forms a metric space in each time period t, at least for equally weighted sample correlation

[23] This will be important in following part, because the rule is constructed in such a way that minimizes amount of transactions

4. Invest certain proportion of total wealth to stocks that satisfies following properties: $u \in V_t \wedge d_{T_t^*}(u) \leq \lambda$.

There are three comments in order here, first of all the constant c is very important and have to be chosen wisely since it affects the amount of stocks included in a portfolio and its "peripherality". Secondly it would be worth to investigate behavior of different networks than MST, for example PMFG as described in (11) since MST have way too much stocks with degree of 1.

The third comment is to item 4, which is not specifically saying what the weights should be. It is on purpose, since the rule will strongly influence the amount of transactions in each period. If it is chosen, for example to put equal weights to each stock that satisfy property in item 4,[24] then the amount of transaction would be huge since proportion of wealth invested in each stock changes after every period as returns have to be taken into consideration. Additionally the amount of stocks that satisfies property 4 is not stable in time. Therefore it is reasonable to find algorithms that overcome these problems and lead to feasible solution with sparse transaction matrix. These algorithms are not discussed in this paper, however when comparing the results in data section, certain methods that limit transactions will be used. Additionally it would be sensible to test whether certain penalization for lower "peripherality of stocks" should be used, since the previously discussed method assume that stocks with degree less than λ are equivalently attractive for portfolio optimization.[25]

At last it is worth to mention that,as further discussed in next section, this model have an outstanding empirical performance, with possibly sparse transaction vector and low amount of stocks held at each period, making it an very good market beating portfolio construction method even for small investors. Additionally the computational requirements are very low, the problem can be solved even for large number of assets[26] and the solution can be found even in basic programs such as MS Excel. Furthermore even though the last two pages were a bit technical,[27] the method is very simple and intuitive, only requirements being basic knowledge of correlation matrices and few notions from graph theory.

[24]Of course with the rule that weights sum to one.

[25]This would be probably more an issue for different networks as PMFG, or measures of centrality, since most of the stocks in MST have degree equal to 1

[26]It took around 27 seconds to find portfolio for 471 stocks in MATLAB on basic notebook. For the Markowitz portfolio as defined in 3.1 it would take much longer, certainly more than 30 minutes.

[27]It was only so because of the academic nature of the text

Chapter 4

Data analysis

This chapter compares empirical results of portfolio selection methods described in chapter 3 against each other and against equally weighted portfolio. Additionally the results for different predictors of mean and covariance are compared.

4.1 Data

The data that are used for testing the models are historical daily close stock prices of S&P 500 member firms with 10 years long history, that are adjusted for stock dilution and dividend payment.[1] These data were downloaded from Yahoo Finance.

From the data at disposal it is clear, that the reweighing of portfolio will be done at the end of each trading day, when the close price enters the portfolio optimization model and new optimal weights for next day are calculated. After this happens it is expected that all transactions necessary for portfolio reweighing will be executed at the very same close price.

Unfortunately there is one bigger issue connected to the data. It is that these data capture historical price development only for stocks that survived in the market and are members of prestigious index (i.e. for past winners). This may be a problem since strategy that performs better in picking among winners does not necessarily perform better when choosing from all assets ex ante, when the winners are not known.

[1] More about the computation of adjusted close stock prices is given in (21)

4.2 All models

In this section all the portfolio selection models described in chapter 3, with input of one step ahead covariance predictors from chapter 2 will be presented.

Results are presented in table 4.1, however beforehand some comments are in order here especially to methodology of Markowitz portfolio back testing, where several problems occurred. The biggest issue connected to this model was, that the covariance matrices are very dynamic in time as illustrated in graph B.1, where development of equally weighted portfolio variance in time is presented. It is therefore not reasonable to set unique value of acceptable variance for the whole period, since sometimes there would be no solution feasible and other times the weights would be given only to stocks with best past performance. Simplistic solution was adopted to overcome this problem and it was to set several levels of variance and when the lowest level is not feasible, then the optimization will be done with higher level of variance constraint.[2] Another a bit less urgent problem was that variance constraint is defined in a bit different manner for Transaction penalizing Markowitz approach, therefore the results in table 4.1 for this type of portfolio are based on slightly different variances , furthermore are calculated with different algorithm.[3] Last comment concerning the data is that performance of all the models have to be taken with reserve, since it is based only on 50 stocks, that are listed in appendix A.4, and 2250 trading days.

Table 4.1 contain information on performance of daily reweighed portfolios for period 21.4.2006-27.3.2015 that included 2'250 trading days. The information on cumulative returns over whole period and four sub periods is given, together with "information ratio".[4]

As is clear from the table 4.1 performance of Markowitz portfolios is rather poor for the whole period compared to equally weighted portfolio and MST based portfolio. The cumulative returns being between 111.1%-166%, while the equally weighted portfolio achieved 242% and even better the MST portfolio with 264%. The same story it is with information ratios, where the maximum

[2]These levels were set at 0.0003, 0.0004, 0.000065, 0.0009, 0.00022 and 0.003, only when short selling is allowed, then they are set at higher levels. Moreover, unfortunately these variances were set at different level for Markowitz portfolio based on exponentially weighted covariance matrix, however, information ratios should still provide valuable information.

[3]'sqp' MATLAB is used for ordinary Markowitz portfolio optimization and 'fmincon' for the one that penalize for transaction costs.

[4]Technical details for their computation is given in appendix A.4.

from Markowitz based portfolios is 0.56, while it is 0.62 and 0.67 for equally weighted an MST portfolios respectively.

Both Markowitz portfolio approaches are strongly underperforming during financial crisis[5] On the other hand they perform exceptionally well for third period 4/2011-4/2013. This type of behavior is characteristic for Markowitz approach, going from one extreme to another.

When it is allowed for short selling the performance is even worse as described in the last column.[6] The development of returns is much more appealing for network based portfolios that outperform the equally weighted portfolio during crisis, and even during last two periods, at least when it comes to cumulative returns. Also their performance over whole testing period is better from both the cumulative returns and information ratio perspective.

The last comment is that even though the very complicated procedure described in section 3.2, led to severe decrease in amount of transactions[7] to around 1/5 the amount of ordinary Markowitz portfolio model it did not enhance the performance at all, actually quite the opposite. Moreover the amount of transactions is still not low enough given the performance as will be presented in next section. Therefore subsequent research effort will be put on deeper analysis of MST portfolio properties, that is much less technically demanding and offer superior performance.

[5]Only except for Markowitz portfolio based on weighted covariance matrix.

[6]Data for Markowitz portfolio with transaction adjustment are not present when short selling is allowed since it take far too long to compute them and since it is expected that their performance would hardly outperform the equally weighted portfolio.

[7]By transaction it is understood change in weight higher than 0.0005, caused by either different returns of stock in given period as is the case of equally weighted portfolio and very often even for networks, but also for regularly reweighing of portfolios caused by other constraints.

