Turun yliopisto
University of Turku

OPTIMAL EQUITY PORTFOLIO CONSTRUCTION

The efficiency of mean-variance optimization with in-depth covariance matrix estimation and portfolio rebalancing

Master's Thesis
in Accounting and Finance

Author:
Joonas Hämäläinen

Supervisors:
Professor Mika Vaihekoski
Assistant Professor Antti Fredriksson

18.3.2011
Turku

Turun kauppakorkeakoulu • Turku School of Economics

CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 INTRODUCTION

1.1 Background for the topic

Defining an optimal investment portfolio has been a major topic of interest in financial research for decades, ever since Markowitz's groundbreaking study about portfolio selection in 1952. The idea of holding assets according to a composition based on quantitative measures has spun off a vast amount of literature and is an ever popular subject of discussion in the field of investment management.

Markowitz's theory and quantitative asset management, however, have not triumphed without problems in real world applications. Complex mathematical models have been built to achieve better performance but it seems that the big effort is not paying off; a simple strategy dividing wealth evenly across available assets can still outperform the more complex models when their performance is evaluated in the real world stock market (DeMiguel, Garlappi & Uppal 2009b). This raises a question whether the sophisticated models formulated in research are valuable in practice and if they are feasible to use in portfolio construction.

Hesitation to use portfolio optimization models is observable in practice: Rasmussen (2003, 192–193) identifies some of the reasons why asset managers have not widely adopted the use of quantitative strategies. Among these reasons are the shortcomings of portfolio optimization models, such as portfolio weight instability – the optimal weights can change dramatically when even the slightest changes are made in the input parameters of the model. According to Rasmussen, instead of utilizing portfolio optimization, asset managers tend to stick to basic rules and simple strategies, or rely on their expertise in constructing a portfolio on the basis of qualitative assessment.

On the other hand, a recent survey by Fabozzi, Focardi and Jonas (2006b, 5–21) reports that the use of quantitative methods in equity portfolio management has increased. Several of the survey participants reported using automated portfolio management strategies and moreover, the most widely used method for optimization was Markowitz's mean-variance approach. With the continuously improving performance of personal computers today, even complex optimization problems can be conveniently solved for use in portfolio construction. This has undoubtedly had an influence on asset management firms adopting portfolio optimization methods. (Fabozzi, Focardi & Kolm 2006a, 8–9.)

The mixed views provided by the research in the field and the increasing adoption rate of quantitative strategies by asset management firms create a cloud of uncertainty over the topic. Currently, there appears to be no consensus on the usefulness of quantitative asset management. This makes portfolio optimization a particularly interesting and current topic for research in this thesis.

1.2 Previous research on the subject

Research on Markowitz's mean-variance portfolio theory has been relatively vivacious in the past. Several papers have documented some of the problems of the mean-variance optimization. For example, Michaud (1989) lists reasons why the mean-variance optimization has not been overly popular among asset managers. According to Michaud, the main problem is that the mean-variance optimization requires input parameters to be far more precise than the inaccurate information available is; estimation errors in the inputs are magnified by the optimization process.

The length of data used for parameter estimation appears to be fundamental: Broadie (1993, 21–22) presents that a small amount of historical data leads to estimation errors, but a long time horizon can respectively produce erroneous estimates because of the possibility that the parameters are nonstationary, i.e. changing over time. Broadie states that when choosing the length of data to use, there is a trade-off between estimation error and stationarity.

Best and Grauer (1991) examine the effect of changes in asset return means to portfolio weights and performance. They find that a very small change in an individual asset mean can change the composition of the portfolio completely, yet the performance of the portfolio is not affected much. This obviously has the implication of increasing transaction costs tremendously.

The poor performance of mean-variance efficient portfolios compared to simpler strategies has been documented in the past. Bloomfield, Leftwich and Long (1977) compare tangency portfolios to a simple portfolio that places equal weight on each asset. They find out that the optimizing models are not worth implementing, reporting that the tangency portfolios had higher volatility than the simple strategy. Similar results are obtained by Jorion (1991) who compares optimized portfolios with different parameter estimation methods, finding that none of them outperforms an equal-weight strategy out-of-sample in terms of Sharpe ratio.

Due to the problems and poor performance of mean-variance portfolios, a lot of research has been conducted in search for remedies; this research has focused mainly on handling estimation error. Jorion (1985; 1986) reports that a shrinkage estimator of the expected return, where the sample average is shrinked toward a common mean, improves the estimation and outperforms the classical sample mean. More recently, Pástor (2000) presents an approach where the investor includes belief in an asset pricing model in their portfolio selection, combining views based on historical data and asset pricing models.

Ledoit and Wolf (2003; 2004) suggest shrinkage methods for covariance matrix estimation to improve portfolio performance. They find out that the shrinkage estimators outperform the classical sample-based estimate. Furthermore, Jagannathan and Ma (2003, 1654) state that short-sale constraints in portfolio optimization can improve the performance of sample covariance matrix portfolios to the level of shrinkage portfolios.

A different approach to estimation error is demonstrated by Garlappi, Uppal and Wang (2007) who consider a robust allocation method where constraints reflecting the chance of estimation error are set for the optimization process. They find that the optimal portfolio is a weighted average of the minimum variance and mean-variance portfolios. Furthermore, Kan and Zhou (2007) suggest that estimation error can be reduced by holding additional risky portfolios.

The aforementioned improvements to deal with estimation error have not provided a solution to the problems of portfolio optimization: In a very recent and comprehensive study, DeMiguel et al. (2009b) compare the out-of-sample performance of 14 portfolio optimization models to a simple benchmark strategy where the investor divides their wealth evenly across available assets. They evaluate several modern approaches to deal with estimation error, over seven empirical datasets of monthly returns. None of the tested models is shown to consistently beat the benchmark strategy in terms of out-of-sample Sharpe ratio, certainty-equivalent return or trading volume.

Explaining the poor performance of the optimal portfolios, DeMiguel et al. (2009b, 1919–1920) state that because the parameter estimates produce extreme portfolio weights that can be all but optimal, the equal-weight benchmark strategy can actually produce smaller allocation mistakes. Moreover, DeMiguel et al. (2009b, 1936–1944) show that a mean-variance portfolio can require several hundreds, even thousands of months of historical data for parameter estimation to outperform the equal-weight strategy. Furthermore, the models that are designed to reduce estimation error require only a slightly shorter amount of data to beat the equal-weight strategy. On the other hand, this result appears to be in controversy with Broadie (1993, 21–22) stating that possible nonstationarity in parameters does not favor an extensively long estimation window.

In another, follow-up paper, DeMiguel, Garlappi, Nogales and Uppal (2009a) develop constrained strategies that manage to attain a higher Sharpe ratio than the equal-weight strategy. However, they do not include transaction costs in the portfolio performance, and the turnover rates of the newly developed strategies are considerably higher than that of the equal-weight strategy.

A promising direction is presented by Ding and Fei (2008) who use daily return data and employ a dynamic conditional correlation model to estimate volatility and correlations of assets. Moreover, they rebalance portfolios every two days and apply a turnover constraint to restrict transaction costs. Their results indicate that it is possible to outperform the equal-weight strategy, although the performance evaluation is done with only one dataset, applying strict constraints to the turnover rates. This leaves open questions about factors that improve the portfolio performance.

Overall, the mixed research results create a confused mindset about whether the mean-variance portfolio optimization is worth implementing in practice and if a simple benchmark strategy can consistently be outperformed. This gives reason for investigating optimal portfolio construction further and it is fundamentally the purpose of this thesis.

1.3 Purpose and outline of the thesis

1.3.1 Extensions to previous research

The previous studies regarding portfolio performance bear some shortcomings. Firstly, most of the research has been conducted with monthly asset returns, often very old data spanning across decades, instead of more accurate daily information. Most notably, monthly returns are used for parameter estimation in the comprehensive study by DeMiguel et al. (2009b). Secondly, transaction costs incurred by portfolio rebalancing are traditionally not included in the performance results but reported separately. Currently, there appears to be no study that examines the performance of portfolio optimization strategies across several datasets, utilizing daily returns and including transaction costs in the analysis.