Table 4.1: Comparison of all models

			No short selling						Leverage 1.6
Date			4/2006 - 4/2009	4/2009 - 4/2011	4/2011 - 4/2013	4/2013 - 3/2015	4/2006 - 3/2015	Transactions	4/2006 - 3/2015
	Equally weighted portfolio	Cumulative ret.	-14,3%	113,8%	25,9%	48,3%	242,0%	16253	
		Mean/stdev ann.	-0,16	2,23	0,60	1,76	0,62		
Sample Covariance	Markowitz	Cumulative ret.	-36,3%	86,9%	71,3%	28,0%	161,0%	50486	29,4%
		Mean/stdev ann.	-0,52	1,92	1,58	0,94	0,53		0,14
	Transaction cost portfolio	Cumulative ret.	-38,6%	75,2%	61,1%	21,9%	111,1%	10438	
		Mean/stdev ann.	-0,58	1,81	1,46	0,77	0,45		
	Network	Cumulative ret.	-9,0%	106,1%	31,0%	48,5%	264,0%	19204	
		Mean/stdev ann.	-0,10	2,42	0,71	1,68	0,67		
Covariance with Shrinkage	Markowitz	Cumulative ret.	-34,0%	67,8%	77,1%	26,4%	147,9%	52331	25,0%
		Mean/stdev ann.	-0,49	1,58	1,72	0,90	0,51		0,12
	Network	Cumulative ret.	-9,0%	106,1%	31,0%	48,5%	264,0%	19204	
		Mean/stdev ann.	-0,10	2,42	0,71	1,68	0,67		
Weighted Covariance	Markowitz	Cumulative ret.	-16,4%	65,2%	42,8%	34,9%	166,0%	51854	26,3%
		Mean/stdev ann.	-0,22	1,44	1,08	1,17	0,56		0,13
	Transaction cost portfolio	Cumulative ret.	-25,8%	54,3%	49,1%	24,6%	112,7%	10662	
		Mean/stdev ann.	-0,39	0,96	1,43	0,83	0,47		
	Network	Cumulative ret.	-12,3%	107,7%	26,6%	52,7%	251,9%	23482	
		Mean/stdev ann.	-0,13	2,53	0,62	1,84	0,63		

4.3 MST portfolio vs. Equally weighted portfolio

This section is dedicated to more detailed examination of Network based portfolios and comparison of its performance to equally weighted portfolio. The technical details to this section are included in A.4.

One of the biggest issue with previous analysis is that the models were tested on only 50 stocks, which enabled us to compare methods all together, however the amount of stocks was too low for construction of networks with meaningful properties. Reason for including such a limited number of stocks in previous analysis was that the Markowitz approach is computationally very complicated and it took around 18 hours to perform backtesting for the portfolio with penalization for transaction costs. This illustrate another advantage of Network analysis over Markowitz optimization, which is very fast computation even with large number of stocks.

This analysis will be performed for 160 stocks from S&P 500 index with at least 25 years of quotation history. The stock prices are taken for 6'000 trading days in period 11.7.1991-1.5.2015.[8]

Results for networks composed based on two different models of volatility and one extra model based on network analysis not described in the theoretical section will be presented here. Both results with and without including transaction cost will be presented.

4.3.1 No transaction costs

Results for portfolios without the presence of transaction cost are presented in table 4.2 on page 30. Most important results are shown in last column that is summarizing performance of models over the whole 25 year long period and from which it is obvious that network based models strongly outperform the equally weighted portfolio. Not only that cumulative returns are higher for these methods reaching as high as 98 multiple[9] of initially invested wealth,[10] but also the information ratios are higher. This means that on average peripheral stocks exhibit better properties than "ordinary stock". For method described in section 3.3 the cumulative returns are around 1.5 times higher than for Equally weighted portfolio and information ratio is 1.22 times higher, while the

[8] The dates are in format d.m.yyyy

[9] These are results of the model that is not specifically described in this work, however is based on network analysis and serves as illustrative example of network analysis potential.

[10] The cumulative returns seem to be very high for all models, however this is caused by already mentioned fact, that all the models are tested on "past winners" only.

"Network method" offers more than twice as good returns as Equally weighted portfolio. Development and comparison of cumulative portfolio returns for each method are plotted on graph 4.1, page 31.

Another observation from table 4.2 is that sample covariance model 2.1.2 seems better for constructing networks than the weighted covariance matrix from section 2.1.3. This statement however have to be taken with reserve, since it is not further supported in table 4.3 after transaction costs are introduced.

The other 6 columns in table 4.2 shows performance of these models in 1'000 trading days long sub periods. From the development of these quantities it is clear that all the models are closely connected to the market performance.[11] This claim is however more obvious from scatter plots B.4 presented in appendix B, that captures the relationship between models by plotting their joint daily returns. The correlation among these models is very strong and positive, which is in accordance with their construction. Portfolios are constructed so that equal weights are given to each peripheral stocks, which mean that such a portfolio is obviously predisposed to react on general stock market fluctuations.

Back to table 4.2 the network methods clearly outperform equally weighted portfolio in each sub period achieving better results both in good times, but also in times of financial distress (In periods 1999-2003 when the so called Internet bubble burst it achieved 1.4 higher cumulative returns and in 2007-2011, when the recent financial crisis started, it achieved almost double the cumulative returns). This is very appealing property, since the portfolio managers are very concerned about huge drawdown their strategy might bring in bad times. Other models also perform very good, always, except during period 1995-1999, outperforming the equally weighted portfolio and even during 95-99 period they are very close to equally weighted portfolio in performance.

Another excellent property of network based portfolios is that they achieve such results with only very limited amount of stocks. After certain simplistic measures that adjust model presented in section 3.3 the amount of stock to invest in was arbitrarily set to 40, and for Network Method to 32. This is very good since it shows that "market" can be beaten with only limited amount of assets that is feasible even for high net worth individuals, or small funds.

Additionally the maximum exposure to each stock is limited to 2.5% and 3.125%, which makes the portfolio robust to failure of individual companies.

[11] Where market is represented by equally weighted portfolio.

Table 4.2: Cumulative returns and information ratios for Network models, no transaction costs.

Date		7/1991 - 6/1995	6/1995 - 6/1999	6/1999 - 6/2003	6/2003 - 5/2007	5/2007 - 5/2011	5/2011 - 5/2015	7/1991 - 5/2015
Equally weighted portfolio	Cumulative returns	132%	175%	44%	117%	31%	76%	4544%
	Mean ann. return	23,67%	28,87%	9,65%	21,51%	7,01%	15,37%	17,42%
	Mean/stdev ratio ann.	2,35	1,97	0,47	1,90	0,24	0,92	0,9539
Sample Covariance Network	Cumulative returns	180%	172%	45%	133%	33%	103%	6885%
	Mean ann. return	29,52%	28,60%	9,80%	23,68%	7,56%	19,00%	19,45%
	Mean/stdev ratio ann.	2,74	1,97	0,49	2,13	0,31	1,28	1,1633
Weighted Covariance Network	Cumulative returns	172%	164%	45%	129%	36%	101%	6467%
	Mean ann. return	28,67%	27,57%	9,82%	23,09%	8,20%	19,13%	19,14%
	Mean/stdev ratio ann.	2,71	1,90	0,49	2,10	0,33	1,28	1,14
Network Method	Cumulative returns	169%	198%	62%	128%	59%	109%	9719%
	Mean ann. return	28,20%	31,49%	12,93%	22,94%	12,36%	20,32%	21,16%
	Mean/stdev ratio ann.	2,69	2,18	0,65	2,07	0,51	1,39	1,28

Table 4.3: Cumulative returns and information ratios for Network models in presence of transaction costs.

Flat cost per trade	$4,00	11.7.1991 - 1.5.2015								
	Initial wealth invested	$ 1 000 000,00	$ 100 000,00	$ 50 000,00	$ 20 000,00	$ 10 000,00	No Cost	Transactions	Stocks Held	Max. Weight
Equal initial investment with reweighing	Cumulative returns	3864%	3651%	3412%	2698%	1508%	3889%	6692	160	0,63%
	Mean annualized return	16,52%	16,25%	15,92%	14,79%	12,02%	16,55%			
	Information ratio ann.	0,91	0,90	0,88	0,82	0,66	0,92			
Sample Covariance Network	Cumulative returns	5216%	4594%	3903%	1829%	0%	5286%	8177	40	2,90%
	Mean annualized return	18,00%	17,37%	16,57%	12,93%	0,00%	18,06%			
	Information ratio ann.	1,10	1,06	1,01	0,79	0,00	1,10			
Weighted Covariance Network	Cumulative returns	5270%	4632%	3923%	1797%	0%	5341%	12934	40	2,90%
	Mean annualized return	18,05%	17,41%	16,60%	12,85%	0,00%	18,11%			
	Information ratio ann.	1,10	1,06	1,01	0,78	0,00	1,10			
Network Method	Cumulative returns	7501%	6817%	6056%	3774%	0%	7577%	10597	32	3,70%
	Mean annualized return	19,80%	19,32%	18,74%	16,41%	0,00%	19,85%			
	Information ratio ann.	1,23	1,20	1,16	1,02	0,00	1,23			

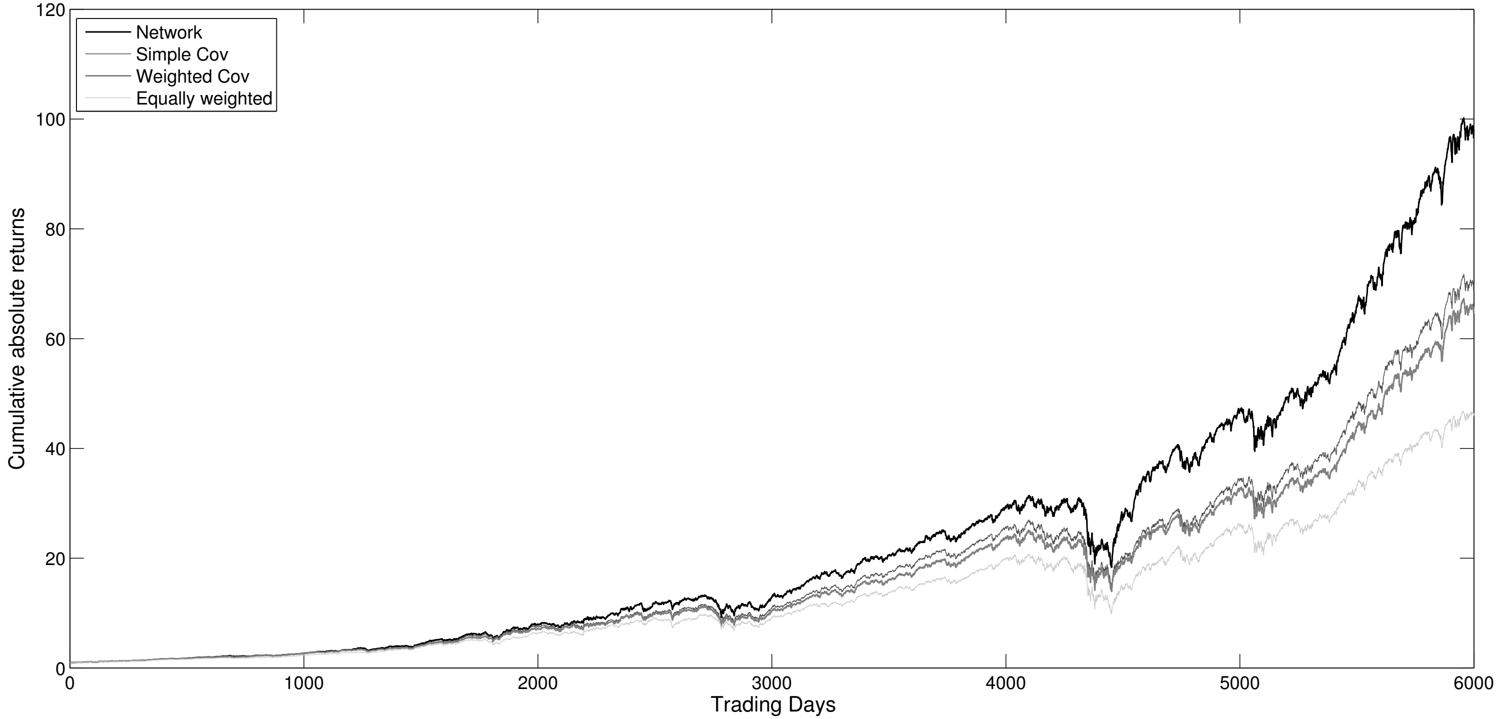

Figure 4.1: Cumulative returns over 6000 trading days for three network models compared to equally weighted portfolio without transaction costs.

4.3.2 Network based portfolio performance in the presence of transaction costs

Problem connected to Markowitz portfolio selection was not only its inferior performance in context of poor returns, however the model was also very wasteful when it came to transactions, where as much as 112'070 transactions occurred during 2'250 days making average of 49.8 transactions a day, with only 50 stocks. This is an awful average that reflects the fact that no penalization for transactions was made. And even after penalizing for transaction cost in the model, transactions remained at quite high levels.

The situation is much better for network based models, after certain actions for selecting between the most peripheral stocks are taken. Specifically the stocks are ordered and according to whether they are peripheral and what is their order, they get the weights assigned.[12] Additionally some adjustments are taken to account for changes of weights caused by different returns of individual stocks with positive weights in a given period. It is so because different returns of stocks results in that the weights of individual stocks in portfolio deviate from their equal state, which is equally weighted peripheral stocks portfolio. The deviation is of such a form that past winners are given higher weights, which negatively affects the cumulative returns of portfolio.[13] Last comment concerning adjustments is that transaction cost were incorporated in optimization process, however only for portfolio with initial investment of $ 1'000'000, therefore it is expected that for portfolios with different initial investment they are covered from external sources.

After applying these measures the MST portfolios need only as little as 8'177 and 12'934 transactions for sample and weighted covariance models respectively, which is shown in table 4.3. This is an excellent result for 6500 trading days with 160 stocks when the average is 1.26 and 1.99 transactions per trading day. Unfortunately measures applied to limit for transactions that are described in appendix A.4 are very costly when it comes to reduction of cumulative returns. Therefore it strongly depends on portfolio size how much to limit for transactions, and since the larger portfolios realize economies of scale when it comes to transaction costs as they are usually set at fixed level per trade, it might be reasonable to perform reweighing more frequently for these portfolios.

[12]More about this in appendix A.4

[13]Data supporting this claim are not presented here, but it is based on authors observations.

The adjustment were made not only for network based portfolios but even for equally weighted portfolio.[14] They were done following the same logic of adjusting network portfolios to make the comparison as realistic as possible, the basic logic being that when weight of particular stock reaches certain predetermined value the weight will be lowered on account of stocks with lowest weights. This portfolio is denoted in the table 4.3 as "Equal Initial investment with reweighing" and will be referred to as market representing portfolio in the rest of this section.

Table 4.3 present results for these adjusted models where transaction costs are already incorporated in returns, where the costs for each transaction are measured as percentage of total wealth and are deducted from returns in period when they occur. To be as realistic as possible these cost are taken from large brokerage firm with lowest transaction costs in 2015 according to (17)[15] and are set at $ 4 flat per trade.[16] The results are then computed over the whole period for different scales of portfolios. First of all it is worth to notice column "No cost" in table 4.3, which shows the results of adjusted models with $ 0 transaction costs and compare it to table 4.2 overall returns. It shows how costly it is to bound the amount of transactions to low levels, as annualized returns dropped by 1-1.5% for all the models, when compared to unadjusted portfolios from table 4.2.

The last three column mainly shows the stability of each portfolio. Transactions are very low for all the models as discussed above and the maximum exposure to risk is limited at reasonably save levels in range of 0.63-3.7% with the possibility for further reduction by including more stocks into decision process or allowing for lower portfolio peripherality. Other extremely important property is that network based portfolios tend to outperform the market representing portfolio with only 0.25 and 0.2 multiple of its stocks, being very parsimonious to wealth requirements. Last comment concerning stability of MST portfolio, that is not presented in the table, is that around 80% of wealth is invested into 40 particular stocks throughout time.

[14] Such a portfolio is extremely costly, since reweighing would have to be made after every period, when daily returns are incorporated in the portfolio. For this particular example as much as 959'680 transaction would have to be realized in order to keep such a portfolio (i.e. 159.947 transactions a day).

[15] Clearly the transaction costs were decaying in last decades with the development of information technologies, so the results are not historically feasible, however since it is reasonable to expect that transaction costs will be declining even further in future, this might not be a problem for validity of results presented in this work.

[16] No other trading costs, as for example bid-ask spread or capital gain taxes are controlled for.