Furthermore, closer examination of the effects of rebalancing frequency and estimation window length on portfolio performance is needed in order to find out more about the factors affecting performance. The thesis extends previous research considering the aforementioned matters, evaluating the performance of modern extensions to mean-variance framework. We use daily returns for parameter estimation, include transaction costs in the portfolio performance and evaluate the portfolio strategies in different economic conditions, across four datasets.

Fundamentally, the research problem is to find out if it is possible, contrary to most previous research, to consistently outperform a simple strategy with optimization models. That is, whether the more complex models are valuable in practice or not. Furthermore, by examining the rebalancing of the portfolio and the length of estimation window more carefully, we aim to find out what their contribution to portfolio performance is. In a broader context, the thesis aims to shed light on the confused mindset about portfolio optimization that prevails in the field.

1.3.2 Boundaries of the thesis

An important adjustment to the objective of the thesis is that we focus on portfolio optimization models that are based on statistical data in the form of asset returns. No other source of information is included in the process of determining the optimal portfolio composition; the aim is to find out if it is possible to extract information solely from the return distributions to outperform the benchmark. For example, strategies that require the investor's subjective views about the performance of assets are omitted. Furthermore, the investable assets considered in the thesis are equities, or equity portfolios represented by stock market indexes.

Previous research has formulated an immense amount of different parameter estimation models and optimization strategies. All possible models cannot be tested in the scope of the thesis and therefore, only the most interesting and promising of those that fit into the boundaries described above are selected. Several well-known methods are omitted because of their documented poor performance: For example, an optimization method that utilizes Monte Carlo simulation, known as portfolio resampling, has had its performance critiqued (see, for example, Becker, Gürtler & Hibbeln 2009; Scherer 2006). Another well-known method not considered in the thesis is robust optimization which includes estimation error in the optimization process: It has not proven to be effective and is actually equivalent to shrinkage estimation, only involving more difficult computations (Scherer 2007).

In previous research, it has been stated that expected returns are most prone to estimation error: For example, DeMiguel et al. (2009b, 1936–1941) find out, by studying estimation error analytically, that a large part of the estimation error is due to the estimation of expected returns. Furthermore, Chopra and Ziemba (1993) show that errors in the expected returns can be over ten times more effective than errors in elements of the covariance matrix, depending on the risk aversion of the investor. Jagannathan and Ma (2003, 1654) find empirically that tangency portfolios do not outperform the minimum variance portfolio in terms of out-of-sample Sharpe ratio, implying that the estimation error in expected returns is large. Similarly, Jorion (1991, 724–725) reports that the minimum variance portfolio performs better than the tangency portfolio, suggesting that historical covariances can be useful in estimating future parameter values, but that the past means are not reliable for forecasting. Considering the above mentioned studies, the thesis focuses on estimating the covariance matrix and leaves out the estimation of expected returns; thus the optimization problem becomes one trying to obtain the minimum variance portfolio.

The thesis assumes the position of an institutional investor aiming to construct a portfolio that performs better than the benchmark in terms of relatively simple performance evaluation criteria. Thus, the thesis is not concerned with utility functions or selecting an optimal portfolio for a particular investor. Instead, the different models are compared against the benchmark using measures, such as the Sharpe ratio, that all investors are assumed to care about in a uniform manner.

1.3.3 Outline of the empirical study

In the empirical part of the thesis, we test three covariance matrix estimation models with different estimation window lengths and rebalancing intervals, evaluating the performance of minimum variance portfolios out-of-sample on real world stock market data, across four datasets. To achieve this, we collect and analyze data in the form of international stock market indexes that serve as investable assets. The statistical programming language R is utilized to build a program that performs the parameter estimation, invests wealth across the assets by solving the minimum variance portfolio, and automatically rebalances the portfolio throughout the investment period.

The tested parameter estimation models are: the classical sample covariance matrix model, the constant-correlation shrinkage model presented in Ledoit and Wolf (2004, 112–113) and the dynamic conditional correlation model that is proposed in Engle (2002, 341–342). Each model is tested with three different lengths of estimation window and different rebalancing frequencies, including a threshold rebalancing criterion to constrain portfolio turnover. The performance of the portfolios is evaluated by out-of-sample Sharpe ratios and turnover amounts, as in DeMiguel et al. (2009b).

Furthermore, we conduct an additional study on the optimal length of the estimation window in forecasting the future covariances of the asset returns. A mathematical estimation method is developed for assessing the difference between the forecast and realized values. The programming language R is used to calculate this difference throughout the datasets.

The thesis takes an interdisciplinary approach on empirical research, utilizing tools and research results from several different fields including finance, statistics, applied mathematics, and computer programming. This shows that modern quantitative finance leans on several branches of science and is becoming an increasingly demanding subject in the world of economics.

1.3.4 Structure of the thesis

The thesis proceeds as follows: The second chapter reviews the theoretical foundation for constructing optimal portfolios, first presenting Markowitz's mean-variance portfolio theory, after which the main issues of the theory are discussed. Attempts to remedy the problem of estimation error are considered; the shrinkage estimation model and autoregressive conditional heteroscedastic models are introduced. These models are utilized in the empirical part of the thesis. Finally, chapter two discusses portfolio rebalancing and the consequent transaction costs.

The third chapter presents the data for the empirical study and descriptive statistics for each dataset. The portfolio optimization process is described in detail and the parameter estimation models that are evaluated in the empirical study are presented. Furthermore, the performance evaluation methods used for comparison of the portfolios are introduced.

The fourth chapter presents the empirical results, first analyzing each estimation model's performance independently for each dataset and then proceeding to an overall view of the results across all datasets. Moreover, an additional study on the optimal length of the estimation window is conducted and the results examined. Finally, the fifth chapter concludes the thesis by presenting the main findings and conclusions. The programming code for the empirical study is provided in the appendix at the end of the thesis.

2 THEORETICAL FOUNDATION

2.1 Modern portfolio theory

2.1.1 Diversification and mean-variance efficient portfolios

An investor is faced with a decision of how to allocate wealth across available assets, such as equities. A formal solution to this problem was developed by Markowitz (1952) who introduced the trade-off between expected return and variance, and the idea about dividing wealth across assets based on their co-movements. In essence, the investor seeks to minimize the variance of the investment portfolio with any given expected return, or vice versa, maximize the expected return of the portfolio with any given variance. Roy (1952) presented much of the same idea in his paper, however, Markowitz is cited as the father of the so-called modern portfolio theory. (Elton & Gruber 1997, 1743–1745.)

The key proposition of modern portfolio theory is that the investor should select assets in the portfolio depending on their co-movement with other assets. Any asset's individual properties are not the only thing affecting the choice. For example, stocks of companies operating in the same industry are likely to move in the same direction, following the movements of the entire industry in general; this means that the stocks are positively correlated. Rather than investing the entire wealth in one particular stock or industry, the investor should allocate wealth across assets that exhibit low correlation. By diversifying the portfolio in this way, a lower volatility with equal or higher expected return can be achieved for the overall portfolio, than any of the individual assets comprising the portfolio exhibit. This leads to the concept of an efficient frontier of portfolios, depicted in Figure 1, which perform better than any other combinations of assets available. (Elton & Gruber 1997, 1744–1745; Fabozzi et al. 2006a, 20–21; Markowitz 1952.)

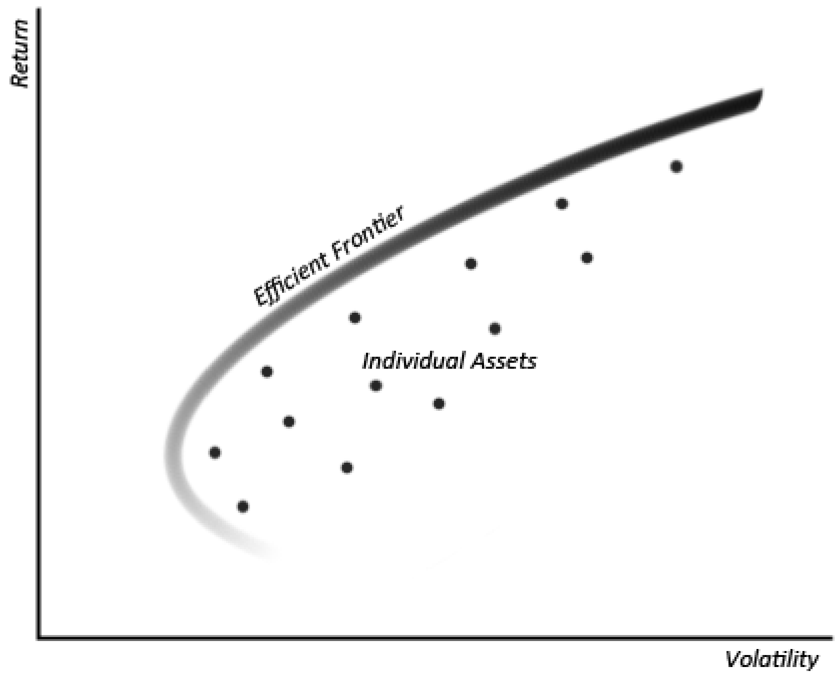

Figure 1 Efficient frontier (Fabozzi et al. 2006a, 20)

As can be seen in Figure 1, any of the individual assets have either lower return with equal volatility, or equal return with higher volatility, than any point on the efficient frontier. The efficient frontier is essentially a curve connecting portfolios with the lowest volatility for all given expected returns. The portfolios on the efficient frontier are called mean-variance efficient portfolios; they represent the best trade-off between risk and expected return. Portfolios below the efficient frontier are inefficient and portfolios above the frontier cannot be obtained. The portfolio on the efficient frontier with the least amount of volatility is called the minimum variance portfolio. (Fabozzi et al. 2006a, 20–21.)

As the number of assets in the portfolio increases, the portfolio variance diminishes, however, the variance cannot go all the way to zero because of the nonvanishing correlations of the assets (Fabozzi et al. 2006a, 17–18). Evans and Archer (1968) show that a portfolio containing ten or more assets is adequate to diversify away most of the unsystematic risk. On the other hand, there is a drawback to diversification: if the portfolio contains a large amount of assets, it will cause transaction costs to be high when the portfolio is rebalanced frequently (Rasmussen 2003, 97).

It should be noted that Markowitz's theory about portfolio selection is a normative theory, which means that it describes what investors should do, instead of what is actually taking place in the real world. Even in many asset management firms, portfolio selection is based on qualitative assessments of the asset managers instead of a quantitative optimization procedure. (Fabozzi et al. 2006a, 15–16.)

2.1.2 Portfolio optimization process

The mean-variance portfolio optimization requires that the expected returns, volatility, and correlation of the assets are known, or estimated from data. Once these parameters have been determined and possible constraints, such as short-sale restrictions, on portfolio choice have been set, one can compute the efficient frontier by applying an optimization algorithm. This process is illustrated in Figure 2. (Fabozzi et al. 2006a, 21.)

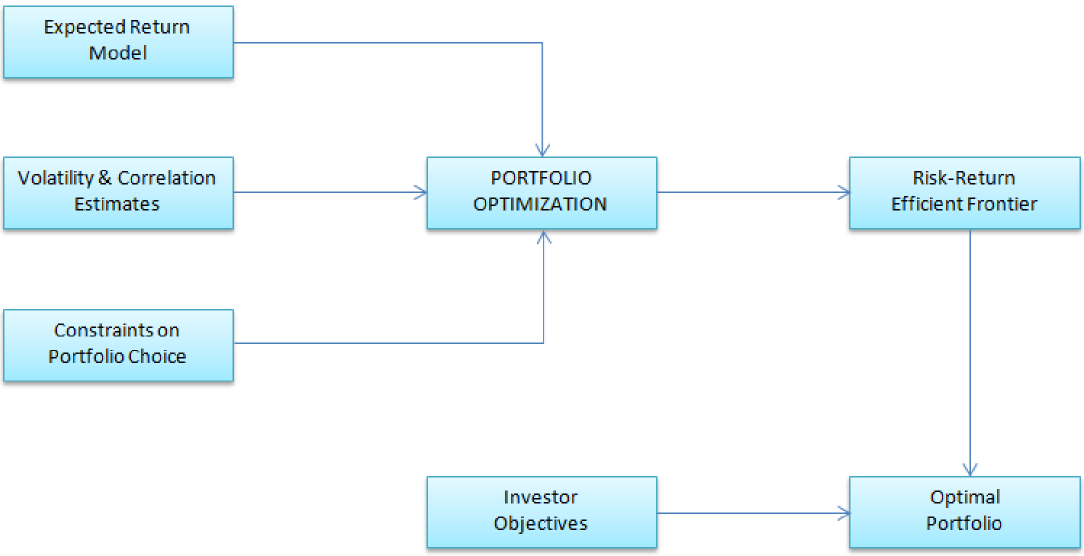

Figure 2 Portfolio optimization process (Fabozzi, Gupta & Markowitz 2002, 8)

In some cases the optimization problem has an analytical solution, but often, especially when using inequality constraints, the optimization problem has to be solved with numerical methods. Fortunately, these problems are relatively easy to solve with today's computer performance. Adding the investor's objectives in the process described in Figure 2, the optimal portfolio on the efficient frontier can be selected based on the investor's preferences. (Fabozzi et al. 2006a, 21, 24–25.)

The computation of efficient portfolios is described mathematically as follows. Let $\boldsymbol{w} = (w_1, w_2, \dots, w_N)$ denote a vector of portfolio weights for N assets. Each weight w_i describes a percentage of total wealth that is invested in asset i. Furthermore, it is assumed that the investor places their entire wealth in the portfolio of risky assets, so that

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

(Fabozzi et al. 2006a, 22.)

Next, let vector $\boldsymbol{\mu} = (\mu_1, \mu_2, \dots, \mu_N)$ denote the expected returns for the N assets. To describe the variances, covariances, and correlations of the assets, an $N \times N$ covariance matrix is defined as

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_{11} & \cdots & \sigma_{1N} \\ \vdots & & \vdots \\ \sigma_{N1} & \cdots & \sigma_{NN} \end{bmatrix}, \tag{2}$$

where σ_{ij} represents the covariance between asset i and asset j, defined as $\sigma_{ij} = \rho_{ij}\sigma_i\sigma_j$, where ρ_{ij} denotes the correlation between asset i and asset j ranging between minus one and one. In the case where $i = j$, σ_{ij} denotes the variance of asset i, or σ_i^2. These assumptions being made, the expected return and variance of the investment portfolio are defined as

$$\mu_p = \boldsymbol{w}\boldsymbol{\mu}' \tag{3}$$
$$\sigma_p^2 = \boldsymbol{w}\boldsymbol{\Sigma}\boldsymbol{w}', \tag{4}$$

respectively. (Fabozzi et al. 2006a, 22–23.)

By defining the portfolio weights, the investor determines the composition of the investment portfolio. As illustrated in Figure 2, the investor's objectives determine the optimal portfolio on the efficient frontier. For example, the investor may be interested in finding the portfolio that minimizes variance given a certain target return μ_0. In this case, the mean-variance optimization problem is of the form:

$$\min_{\boldsymbol{w}} \boldsymbol{w}\boldsymbol{\Sigma}\boldsymbol{w}' \tag{5}$$

subject to the constraints

$$\boldsymbol{w}\boldsymbol{\mu}' = \mu_0 \tag{6}$$
$$\boldsymbol{w}\boldsymbol{\iota} = 1, \tag{7}$$

where $\boldsymbol{\iota}$ denotes a vertical vector of ones. (Fabozzi et al. 2006a, 21–23.)

By choosing a different target return and solving the optimization problem, another portfolio on the efficient frontier will be obtained. If the optimization problem in Equation 5 is solved for all possible values of μ_0, all the portfolios on the efficient frontier are obtained (Fabozzi et al. 2006a, 24). It is possible to be indifferent regarding the target return, or equivalently, assume that all expected returns for the assets are equal as described in Jorion (1985, 270–271). In this case, the constraint of Equation 6 is removed from the optimization problem and the solution to this new problem is the minimum variance portfolio.