The other columns show the performance measures for different levels of initially invested wealth. It is apparent that costs have little effect on portfolio performance when large initial wealth was invested[17] differences being expressed in fractions of annualized return percentages. However as the amount of initial investment decreases to the wealth levels approachable by small individual investors, the transaction costs start to reduce annualized returns significantly, to the extreme of $ 10'000 portfolio that would end up in loss of all money invested for all network portfolios. The effect of initial wealth on cumulative returns of portfolio is depicted on graph B.3, where the case of sample covariance based MST portfolio is modeled.

To finalize this section, it is apparent that network based portfolios outperform the market representing portfolio for levels of wealth above $ 50'000. For lower levels of wealth the advantage of less transactions is huge, and lead to underperformance of network portfolios. However it is important to say that for such low levels of wealth the rebalancing rule would be readjusted to minimize transactions as much as possible. The performance comparison of all models is depicted on graph B.2, where annual returns for years 1991-2015 are presented. From the graph it is clear that all models are strongly correlated and that the reaction of network based portfolios to crises are very similar to market ones, with the exception of Network method, that have absolutely astonishing properties, since its annual returns were, except for 2002, always positive during concerned period.

4.3.3 Possible shortcomings of Network based portfolios

The one possible problem is that the selection method does not explicitly control over portfolio variance and it accepts the weighted covariance of peripheral stocks, whatever its value is. One possibility is to introduce riskless asset, however even if such asset exists it would have extremely low return and additionally it would strongly increase transaction costs, since the wealth would have to be moved in every period, therefore this approach is avoided in this work. Moreover it seems rather idealistic goal to control for variance even in Markowitz portfolio since, as implied from figure B.1 covariance matrices are extremely unstable in time, therefore either new and better measures of risk have to be presented, or the Markowitz optimization will not be reliable in controlling of risk as well.

[17] The term large investment is used for $ 1'000'000 and more.

For the above presented reasons it is taken as much higher benefit that maximal exposure to risk of individual assets is constrained at very low levels that could be arbitrarily changed by including more stocks. Additionally the stability of portfolio composition is very valuable, not only that it limits transactions, but stable portfolios seem to strongly outperform the ones that changes according to short term noise as is the case in Markowitz portfolio that tends, by construction, change according to what stocks performed better recently.

4.4 Main imperfections of presented data

There are several issues connected to analysis performed above. First of all as already mentioned for section 4.2/4.3 analysis is performed on only 50/160 stocks that survived at least 10/25 years and made it to S&P 500 index, which is definitely not a representative sample of overall stock market. Additionally taxes were not incorporated into calculation of cumulative returns, which certainly poses certain problem since long term investment gains are usually taxed at preferential rate.

Chapter 5

Conclusion

The goal of this thesis was to find a portfolio selection method that outperform equally weighted portfolio while being reasonably stable in time. Three methods were tested, namely Markowitz model, Markowitz model with penalization for transaction costs and Minimum spanning tree based portfolio. It was found that both Markowitz approaches underperformed the equally weighted portfolio even without the presence of transaction costs. Moreover very often these portfolios were exposed largely to risk of individual companies, usually holding few stocks with large weights. The MST based portfolio described in section 3.3 on the other hand outperformed the equally weighted one reaching to 19.45% of annualized returns in period 11.7.1991 - 1.5.2015, when the equally weighted portfolio reached only 17.42%. Furthermore it was found that the set of peripheral stocks is very stable in time, giving the possibility to make transaction vector sparse for each time period. Other outstanding property of MST method is that it outperforms the equally weighted portfolio holding only fraction of its stocks.[1] Therefore the network analysis is ideal candidate for further research in context of portfolio optimization.

It was also found that the covariance models presented in chapter 2 slightly affect the portfolio performance even in the long run, however the materiality was quite low. The biggest difference was observed for MST portfolio, which, when it was constructed from the sample covariance matrix returned 19.45% annually while for weighted covariance matrix it was only 19.14%.

The main takeaway from this thesis is that it is better to invest in peripheral stocks, rather than holding market portfolio and that further research in

[1]This fraction was set to 0.25 in our analysis.

network analysis constructed portfolios is meaningful.[2]

[2]Where peripheral stocks in this context are stock with lowest degree in MST, for more see section 3.3.2.

Bibliography

[1] Markowitz, H. (Mar., 1952). Portfolio Selection. *Journal of finance*, Vol. 7, No. 1. [Online]. Retrieved May 10, 2014, from http://www.jstor.org/discover/10.2307/2975974?uid=3737856&uid=2&uid=4&sid=21106683487283

[2] Markowitz, H. (1959). *Portfolio Selection: Efficient diversification of investments*. New York: John Wiley & Sons, Inc..

[3] Wooldridge J.M. (2009). *Introductory Econometrics a Modern Approach, 4th edition.* Mason: South-Western Cengage Learning.

[4] Ledoit, O., & Wolf, M. (Oct., 2001). *Improved Estimation of the Covariance Matrix of Stock Returns With an Application to Portfolio Selection.* [Online]. Retrieved September 7, 2014, from http://www.econ.upf.edu/docs/papers/downloads/586.pdf

[5] Hájková, V., Johanis, M., John, O., Kalenda, O. & Zelený, M. (2012). *Matematika, 1st edition.* Praha: MATFYZPRESS.

[6] Magnus, J.R., & Neudecker, H. (Jan., 2007). *Matrix Differential Calculus with Applications in Statistics and Econometrics, 3rd edition.* Chichester: John Wiley & Sons, Ltd..[Online]. Retrieved January 5, 2015, from http://www.janmagnus.nl/misc/mdc2007-3rdedition

[7] Business Insider (2012). *15 of the Most Spurious Correlations in the Stock Market.* [Online]. Retrieved January 6, 2015, from http://www.businessinsider.com/15-spurious-stock-correlations-2012-4?op=1

[8] Rachev, S.T., Stoyanov, S.V., Biglova, A. & Fabozzi, F.J. (2003). *An Empirical Examination of Daily Stock Return Distributions for U.S.Stocks.*

[Online]. Retrieved January 31, 2015, from http://www.ams.sunysb.edu/~rachev/publication/tr_an_empirical_examination.pdf

[9] Boyd, S., & Vandenberghe, L. (2004). *Convex Optimization.* New York: Cambridge University Press. [Online]. Retrieved January 6, 2015, from https://web.stanford.edu/~boyd/cvxbook/bv_cvxbook.pdf

[10] Pozzi, F., Di Matteo, T., & Aste, T. (June, 2012). Exponential smoothing weighted correlations. *The European Physical Journal B*,Vol. 85. [Online]. Retrieved October 25, 2014, from http://link.springer.com/article/10.1140%2Fepjb%2Fe2012-20697-x

[11] Shlens, J. (April, 2014). *A Tutorial on Principal Component Analysis.* [Online]. Retrieved October 27, 2014, from http://arxiv.org/pdf/1404.1100.pdf

[12] Kwan, C.Y.C. (Dec., 2011). *An Introduction to Shrinkage Estimation of the Covariance Matrix: A Pedagogic Illustration.* [Online]. Retrieved January 8, 2015, from http://link.springer.com/article/10.1140%2Fepjb%2Fe2012-20697-x

[13] Ledoit, O., & Wolf, M. (June, 2003). *Honey, I Shrunk the Sample Covariance Matrix.* [Online]. Retrieved January 10, 2015, from http://www.econ.upf.edu/docs/papers/downloads/691.pdf

[14] Bartoszynski, R., & Niewiadomska-Bugaj, M. (2008). *Probability and Statistical Inference, 2nd edition.* New Jersey: John Wiley & Sons, Inc..