2.2 Main issues of the mean-variance framework

2.2.1 *Asset return distributions*

The financial data that are used in the mean-variance optimization are essentially series of asset prices from certain time periods; the asset prices themselves are not as important as the returns on those assets (Rasmussen 2003, 9). Returns are calculated from the prices of the assets, so that a one-period simple return is defined as

$$R_t = \frac{P_t}{P_{t-1}} - 1, \tag{8}$$

where P_t denotes the asset's price at time t, and P_{t-1} denotes the asset's price in the previous time period, or $t - 1$ (Tsay 2010, 3). Moreover, it is often practical to examine logarithmic asset returns and it appears to be the standard practice in financial research. The logarithmic return of an asset is defined as the natural logarithm of gross return, or equivalently, as the difference between the logarithms of the prices:

$$r_t = \ln(1 + R_t) = \ln(P_t/P_{t-1}) = p_t - p_{t-1}, \tag{9}$$

where $p_t = \ln(P_t)$ and $p_{t-1} = \ln(P_{t-1})$ (Tsay 2010, 5).

Markowitz's mean-variance theory is concerned with only the first two moments of the asset return distribution, namely the mean and standard deviation. Thus, the theory implicitly assumes that either returns are jointly normally distributed and the mean and variance alone describe the return distribution, or that the investor only cares about the mean and variance of the investment portfolio (Fabozzi et al. 2006a, 30).

However, it is well known that asset returns do not follow a normal distribution. Fama (1965, 45–55) was one of the first to study the empirical distributions of stock price changes. The study reports that each stock had more frequency in the center of the distributions than the normal distribution assumes, and also had longer tails than the normal distribution. The same phenomenon has been documented in stock price indexes: Behr and Poetter (2009, 108–112) study the distributions of daily European stock index returns and find that nearly all of the indexes exhibit negative skewness and notable excess kurtosis. Peiró (1994) suggests that Student's t-distribution fits best to describe stock returns. Figure 3 depicts a typical stock return distribution with excess kurtosis and the normal distribution that is generally used as an assumption for stock returns.

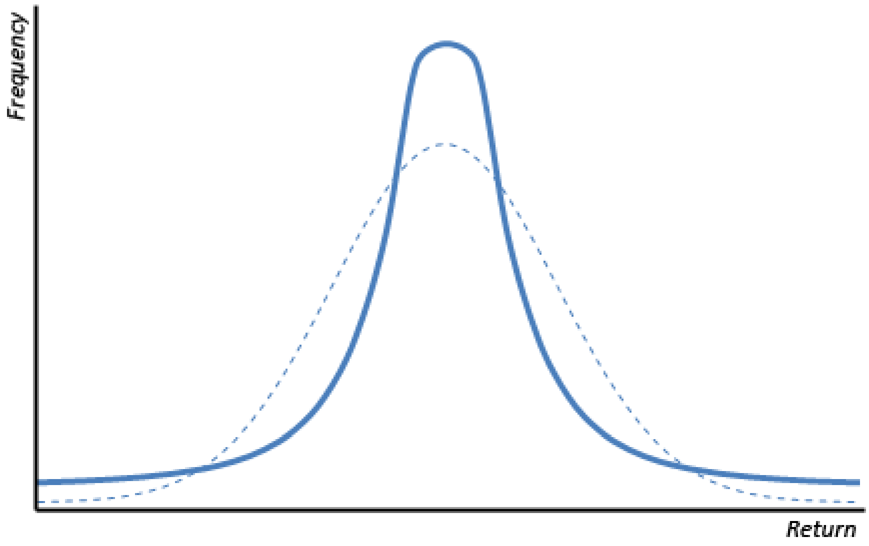

Figure 3 Typical stock return distribution compared to the normal distribution (Rasmussen 2003, 35)

From Figure 3, it can be seen how a typical distribution of stock returns (illustrated by the solid line) has longer tails and more mass in the center of the distribution. The normal distribution (illustrated by the dashed line), in contrast, lacks extreme values and is more evenly dispersed around the mean. Besides looking at graphs of return distributions, the normality of a distribution can be tested with a statistical method referred to as the Jarque–Bera test statistic:

$$JB = n\left(\frac{\hat{s}^2}{6} + \frac{\left(\hat{k} - 3\right)^2}{24}\right), \tag{10}$$

where n denotes the number of observations, $\hat{s}$ denotes the sample skewness, and $\hat{k}$ denotes the sample kurtosis. The higher the value of the test statistic, the more the sample distribution deviates from the normal distribution. (Jarque & Bera 1987, 165.)

Because return distributions often exhibit skewness and kurtosis, theoretically these moments could be taken into account in the portfolio optimization problem. Scott and Horvath (1980) state that a rational investor can be assumed to prefer high odd moments, or high mean and positive skewness, and low even moments, or low standard deviation and less heavy tails in the distribution. Interestingly, Fabozzi et al. (2006a, 131–133) show that a portfolio minimizing the standard deviation may in certain cases also minimize the skewness and maximize the kurtosis.

However, including higher moments in portfolio optimization is a risky procedure. Outlier observations' impact is larger because they are raised to third and fourth powers.

Heavy tails that are typical for asset return distributions further complicate the problems with higher moment estimation. (Fabozzi et al. 2006a, 141–142.)

2.2.2 Estimation error

The mean-variance optimization assumes that the input parameters are precise descriptions of reality. However, in practice, the future values of assets' mean returns and covariances cannot be known. Instead, the values need to be estimated, often from historical data. The problem is that these estimates are not accurate; they contain estimation error which is magnified in the optimization process and the resulting portfolio weights are often unstable and irrational. (Fabozzi et al. 2006a, 215–216; Michaud 1989.)

The sample mean is the most straightforward estimate of the expected return. The sample mean vector can be computed from asset return vectors as follows:

$$\hat{\boldsymbol{\mu}} = \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{r}_t, \tag{11}$$

where T denotes the number of observations, and $\boldsymbol{r}_t$ denotes a vector of returns at time t (Tsay 2010, 19). Moreover, the sample covariance matrix is computed as follows:

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{T-1}\sum_{t=1}^{T} (\boldsymbol{r}_t - \hat{\boldsymbol{\mu}})(\boldsymbol{r}_t - \hat{\boldsymbol{\mu}})' \tag{12}$$

(Tsay 2010, 19).

Because the distributions of financial return series exhibit heavy tails, the sample mean is not a good estimator of the expected return. For distributions that do not have heavy tails, the sample mean's performance can be improved by increasing the sample size. However, this is not the case with typical asset return distributions. In addition, nonstationarity often present in financial time series leads to the mean being a poor forecast of future returns. (Fabozzi et al. 2006a, 224.)

Estimating the sample covariance matrix from historical data is an easy method but it does not take into account how volatility evolves over time. Volatility estimated from historical data proposes a problem of an extreme observation, or a jump, affecting the value of the estimate long after the actual event. When the estimation window rolls past the event, there is a sudden change in the estimate when the extreme observation is left out. (Rasmussen 2003, 147–148.)

In general, the optimal allocation is not much use if the optimization inputs, that is the expected return and covariance matrix forecasts, are not accurate. Estimation error in the inputs tends to be exaggerated in the optimization process: assets that feature high expected return and low variance will be overweighted, and vice versa, assets with low expected return and high variance will be underweighted. These assets are also the ones that most likely feature a lot of estimation error. (Michaud 1989, 33–34.)

Furthermore, estimation error in expected returns can be over ten times more effective in portfolio optimization than estimation error in the covariances (Chopra & Ziemba 1993). Because of estimation error, mean-variance optimized portfolios usually perform worse than a simple equal-weight portfolio, as reported in DeMiguel et al. (2009b).

The length of historical return series used for parameter estimation appears to be fundamental. Besides estimation error, there is another problem when estimating model parameters from historical data: nonstationarity of the parameters. This means that the parameters of the model change over time. In other words, when estimating the parameters from a long return series, the returns of an asset recorded a long time ago might not tell a lot about today's return properties. On the other hand, if using only a short amount of data, estimation error increases. Thus, there is a trade-off between estimation error and stationarity when choosing the length of the estimation window. (Broadie 1993, 21–22.)

2.3 Providing better estimates

Parameter estimation models in finance are always approximate models. They assume probability distributions that are idealizations of reality, and thus, are never fully accurate. (Fabozzi et al. 2006a, 193.) However, estimation error in the parameters can be reduced by applying statistical methods.

2.3.1 Shrinkage estimation

Shrinkage estimation is a statistical method that can be used to provide better estimates for parameters such as expected returns. In essence, the idea of the approach is to average the sample-based estimate toward a more general value. The so-called James–Stein shrinkage estimator for expected returns is defined as:

$$\hat{\boldsymbol{\mu}}_{JS} = (1-\delta)\hat{\boldsymbol{\mu}} + \delta\mu_0\boldsymbol{\iota}, \tag{13}$$

where δ denotes a shrinkage weight term that is determined separately, $\hat{\boldsymbol{\mu}}$ denotes the vector of sample means, μ_0 denotes the shrinkage target, and $\boldsymbol{\iota}$ denotes a vector of ones. (Jorion 1986, 279–284.)

The shrinkage method can also be applied to the estimation of the covariance matrix. The idea is to shrink the sample covariance matrix toward an estimator that is more structured. A structured estimator has little estimation error but can be very biased. The idea is to find a compromise between the sample covariance matrix and the structured estimator, or the shrinkage target. (Ledoit & Wolf 2004, 112–113.)

Ledoit and Wolf (2004, 112–113) propose the constant-correlation covariance matrix as the shrinkage target. Their shrinkage estimator for the covariance matrix is defined as:

$$\boldsymbol{\Sigma}_{LW} = (1-\delta)\widehat{\boldsymbol{\Sigma}} + \delta\boldsymbol{\Sigma}_{CC}, \tag{14}$$

where $\widehat{\boldsymbol{\Sigma}}$ denotes the sample covariance matrix, $\boldsymbol{\Sigma}_{CC}$ denotes the shrinkage target which is the constant-correlation covariance matrix, and δ denotes the shrinkage weight between zero and one, for which an optimal value can be calculated. According to Ledoit and Wolf, when the two estimators are combined correctly, a more accurate estimate than either of the extremes can be achieved.

The constant correlation for the shrinkage target can be calculated as follows:

$$\hat{\rho} = \frac{2}{(N-1)N}\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\hat{\rho}_{ij}, \tag{15}$$

where N denotes the number of available assets, and $\hat{\rho}_{ij}$ denotes the sample correlation between asset i and asset j. A constant-correlation matrix $\boldsymbol{R}_{CC}$ is then constructed, where the constant correlation $\hat{\rho}$ is in every element of the matrix, except on the diagonal which is formed of ones. Finally, the constant-correlation covariance matrix is formed as follows:

$$\boldsymbol{\Sigma}_{CC} = \boldsymbol{D}\boldsymbol{R}_{CC}\boldsymbol{D}', \tag{16}$$

where $\boldsymbol{D}$ denotes a diagonal matrix of the standard deviations of the returns. (Fabozzi et al. 2006a, 277–278; Ledoit & Wolf 2004, 115–116.)

Ledoit and Wolf (2004, 113–115) show that their constant-correlation shrinkage estimator outperforms the sample covariance matrix, and performs as well as an index shrinkage estimator for the covariance matrix presented in Ledoit and Wolf (2003). This other shrinkage model, in turn, outperformed other covariance matrix estimators when minimum variance portfolios were evaluated out-of-sample (Ledoit & Wolf 2003, 613–

620). Thus, the constant-correlation shrinkage estimator can be considered as one of the best shrinkage estimators available for the covariance matrix. For this reason, we employ it in the empirical part of the thesis.

2.3.2 Autoregressive conditional heteroscedasticity models

As previously discussed, we are interested in the future values of parameters instead of their past values obtained from historical data. Volatility exhibits characteristics that make it possible to attempt forecasting its future level. One of the properties of volatility is that it tends to cluster in asset returns: high or low levels of volatility can persist for certain time periods. Furthermore, the evolving of volatility over time usually happens in a continuous manner, without jumps. Conditional heteroscedasticity modeling is about forming a dynamic equation that writes the evolution of conditional variance in the form of a time series model. In essence, the models attempt to capture the serial dependence that return series exhibit. (Tsay 2010, 110–113.)

To find out if returns are serially dependent and if volatility modeling can be used to take advantage of this feature, a test for autocorrelation in the squared returns can be performed (Tsay 2010, 111–112). The sample autocorrelation for lag k is defined as:

$$\hat{\varrho}(k) = \frac{\hat{y}(k)}{\hat{y}(0)}, \tag{17}$$

where

$$\hat{y}(k) = \frac{1}{T}\sum_{t=1}^{T-k}\big((r_t - \mu_T)(r_{t+k} + \mu_T)\big), \tag{18}$$

where T denotes the amount of observations, r_t denotes the return at time t, and μ_T denotes the mean of returns over the sample (Fabozzi et al. 2006a, 326). The returns and the sample mean of returns in Equation 18 can be replaced with the corresponding values of squared returns to test for serial dependence. In addition, to test for overall autocorrelation across several lags, the so-called Ljung–Box statistic is defined as:

$$Q(m) = T(T+2)\sum_{k=1}^{m}\frac{\hat{\varrho}^2(k)}{T-k}, \tag{19}$$

where m denotes the number of lags included in the statistic (Ljung & Box 1978; Tsay 2010, 32–33).

Engle (1982) presented the autoregressive conditional heteroscedastic (ARCH) model that can capture volatility clustering. The idea behind the ARCH model is that parameter a_t, called the shock term, is serially dependent and can be explained by a function of its lagged values:

$$a_t = \sigma_t \varepsilon_t, \qquad \sigma_t^2 = \omega + \alpha_1 a_{t-1}^2 + \cdots + \alpha_m a_{t-m}^2, \tag{20}$$

where ε_t represents a sequence of independent and identically distributed random variables that have a zero mean and variance of one, $\omega > 0$ and $\alpha_i \geq 0$ denote parameters that need to be calibrated for the data. The α_i are parameters that give weights to the past shock terms. Under the ARCH model, a large shock is likely to be followed by another large shock, featuring a behavior similar to volatility clustering in asset return series. (Tsay 2010, 113, 115–116.)

The ARCH model usually requires a lot of parameters to model a volatility process accurately and it can be troublesome to apply the model in practice (Tsay 2010, 131). Bollerslev (1986, 308–309) introduced the generalized ARCH (GARCH) model for more flexible lag structure and longer memory. A GARCH(q, p) model can be defined as follows:

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i a_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2 \tag{21}$$

where $\omega > 0$, $\alpha_i \geq 0$, $\beta_i \geq 0$, and the sum of α and β parameters is smaller than one; if $p = 0$, the model is reduced to an ARCH(q) model (Tsay 2010, 132). The GARCH process includes lagged conditional variances, whereas the ARCH model has the conditional variance determined only by past sample variances (Bollerslev 1986, 309). Furthermore, numerous extensions and modifications to the GARCH model have been developed (see, for example, Tsay 2010, 140–150).

To use the ARCH model in portfolio context, we describe the return series as a process: $r_t = \mu_t + a_t$, where μ_t denotes the expected return (Tsay 2010, 111). For simplicity, it can be assumed that r_t is normally distributed with zero mean, so that $r_t = a_t$. The parameters of ARCH and GARCH models can be estimated with maximum likelihood methods (Tsay 2010, 120–122).

2.3.3 Multivariate volatility and correlation modeling

The univariate volatility modeling can be generalized to the multivariate case, to model the relationships between volatility processes of several return series. In essence, multivariate volatility modeling attempts to model the time evolution of the covariance matrix $\boldsymbol{\Sigma}_t$. The large amount of parameters that are needed to estimate the entire covariance matrix can often make these models impractical. (Tsay 2010, 505–506.)

Bollerslev (1990) proposes a multivariate model that has time-varying conditional variances but the conditional correlations are constant. This way the estimation of the conditional covariance matrix is noticeably simplified. The time-varying conditional covariance matrix $\boldsymbol{\Sigma}_t$ is partitioned in the following manner:

$$\boldsymbol{\Sigma}_t = \boldsymbol{D}_t \boldsymbol{R} \boldsymbol{D}_t, \tag{22}$$

where $\mathbf{D}_t$ denotes an $N \times N$ diagonal matrix that has the conditional volatilities $\sigma_{1,t}, \dots, \sigma_{N,t}$ on its diagonal, and $\boldsymbol{R}$ denotes the $N \times N$ time-invariant correlation matrix. A univariate GARCH model is assumed for the conditional variances. (Bollerslev 1990, 498–501.)