[15] Pozzi, Z., Di Matteo, T., & Aste, T. (April, 2013). Spread of risk across financial markets: better to invest in the peripheries. *Scientific Reports*, Vol. 3, Article number: 1665. [Online]. Retrieved May 10, 2014, from http://www.nature.com/srep/2013/130416/srep01665/full/srep01665.html

[16] Lobo, M.S., Fazel, M. & Boyd, S. (2006). Portfolio Optimization with Linear and Fixed Transaction Costs. [Online]. Retrieved May 25, 2014, from https://faculty.washington.edu/mfazel/portfolio-final.pdf

[17] TopTenReviews (2015). *Online Stock Trading Review.* [Online]. Retrieved February 16, 2015, from http://online-stock-trading-review.toptenreviews.com/optionshouse-review.html

[18] Harju, T. (2011). *Lecture Notes on Graph Theory.* [Online]. Retrieved January 9, 2014, from http://cs.bme.hu/fcs/graphtheory.pdf

[19] Mantegna, R.N. (1999). Hierarchical Structure in Financial Markets. *The European Physical Journal B*,Vol. 11. [Online]. Retrieved May 10, 2014, from http://link.springer.com/article/10.1007%2Fs100510050929

[20] Bonanno, G., Caldarelli, G., Lillo, F., Micciche, S., Vandewalle, N., & Mantegna, R.N. (2004). Networks of equities in financial markets. *The European Physical Journal B*,Vol. 38. [Online]. Retrieved May 10, 2014, from http://arxiv.org/pdf/cond-mat/0401300.pdf

[21] Yahoo (2015). About historical prices. [Online]. Retrieved February 10, 2015, from https://help.yahoo.com/kb/finance/historical-prices-sln2311.html

Appendix A

Technical comments and proofs

A.1 Covariance and mean modeling

Theorem 1. Let $\mathbf{S}$ denote sample covariance matrix, then rank $r(\mathbf{S}) \leq min\{N, \Delta t - 1\}$.

Proof. From (6) chapter 1.7 it follows that for any two matrices $\mathbf{C} \in M(m \times n)$ and $\mathbf{D} \in M(n \times p)$ it holds that $r(\mathbf{CD}) \leq min\{r(\mathbf{C}), r(\mathbf{D})\}$. Therefore for,

$$\mathbf{S} = \frac{1}{\Delta t}\mathbf{Y}\left(\mathbf{I} - \frac{1}{\Delta t}\mathbf{11'}\right)\mathbf{Y'}$$

it must hold that $r(\mathbf{S}) \leq min\left\{r(\mathbf{Y}), r\left(\mathbf{I} - \frac{1}{\Delta t}\mathbf{11'}\right)\right\}$.

Clearly since $\mathbf{Y} \in M(N \times \Delta t)$ it holds that $r(\mathbf{Y}) \leq min\{N, \Delta t\}$. Additionally since $\mathbf{I} - \frac{1}{\Delta t}\mathbf{11'}$ is an idempotent matrix and from (5) chapter 9 theorem 57 the rank of idempotent matrix equals to its trace which is equal to:

$$tr(\mathbf{I} - \frac{1}{\Delta t}\mathbf{11'}) = tr(\mathbf{I}) - \frac{1}{\Delta t}tr(\mathbf{11'}) = \Delta t - \frac{1}{\Delta t}\Delta t = \Delta t - 1$$

Then

$$min\left\{r(\mathbf{Y}), r\left(\mathbf{I} - \frac{1}{\Delta t}\mathbf{11'}\right)\right\} = min\{N, \Delta t - 1\}$$

□

Theorem 2. Let logarithmic stock returns be generated by time series generating process

$$y_{i,l} = \alpha_i + \beta_i x_t + \epsilon_{i,l} \tag{A.1}$$

With properties as described in section 2.1.4. Then covariance matrix can be expressed by following formula:

$$\mathbf{\Phi} = \sigma_x^2 \boldsymbol{\beta}\boldsymbol{\beta}' + \mathbf{\Delta}$$

Where definitions of individual matrices is defined in 2.1.4.

Proof. Let $\phi_{i,j}$ denote i-th row and j-th column of matrix $\mathbf{\Phi}$. It is necessary to show that $\phi_{i,j} = Cov(y_i, y_j)$.

$$\begin{aligned} Cov(y_i, y_j) &= E[(y_i - E(y_i))(y_j - E(y_j))] = E[(\beta_i(x - E(x)) + \epsilon_i)(\beta_j(x - E(x)) + \epsilon_j)] \\ &=^* \beta_i\beta_j E[(x - E(x))^2] + E[\epsilon_i\epsilon_j] = \sigma_x^2 \beta_i\beta_j + E[\epsilon_i\epsilon_j] \end{aligned}$$

Where $E[\epsilon_i\epsilon_j] = 0, \forall i \neq j$ and $E[\epsilon_i\epsilon_j] = Var(\epsilon_i) = \delta_{i,i}, \forall i = j$. This and third equality $=^*$ holds from properties of residuals (They are uncorrelated to each other, to market returns and have stable variance). Now it is clear that $\mathbf{\Phi}$ is a covariance matrix implied by model A.1. □

A.2 Optimization process

The purpose of this section is to show that Markowitz portfolio optimization problem:

Maximize

$$\mathbf{w}'\boldsymbol{\mu}_e \tag{A.2}$$

Under Constraints

$$\mathbf{w}'\boldsymbol{\Sigma}_e\mathbf{w} \leq \sigma_{MAX}^2 \tag{A.3}$$

$$\mathbf{1}'\mathbf{w} = 1 \tag{A.4}$$

$$\mathbf{1}'|\mathbf{w}| \leq 1.6 \tag{A.5}$$

$$\mathbf{w} \leq 0.2 * \mathbf{1} \tag{A.6}$$

$$-\mathbf{w} \leq 0.1 * \mathbf{1} \tag{A.7}$$

Can be rewritten into SOCP problem. Constraints A.4,A.6 and A.7 are already in the form of SOC constraint. However it is necessary to show that constraint A.5 can be rewritten as a set of linear constraints. This is done by introducing auxiliary variable and the process is described in (16) section 1.2. Additionally constraint A.3 is not in the SOC form, but it can be replaced by the following equivalent condition, that is in SOC form.

Theorem 3. Condition,

$$\mathbf{w}'\boldsymbol{\Sigma}_e\mathbf{w} \leq \sigma^2_{MAX}$$

is equivalent to

$$\left\|\boldsymbol{\Sigma}_e^{1/2}\mathbf{w}\right\| \leq \sigma_{MAX}$$

Where $\boldsymbol{\Sigma}_e^{1/2}\boldsymbol{\Sigma}_e^{1/2} = \boldsymbol{\Sigma}_e$

Proof. Fist of all it have to be shown that $\boldsymbol{\Sigma}_e^{1/2}$ exist.

To prove this it is necessary to show that $\boldsymbol{\Sigma}_e$ is positive semidefinite matrix. For model of $\boldsymbol{\Sigma}_e$ presented in section 2.1.2 the proof is straightforward because according to this model

$$\begin{aligned}\boldsymbol{\Sigma}_e &= \frac{1}{\Delta t}\mathbf{Y}\left(\mathbf{I}-\frac{1}{\Delta t}\mathbf{1}\mathbf{1}'\right)\mathbf{Y}' \\ &= \frac{1}{\Delta t}\mathbf{Y}\left(\mathbf{I}-\frac{1}{\Delta t}\mathbf{1}\mathbf{1}'\right)\left(\mathbf{I}-\frac{1}{\Delta t}\mathbf{1}\mathbf{1}'\right)\mathbf{Y}'\end{aligned}$$

Which clearly must be positive semi-definite. For model presented in section 2.1.3 the proof is given in (10). And for the last one 2.1.4 the positive semi-definiteness comes from the fact that model 2.1.2 is positive semi-definite.

From the positive semi-definiteness property, theorem about spectral decomposition of symmetric matrix (Theorem 51) and from Theorem 52 in (5) chapter 9, it is clear that all eigenvalues of $\boldsymbol{\Sigma}_e$ are nonnegative (Theorem 52) and therefore $\boldsymbol{\Sigma}_e^{1/2}$ exist and is also symmetric (Immediate consequence of Theorem 51).

Now it is straightforward that:

$$\begin{aligned}\sqrt{\mathbf{w}'\boldsymbol{\Sigma}_e\mathbf{w}} &= \sqrt{\mathbf{w}'\boldsymbol{\Sigma}_e^{1/2}\boldsymbol{\Sigma}_e^{1/2}\mathbf{w}} = \sqrt{(\boldsymbol{\Sigma}_e^{1/2}\mathbf{w})'(\boldsymbol{\Sigma}_e^{1/2}\mathbf{w})} \\ &= \left\|\boldsymbol{\Sigma}_e^{1/2}\mathbf{w}\right\|\end{aligned}$$

□

A.3 MST weighing matrix

The weighting matrix that is used to construct MST portfolio is defined for each time period by following formula:

$$\boldsymbol{\alpha}_t = SQ(2 * (\mathbf{11'} - \boldsymbol{\rho_t})) \tag{A.8}$$

Where function $SQ : M(N \times N) \rightarrow M(N \times N)$ is defined as:

$$SQ(A)_{i,j} = \sqrt{A_{i,j}}, \forall i, j \in \{1, .., N\} \tag{A.9}$$

And,

$$\boldsymbol{\rho_t} = [Diag(\boldsymbol{\Sigma}_t)]^{1/2} \boldsymbol{\Sigma}_t [Diag(\boldsymbol{\Sigma}_t)]^{1/2} \tag{A.10}$$

Where $\boldsymbol{\Sigma}_t$ is the one step ahead forecast of covariance matrix at time t derived from any model presented in chapter 2 and function $Diag : M(N \times N) \rightarrow M(N \times N)$ is defined as:

$$Diag(\mathbf{A})_{i,j} = \begin{cases} \mathbf{A}_{i,j}, & \forall i = j \\ 0, & else \end{cases} \tag{A.11}$$

A.4 Data analysis, technical details

In this section it is discussed how the cumulative returns, annualized returns and information ratios were calculated. Moreover the algorithm to minimize number of transactions for MST portfolios is described.

Let $\mathbf{w}_t$ denote a column vector of optimal weights at time t obtained from any portfolio selection model. Let $\mathbf{Ret}_{t+1}$ denote a column vector of ordinary returns at time $t+1$, where its components are calculated according to formula $ret_i^{t+1} = P_i^{t+1}/P_i^t$, where P_i^{t+1} denote price of stock i at time $t+1$. Then the cumulative returns C_{ret} are calculated according to formula:

$$p_t = \mathbf{w}_t' * \mathbf{Ret}_{t+1} \tag{A.12}$$

$$C_{ret} = \prod_{t=1}^{T-1} p_t - 1 \tag{A.13}$$

Where $(p_t - 1) * 100\%$ is daily portfolio return.

Annualized returns r_{ann} and information ratios IR_{ann} are calculated as follows:

$$r_{ann} = (C_{ret} + 1)^{251/(T-1)} \tag{A.14}$$

$$\sigma_{ann} = stdev(p_t)^{\sqrt{251}} \tag{A.15}$$

$$IR_{ann} = \frac{r_{ann}}{\sigma_{ann}} \tag{A.16}$$

Where $stdev(p_t)$ denote sample standard deviation of daily portfolio returns. The constant 251 is used to annualize the data since it is assumed that on average 1 year have 251 trading days.

Algorithm applied to limit amount of transactions for all periods is such that all stocks are ordered and when the centrality measure of a given stock is less than certain quantity that is specified in subsection 3.3.2, where the constant c from this section is set at $c = 4$, and it is also less then $\lfloor N/c \rfloor$-th stock with such property, then it is assigned with positive weight. Initially all weights are set to be equal, however returns of individual stocks tend to differ, which lead to deviation from the state of equal weights, that is required,[1] moreover there is no limit on risk exposure to individual stock.

For above mentioned reasons the maximal weights are limited to 2.9% and when this treshold is exceeded, the weight is adjusted to 2.5%, while the wealth is distributed to 4 stocks with lowest weights and to cover for transaction costs. This approach is very simplistic, and have very poor performance as apparent from comparison of tables 4.2 and 4.3. It is used here only to show that even after controlling for amount of transaction, this approach can surpase the market.

Finally tickers of stocks used for back testing purposes are for section 4.2: 'MMM,ALXN,ALTR,ADI,ADM,AIZ,T,BBT,BMY,CPB,HSIC,CBG,XEC,CTXS,CME,COH,CMA,CMI,DTV,DD,ETFC,ETR,BEN,GT,GWW,TEG,KSU,GMCR,LLL,LEG,MKC,MDT,MSI,NTAP,JWN,ORLY,ORCL,PAYX,POM,PFE,PBI,PFG,PGR,REGN,SRE,SNA,SO,UNP,VNO,WYNN'

And for section 4.3:
'MMM,ABT,ADBE,AET,AFL,GAS,APD,ARG,AA,ALTR,MO,AEP,AXP,AIG,AME,AMGN,APC,ADI,AON,APA,AAPL,AMAT,ADM,T,ADSK,ADP,AVY,BHI,BLL,BAC,BK,BCR,BAX,BBT,BDX,BBY,HRB,BA,BMY,CA,COG,CPB,CAH,CCL,CAT,CELG,CNP,CTL,CERN,SCHW,CVX,CB,CI,CINF,CTAS,CSCO,C,CLX,CMS,KO,CCE,CL,CMCSA,CMA,CSC,CAG,COP,ED,GLW,

[1] It was observed by the author, that the more equal weights are the better performance is achieved.

COST,CSX,CMI,CVS,DHR,DE,D,DOV,DOW,DTE,DD,DUK,DNB,ETN,EC L,EIX,EA,EMC,EMR,ETR,EOG,EQT,EFX,ES,EXC,EXPD,XOM,FDO,FAS T,FDX,FITB,FISV,FLS,FMC,F,BEN,FTR,GCI,GPS,GD,GE,GIS,GPC,GT, GWW,HAL,HOG,HAR,HRS,HAS,HCP,HP,HES,HPQ,HD,HON,HRL,HST,H UM,HBAN,ITW,IR,TEG,INTC,IBM,IP,IPG,IFF,JEC,JNJ,JCI,JPM,KSU,K, KEY,KMB,KLAC,KR,LB,LH,LRCX,LM,LEG,LEN,LUK,LLY,LNC,LLTC,L MT,L,LOW'

Appendix B

Graphs

This section is used for the purpose of displaying important graphs that did not fit into the body of text.