Engle (2002) proposes a dynamic conditional correlation (DCC) model that parameterizes conditional correlations and is estimated in two steps, first estimating univariate GARCH models for each asset, and then in the second step, standardized residuals from the first step are used in the estimation of the time-varying correlation matrix. This way, the amount of parameters that need to be estimated does not depend on the amount of correlated series, which is not the case with multivariate GARCH models in general. Thus, it is possible to estimate the DCC model for even large correlation matrixes. (Engle 2002, 339; Engle & Sheppard 2001, 2.)

The DCC model can be seen to generalize the constant conditional correlation model presented in Bollerslev (1990): in the DCC model, the correlation matrix $\boldsymbol{R}_t$ is time-varying, and thus the conditional covariance matrix is defined as:

$$\boldsymbol{\Sigma}_t = \boldsymbol{D}_t \boldsymbol{R}_t \boldsymbol{D}_t \tag{23}$$

(Engle 2002, 341). The idea behind the dynamic correlation model is that by considering

$$\sigma_{i,t}^2 = E_{t-1}\left(r_{i,t}^2\right), \qquad r_{i,t} = \sigma_{i,t}\varepsilon_{i,t}, \tag{24}$$

where $\varepsilon_{i,t}$ denotes a standardized disturbance with mean zero and variance of one for each asset i, the conditional correlation between returns r_i and r_j that both have mean zero can be expressed as the conditional covariance between the standardized disturbances:

$$\rho_{ij,t} = \frac{E_{t-1}\left(r_{i,t}r_{j,t}\right)}{\sqrt{E_{t-1}\left(r_{i,t}^2\right)E_{t-1}\left(r_{j,t}^2\right)}} = E_{t-1}\left(\varepsilon_{i,t}\varepsilon_{j,t}\right). \tag{25}$$

(Engle 2002, 339–340.)

The DCC model assumes returns that are conditionally multivariate normal with mean zero and covariance matrix $\mathbf{\Sigma}_t$. The dynamic correlation structure is defined by the DCC(M, N)model:

$$\boldsymbol{Q}_t = \left(1 - \sum_{m=1}^{M} \alpha_m - \sum_{n=1}^{N} \beta_n\right)\overline{\boldsymbol{Q}} + \sum_{m=1}^{M} \alpha_m(\boldsymbol{\varepsilon}_{t-m}\boldsymbol{\varepsilon}'_{t-m}) + \sum_{n=1}^{N} \beta_n \boldsymbol{Q}_{t-n} \tag{26}$$

$$\boldsymbol{R}_t = \boldsymbol{Q}_t^{*-1}\boldsymbol{Q}_t\boldsymbol{Q}_t^{*-1}, \tag{27}$$

where $\overline{\boldsymbol{Q}}$ denotes the unconditional covariance matrix of the standardized residuals from the estimation of the individual volatilities in the first step, $\boldsymbol{\varepsilon}_t$ denotes a vector of the standardized residuals, and $\boldsymbol{Q}_t^*$ denotes a diagonal matrix that is comprised of the square roots of the diagonal elements of $\boldsymbol{Q}_t$. Furthermore, $\alpha_m \geq 0$ and $\beta_n \geq 0$ denote parameters that are estimated with maximum likelihood methods. (Engle & Sheppard 2001, 4–10.)

There is a vast number of multivariate GARCH models with different features, such as more accurate performance or easier estimation (see, for example, Tsay 2010, 510–537). The DCC-GARCH model appears to be one of the most functional multivariate volatility models to implement in practice. Therefore, we utilize it in the empirical part of the thesis.

2.4 Portfolio rebalancing and transaction costs

2.4.1 From single-period to multi-period investing

The mean-variance framework is essentially a one-period model, representing myopic, or short-sighted, behavior. The mean-variance framework assumes that a rational investor, at time t, makes a decision on an investment portfolio to hold for a time period of Δt. The decision is based only on the gains and losses that will be realized at time $t + \Delta t$ without taking into account the performance of the portfolio either during or after the investment period Δt. (Fabozzi et al. 2006a, 19.)

The myopic approach can be considered a limited view of the real world situation; the investor's problem is often multi-period. However, there is evidence that the multi-period problem can be described by, under reasonable assumptions, dividing it into a sequence of single-period problems (see, for example, Fama 1970). (Elton & Gruber 1997, 1745–1746.)

The portfolio can be rebalanced periodically in response to the multi-period problem. The need to rebalance the portfolio can arise due to new information that becomes available as time passes, or simply because of changing investment preferences. Naturally, changes in the economic environment affect asset return distributions. The statistical properties of the returns can be re-estimated upon rebalancing to take new information into account. In addition, asset prices fluctuate as time passes, therefore changing the portfolio composition. The original target composition can be restored by rebalancing the portfolio. (Fabozzi et al. 2006a, 88–91.)

The frequency for rebalancing a portfolio depends on many factors, such as the investment strategy and transaction costs. It is important to find a balance between the deviation from the desired portfolio composition and transaction costs incurred by rebalancing: if the portfolio is kept as close to the optimal weights as possible, it requires very frequent updating. However, frequent updating increases transaction costs which decrease the return of the portfolio. (Fabozzi et al. 2006a, 91.)

As described in Fabozzi et al. (2006a, 91–92), the portfolio can be rebalanced with a certain calendar frequency, such as weekly, monthly, or quarterly. Another approach described is threshold rebalancing, where the portfolio is rebalanced back to the optimal allocation when its weights are further than a specific range from the optimal weights. In addition, there are several more complex criteria for rebalancing, such as the optimal rebalancing considered in Sun, Fan, Chen, Schouwenaars and Albota (2006). However, this type of dynamic programming application is mathematically very complex and difficult to apply in practice.

2.4.2 Transaction costs

When securities are bought or sold in the market, transaction costs are incurred. Transaction costs can be categorized into explicit and implicit costs. Explicit costs include brokerage commissions and fees, taxes, and bid-ask spreads. They are observable and are known beforehand. Implicit costs, such as market impact costs and price movement risk, are not directly observable and unknown in advance. Many of the explicit transaction costs are fixed costs that do not depend on factors such as trade size or market conditions. On the other hand, implicit costs are variable: trades have an impact on the market price, especially large trade volumes. These so-called market impact costs are difficult to measure. (Fabozzi et al. 2006a, 51–65.)

Domowitz, Glen and Madhavan (2001, 223–231) examine trading costs around the world during the latter half of the 1990s. They report that the average one-way trading cost, measured in percentage of trade value, is 71.3 basis points, of which 46.0 points are explicit costs. In general, they find that emerging markets feature higher trading costs than those observed in developed markets. Furthermore, Domowitz et al. document a lot of variation between the explicit and implicit costs of different countries; the USA features the lowest explicit costs of only 8.3 basis points. Total costs for all countries on average are in a declining trend.

Transaction costs appear to have continued declining in the 21st century: Byrne (2010) reports the results of a recent global transaction cost survey showing that the average cost of trading has been falling. In the survey results, the overall average equity trading cost is 19.63 basis points in the USA. The lowest costs in the world are reported in Japan and Sweden, both at 18.34 basis points. Furthermore, the average cost globally is 38.02 basis points.

Considering the complexity underlying the implicit costs and the amount of variation in total transaction costs in general, trying to compute exact values of transaction costs for an investment portfolio would be a very burdensome task. For the empirical study taking transaction costs into account, it is reasonable to assume a fixed transaction cost per trade based on the recent survey and research results.

2.4.3 Constraining portfolio weights

Transaction costs are traditionally not included in the mean-variance optimization procedure, however, they have the potential to cut down portfolio returns as a consequence of frequent rebalancing (Fabozzi et al. 2006a, 74). Including constraints in portfolio optimization is one way of reducing portfolio turnover, thereby reducing transaction costs.

Typical constraints in the portfolio optimization process include a short-sale constraint, since asset managers can be prohibited from selling stocks short. Furthermore, to prevent irrational portfolio weights and large concentrations in particular assets, a range can be determined for each asset weight w_i by imposing the following constraint:

$$L_i \leq w_i \leq U_i, \tag{28}$$

where L_i and U_i denote the lower and upper bounds for asset i, respectively. (Fabozzi et al. 2006a, 31, 100–101.)

Jagannathan and Ma (2003, 1652–1654) state that short-sale constraints produce the same effect as reducing the estimated covariances of the assets. Upper bounds on the weights of the assets, on the other hand, are equivalent to increasing their covariances. Jagannathan and Ma propose that short-sale constraints improve the performance of the sample covariance matrix portfolio to the level of shrinkage estimator portfolios.

3 DATA AND METHODOLOGY

In an attempt to answer the research problem and to shed light on the topic in general, the thesis studies the performance of minimum variance portfolios and evaluates them against a benchmark strategy. An empirical study is conducted where the investor's wealth is divided across assets according to the optimization process with different parameter estimation models, rebalancing the portfolios throughout the investment period. The body of the empirical study follows previous research, such as DeMiguel et al. (2009b). This type of approach requires a sufficient amount of data, a computer algorithm that runs the portfolio optimization and investment process, as well as performance evaluation criteria to compare the portfolios against the benchmark.

3.1 Data for the empirical study

To study the performance of portfolio strategies, a group of assets for investing is selected. The investor's wealth will be allocated across these assets according to the portfolio optimization process. The group of assets must be large enough to provide benefits of diversification but at the same time restricted in size for the optimization to be computationally workable. Daily prices for the assets during a certain period of time are then gathered. It needs to be emphasized that we utilize daily information instead of monthly data that is used in most previous research.

For the empirical study, a total of 23 MSCI standard country equity indexes are selected as investable assets. These are daily price indexes that are formed from the countries' equity market performance. In essence, the indexes can be thought of as portfolios of equities in the countries' stock markets. All the indexes are so-called total return indexes, meaning that possible dividends are reinvested in the index and are thus included in the performance of the index. (MSCI Index Calculation Methodologies 2010, 5–9, 21.)

Furthermore, all indexes are quoted in euro currency, which sets country indexes outside the European Economic and Monetary Union (EMU) exposed to currency exchange rate changes. Since investors diversifying internationally are often exposed to currency risk in reality, it is accepted as part of the investment process. The actual data collected from Datastream are daily closing prices of the selected MSCI equity indexes during the period of June 2001 to May 2010. The data include only closing prices quoted on weekdays, weekends are excluded as stock markets are closed.

For parameter estimation in the optimization process, we need to calculate daily returns from the quoted prices of the assets. Hence, we compute daily logarithmic returns for each equity index according to Equation 9. Furthermore, the logarithmic return data is trimmed by removing days when more than half of the indexes exhibit zero returns. This is usually the case when the markets have been closed during public holidays. The trimming is done to prevent possible problems with the estimation of the covariance matrix and also to make the data less misleading for the parameter estimation.

To thoroughly evaluate the performance of an optimization model, it needs to be tested in different circumstances: during periods of high and low volatility, expansionary and recessionary times, and when the investable assets exhibit different levels of correlation (Fabozzi et al. 2006a, 424). To take into account these aspects, the assets are divided into two groups: one with high correlations and one with relatively low correlations. Furthermore, the total of nine years of daily return data is split into two four-year periods plus one year for startup parameter estimation, so as to examine portfolio performance separately in different market conditions. This forms a total of four datasets for the empirical study listed in Table 1.

Table 1 Datasets for the empirical study

The details of each dataset are reported in this table. The investment period is the time frame that is used for trading and portfolio performance evaluation, and the number of trading days is the total amount of days in this time frame. The average pairwise correlation is the arithmetic mean of the correlations between each asset pair. Correlation range denotes the lowest and highest of these pairwise correlations. Moreover, each dataset is assumed a fixed level of trading costs that are listed in basis points (bps).

Dataset	No. of assets	Investment period	No. of trading days	Avg. pairwise correlation	Correlation range	Assumed trading costs
EMU1	11	6/2002 – 5/2006	1 023	0.58	0.32–0.95	20 bps
EMU2	11	6/2006 – 5/2010	1 019	0.72	0.52–0.93	20 bps
BEM1	12	6/2002 – 5/2006	1 035	0.29	0.05–0.60	40 bps
BEM2	12	6/2006 – 5/2010	1 031	0.40	0.14–0.79	40 bps

The dataset called EMU1 consists of the return series of 11 EMU country indexes: Austria, Belgium, Finland, France, Germany, Greece, Ireland, Italy, Netherlands, Portugal, and Spain. The investment period, i.e. the time frame in which the portfolio performance is examined, for the dataset EMU1 spans from the beginning of June 2002 to the end of May 2006. One of the advantages of studying the EMU countries grouped in one dataset is that all the countries share the same currency. This presents an opportunity to examine portfolio returns without currency exchange rates affecting them. The dataset EMU2 consists of the same countries as the dataset EMU1 but the investment period is the latter four-year portion of the return data, spanning from the beginning of June 2006 to the end of May 2010.

The dataset BEM1 consists of big emerging market countries: Brazil, China, Egypt, India, Indonesia, Mexico, Philippines, Poland, Russia, South Africa, South Korea, and

Turkey. The investment period for this dataset is the same as for the dataset EMU1: from the beginning of June 2002 to the end of May 2006. The dataset BEM2 consists of the same countries as the dataset BEM1, with the time frame for trading spanning from the beginning of June 2006 to the end of May 2010. The datasets EMU1 and BEM1 are parallel in the sense that their investment periods are the same. This applies to the relationship of the datasets EMU2 and BEM2 as well.

The data not included in either of the two investment periods, spanning from June 2001 to May 2002, is used as startup data for parameter estimation for the datasets EMU1 and BEM1. For this reason, it is not included in the actual datasets that are used for out-of-sample performance evaluation. For the datasets EMU2 and BEM2, the startup data is comprised of the time period spanning from June 2005 to May 2006, which overlaps with the datasets EMU1 and BEM1; the last year in these datasets is used as startup data for the latter datasets.

As shown in Table 1, the four datasets have different magnitudes of average pairwise correlations: the datasets EMU1 and EMU2, comprised of the 11 EMU countries, exhibit relatively high correlations between the returns, averaging at 0.58 and 0.72, respectively. The datasets BEM1 and BEM2 feature lower average correlations: 0.29 and 0.40, respectively. This presents an opportunity to analyze portfolio performance under different correlation conditions.

The empirical study includes trading costs as part of the investment process: every time an asset is bought or sold, a fixed percentage of the traded amount is deducted from the investor's wealth. The datasets EMU1 and EMU2 are assumed to entail a fixed trading cost of 20 basis points of the trading volume. The datasets BEM1 and BEM2 are assumed a trading cost of 40 basis points of the trading volume. These rates are selected based on recent studies and information about real world trading costs (see, for example, Byrne 2010; Domowitz et al. 2001). Setting two different trading cost categories presents an opportunity to examine how different magnitudes of trading costs affect the performance of portfolios.

Compared to previous studies, such as DeMiguel et al. (2009b), that use monthly return data, we deal with significantly more information-rich data. As is evident from Table 1, each dataset consists of over thousand observations. Monthly data would require an investment period of nearly 100 years to include as many observations as are present here in the daily return data.

3.2 Descriptive statistics

3.2.1 *Return distributions*

In order to get a better description of the assets available for investment, the mean, standard deviation, skewness, and kurtosis of the logarithmic returns for each country in the four datasets are determined. In addition, the Jarque–Bera test statistic is calculated for each country's returns according to Equation 10 to summarize the deviation from the normal distribution. Table 2 presents this information for the dataset EMU1.

Table 2 Descriptive statistics for the dataset EMU1

This table presents the mean, standard deviation, skewness, and kurtosis of the logarithmic return distributions for each country in the dataset EMU1. The departure from the normal distribution is summarized by the Jarque–Bera test statistic and the corresponding p-value.