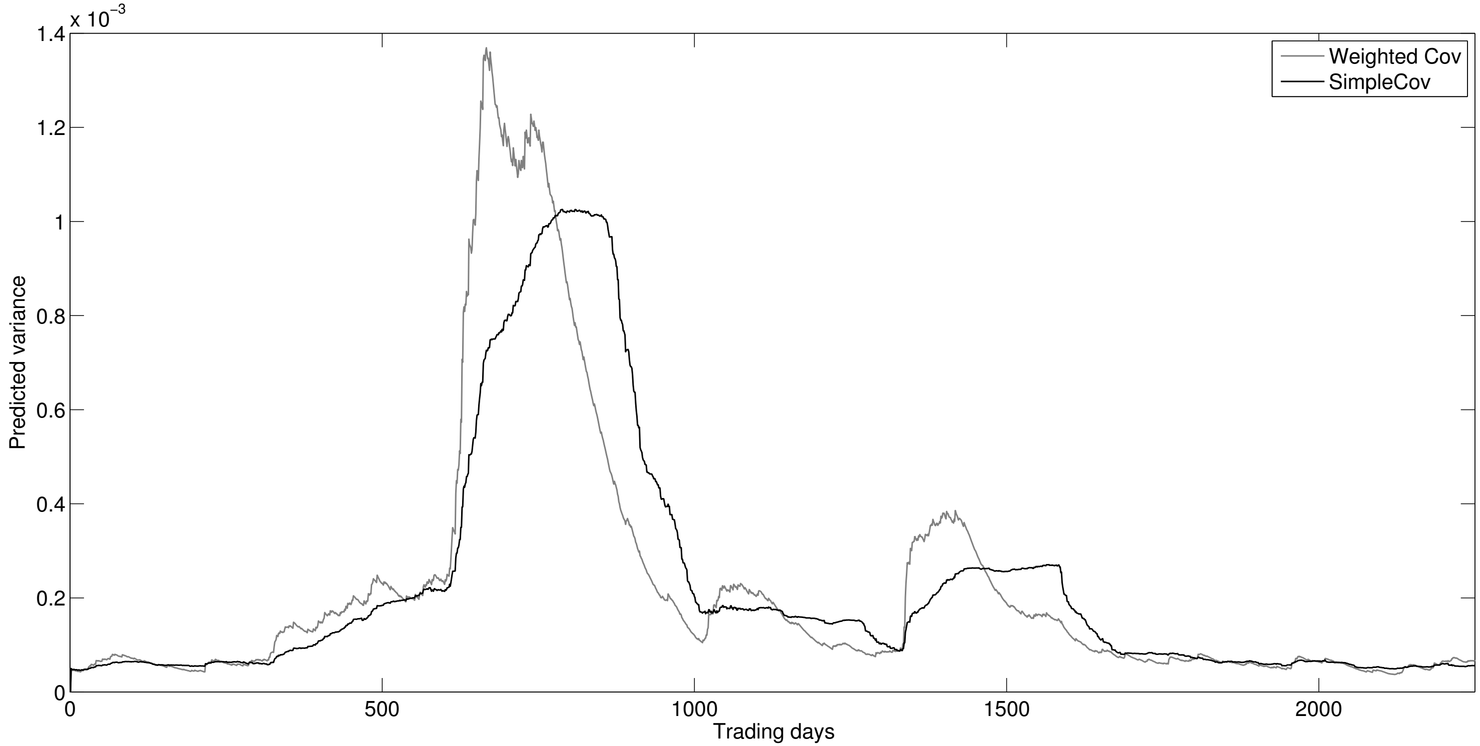

Figure B.1: Development of one step ahead variance forecast for equally weighted portfolio during period 21.4.2006 - 27.3.2015, including 2250 trading days for two models of covariance. Apparently the weighted covariance model reacts much faster on market fluctuations as it puts higher weights on recent observations.

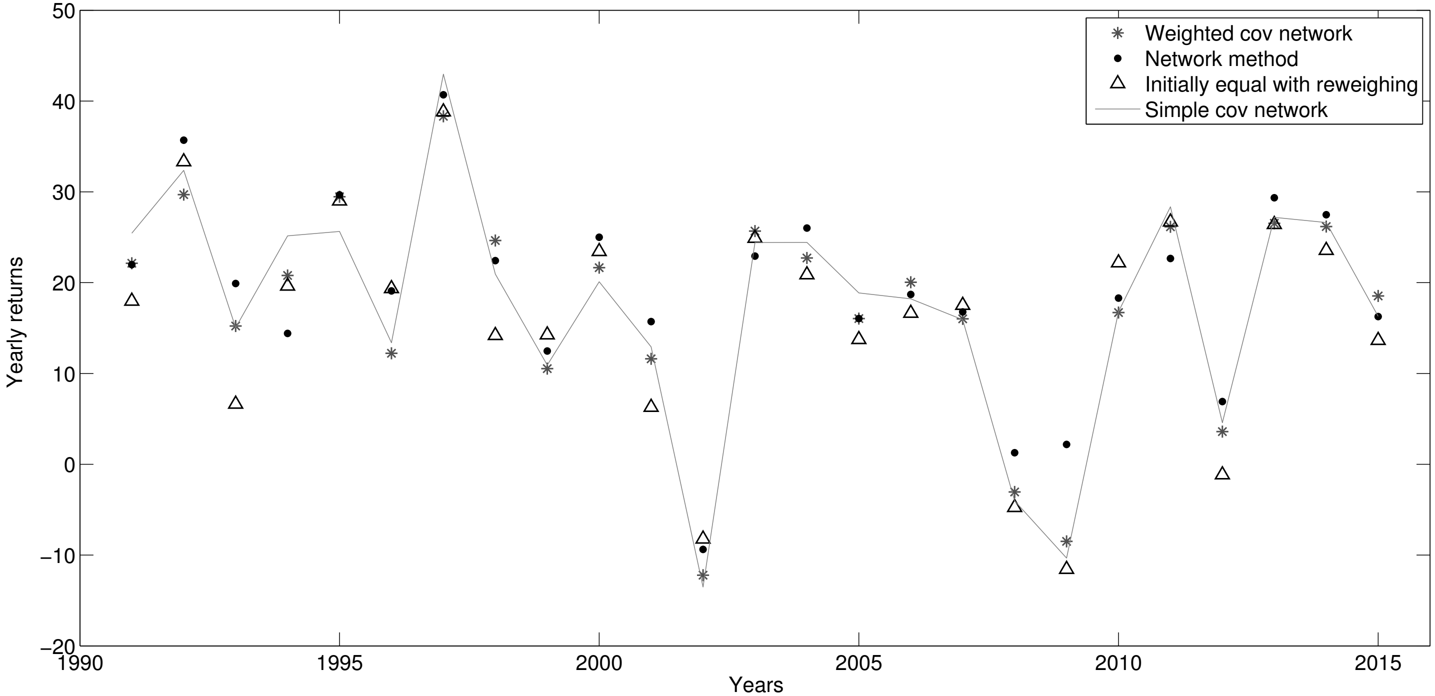

Figure B.2: The returns for each year starting in 1991 ending in 2015 are depicted for each portfolio selection methods with \$ 4 flat transaction costs and \$ 100'000 initial portfolio. Among others there is obvious correlation between these models.

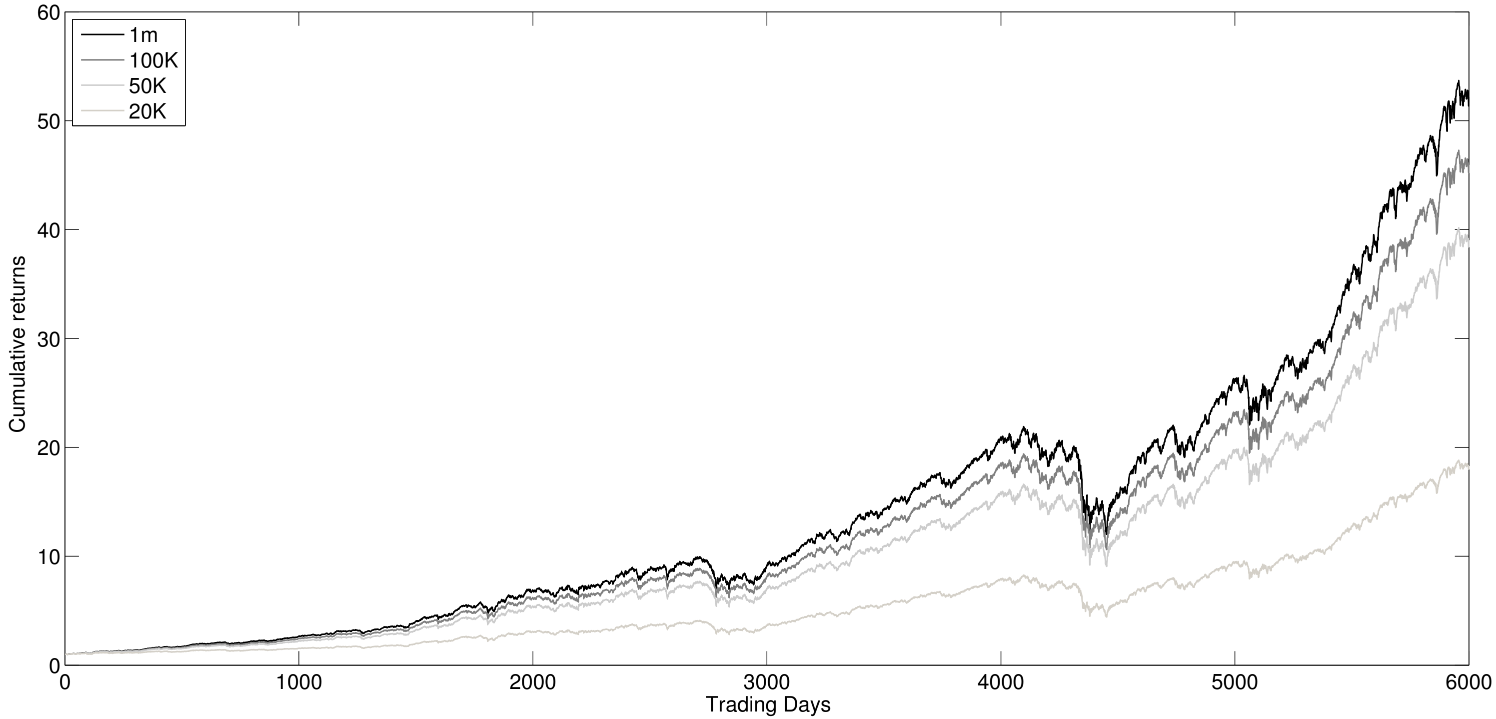

Figure B.3: Cumulative returns over 6000 trading days of sample covariance matrix based Network for different levels of initial invested wealth (Specified in legend) with transaction costs of $ 4 flat per trade.

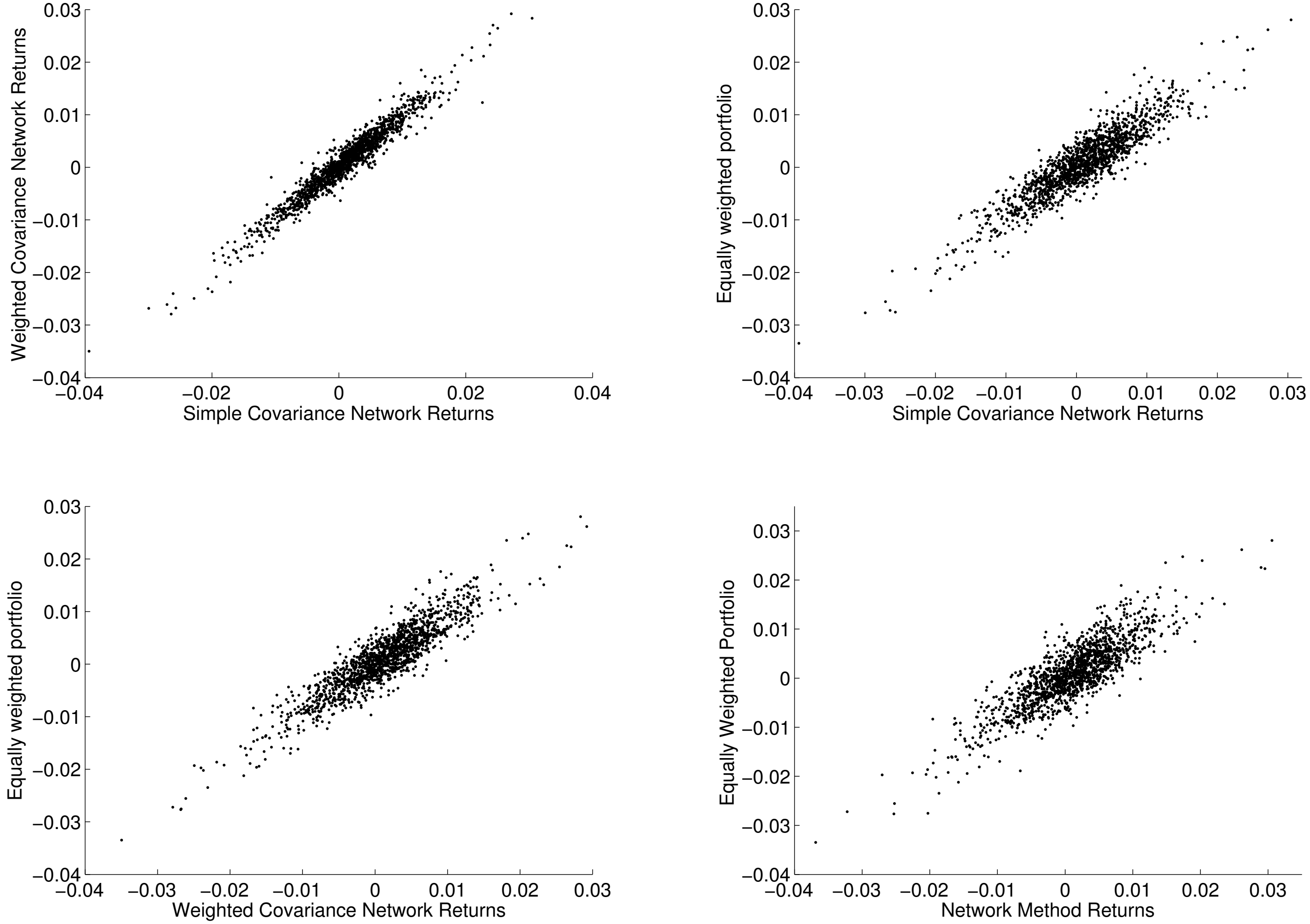

Figure B.4: Scatterplots of daily returns from portfolio selection models described in section 4.3. (Name of model is given at each axis)

Bachelor Thesis Proposal

Author	Robert Ševinský
Supervisor	PhDr. Ladislav Krištoufek, Ph.D.
Proposed topic	Portfolio selection based on hierarchical structure of its components

Topic characteristics The portfolio theory has been in center of interest among financial theorist for a long time. The goal of this thesis is to build on knowledges in this area and to explore new approach towards portfolio selection based on portfolio taxonomy and comparison of this approach with classical method of portfolio selection described by Harry Markowitz in the presence of transaction cost. At the beginning of my work I will describe Markowitz approach to portfolio selection. In the next parts of my work I will focus on the building process of portfolio hierarchical structure using methods Minimal spanning tree and Hierarchical tree. Subsequently I assemble a model by which I assign weights to every asset in portfolio based on its position in hierarchical structure. The advantage I expect this approach has in comparison to portfolio selection based on Markowitz is that its structure should be more stable in time and therefore associated with lower transaction costs. At the end I will compare performance of both approaches using historical data.

Hypotheses

1. Is hierarchical portfolio more stable in time than Markowitz portfolio?
2. Is performance of hierarchical portfolio better than Markowitz portfolio in the presence of transaction costs?
3. Do the portfolio selection based on hierarchical structure of its components have place in modern portfolio theory with respect to its performance?

Methodology Covariance matrix, Minimal-spanning tree, Hierarchical tree.

Outline

1. Introduction
2. Methodology
   Markowitz portfolio
   Minimum spanning trees (MST) & hierarchical trees (HT)
3. Literature review
4. Data & results
5. Conclusions

Core bibliography

1. Bonanno, G., Caldarelli, G., Lillo, F., Miccich., Vandewalle, N. & Mantegna, R.N. (2004) Networks of equities in financial markets. European Physical Journal B 38: 363-371
2. Onnela, J.P., Chakraborti, A., Kaski, K., Kert, J. & Kanto, A. (2003) Dynamics of market correlations: Taxonomy and portfolio analysis. Physical Review E 68, 056110.
3. Pozzi, F., Di Matteo, T. & Aste, T. (2013) Spread of risk across financial markets: better to invest in the peripheries. Scientific Reports 3.
4. Mantegna, R.N. (1999) Hierarchical structure in financial markets. European Physical Journal B 11: 193-197
5. Markowitz, H. (1952) Portfolio Selection. The Journal of Finance, Vol. VII, No. 1: 77-91

Author

Supervisor