	Mean (%)	St. Dev.	Skewness	Kurtosis	JB-test	p-value
Austria	0.0975	0.0094	-0.522	6.429	547.6	< 0.001
Belgium	0.0391	0.0134	0.281	10.399	2347.2	< 0.001
Finland	0.0309	0.0194	-0.390	7.857	1031.3	< 0.001
France	0.0222	0.0144	-0.021	6.559	540.0	< 0.001
Germany	0.0193	0.0164	-0.004	5.982	379.0	< 0.001
Greece	0.0615	0.0114	0.001	4.608	110.3	< 0.001
Ireland	0.0335	0.0111	-0.677	9.306	1773.1	< 0.001
Italy	0.0375	0.0111	-0.178	6.113	418.6	< 0.001
Netherlands	0.0033	0.0154	-0.053	6.944	663.7	< 0.001
Portugal	0.0328	0.0090	-0.264	6.278	470.0	< 0.001
Spain	0.0434	0.0130	0.162	6.782	614.2	< 0.001
Average	*0.0383*	*0.0131*	*-0.151*	*7.024*		

From the skewness and kurtosis columns in Table 2, it is evident that the return distributions for all country indexes deviate from the normal distribution: most countries exhibit negative skewness, and all countries have heavy-tailed distributions since the kurtosis values are larger than three which is the standard level of kurtosis for the normal distribution. In addition, the Jarque–Bera test statistics for all countries have very high values, ranging from the lowest of 110.3 for Greece to the highest of 2347.2 for Belgium. This further confirms that the logarithmic return series in the dataset are not normally distributed. The corresponding p-values show very strong statistical significance for the test statistics for all 11 countries in the dataset EMU1.

Furthermore, the means and standard deviations among the countries vary considerably: Austria has the highest daily mean return of 0.0975%, whereas Netherlands exhibits a return of only 0.0033%. It is important to note that while Austria has the highest mean return, it also has the second lowest standard deviation of 0.0094. This makes it a very attractive asset for investment and the minimum variance optimization is likely to place a large weight on it. Vice versa, Germany has a relatively low return of 0.0193% and exhibits the second highest standard deviation of 0.0164, which makes it an undesirable asset for investment. These aforementioned contrasts may heighten the utility of the portfolio optimization process.

The return distribution statistics for the dataset EMU2 are quite different despite the fact that returns for the same countries are examined. The investment period for the dataset EMU2 includes the recent financial crisis and this is apparent in the statistics in Table 3.

Table 3 Descriptive statistics for the dataset EMU2

This table presents the mean, standard deviation, skewness, and kurtosis of the logarithmic return distributions for each country in the dataset EMU2. The departure from the normal distribution is summarized by the Jarque–Bera test statistic and the corresponding p-value.

	Mean (%)	St. Dev.	Skewness	Kurtosis	JB-test	p-value
Austria	-0.0677	0.0229	-0.036	7.488	855.3	< 0.001
Belgium	-0.0514	0.0179	-0.926	12.877	4287.9	< 0.001
Finland	-0.0244	0.0199	0.043	6.305	464.2	< 0.001
France	-0.0157	0.0171	0.164	9.421	1755.1	< 0.001
Germany	0.0000	0.0165	0.213	9.687	1906.1	< 0.001
Greece	-0.0874	0.0220	-0.037	6.056	396.7	< 0.001
Ireland	-0.1205	0.0252	-0.388	8.861	1484.2	< 0.001
Italy	-0.0444	0.0172	0.251	10.034	2111.3	< 0.001
Netherlands	-0.0024	0.0165	-0.078	9.052	1556.3	< 0.001
Portugal	-0.0194	0.0143	0.343	15.175	6313.7	< 0.001
Spain	-0.0001	0.0181	0.343	11.674	3214.4	< 0.001
Average	*-0.0394*	*0.0189*	*-0.010*	*9.694*		

The mean return is lower than or equal to zero for every single country in the dataset EMU2. The standard deviations are larger for all countries compared to the dataset EMU1. Furthermore, the average kurtosis is much higher for the dataset EMU2 than for the dataset EMU1, meaning that the return distributions have heavier tails in the latter investment period. The Jarque–Bera test statistics are much larger in general compared to the dataset EMU1, most notably for Portugal with a score of 6313.7 which is a sign of very strong non-normality. The corresponding p-values indicate very strong statistical significance for the test statistics.

Similarly to the dataset EMU1, the dataset EMU2 includes countries that are significantly worse investments than other countries: For example, Ireland has the lowest mean return of -0.1205% and the highest standard deviation of 0.0252, making it an undesirable asset for investment. Vice versa, Germany has the highest mean return and the second lowest standard deviation, making it stand out as a good investment. These clear differences are likely to be detected by the optimization process, possibly giving the minimum variance portfolios an advantage over the simpler benchmark strategy.

Table 4 presents the statistics for the return distributions in the dataset BEM1. On average, the countries exhibit much higher mean returns than the countries in the parallel dataset EMU1. The skewness and kurtosis values demonstrate that all countries' return distributions depart from the normal distribution. The Jarque–Bera test statistics vary a lot between the countries: for example, Poland has a low score of 33.8 and Indonesia has a much higher score of 6734.3. Nevertheless, the p-values denote very strong statistical significance for the test statistics for all the countries.

Table 4 Descriptive statistics for the dataset BEM1

This table presents the mean, standard deviation, skewness, and kurtosis of the logarithmic return distributions for each country in the dataset BEM1. The departure from the normal distribution is summarized by the Jarque–Bera test statistic and the corresponding p-value.

	Mean (%)	St. Dev.	Skewness	Kurtosis	JB-test	p-value
Brazil	0.0938	0.0213	-0.089	5.528	277.0	< 0.001
China	0.0510	0.0151	-0.262	4.290	83.6	< 0.001
Egypt	0.1915	0.0190	0.139	6.934	670.6	< 0.001
India	0.0914	0.0147	-1.029	11.424	3242.5	< 0.001
Indonesia	0.0795	0.0197	-1.322	15.213	6734.3	< 0.001
Mexico	0.0543	0.0148	-0.244	4.921	169.5	< 0.001
Philippines	0.0195	0.0141	0.083	4.243	67.8	< 0.001
Poland	0.0662	0.0156	-0.259	3.718	33.8	< 0.001
Russia	0.0948	0.0208	-0.353	6.439	531.5	< 0.001
South Africa	0.0644	0.0138	-0.472	4.881	191.0	< 0.001
South Korea	0.0495	0.0181	-0.427	4.324	107.1	< 0.001
Turkey	0.0862	0.0280	-0.405	8.892	1525.4	< 0.001
Average	*0.0785*	*0.0179*	*-0.387*	*6.734*		

The clear separation of good and bad countries for investment that was evident in the datasets EMU1 and EMU2 is not observable in the dataset BEM1. In fact, the country with the highest mean return, Egypt with 0.1915%, also exhibits a relatively high standard deviation. Similarly, Russia with the second highest mean return of 0.0948% exhibits the third highest standard deviation of 0.0208. Furthermore, Philippines has the second lowest standard deviation of 0.0141 and also has the lowest mean return of 0.0195%. The minimum variance portfolio is likely to underweight the aforementioned countries with high mean returns and overweight the latter country with a low mean return, which may cause the overall portfolio return to be relatively low.

Table 5 exhibits the statistics for the return distributions in the dataset BEM2. Nearly all the countries have positive mean returns, which is in contrast to the parallel dataset EMU2. The average standard deviation is higher than those of the other datasets. As with the other datasets, the skewness and kurtosis columns indicate that the return distributions depart from the normal distribution. Particularly kurtosis is higher, on average, than in the dataset BEM1. The Jarque–Bera test statistics further confirm the non-normality of the distributions and the corresponding p-values signal statistical significance for the statistics for all the countries.

Table 5 Descriptive statistics for the dataset BEM2

This table presents the mean, standard deviation, skewness, and kurtosis of the logarithmic return distributions for each country in the dataset BEM2. The departure from the normal distribution is summarized by the Jarque–Bera test statistic and the corresponding p-value.

	Mean (%)	St. Dev.	Skewness	Kurtosis	JB-test	p-value
Brazil	0.0761	0.0278	-0.283	9.474	1814.3	< 0.001
China	0.0647	0.0237	0.130	7.435	847.8	< 0.001
Egypt	0.0426	0.0195	-1.182	12.033	3745.1	< 0.001
India	0.0565	0.0231	0.342	10.305	2312.6	< 0.001
Indonesia	0.0882	0.0230	0.041	8.130	1130.7	< 0.001
Mexico	0.0388	0.0213	0.087	7.549	890.1	< 0.001
Philippines	0.0577	0.0191	-0.316	6.917	676.4	< 0.001
Poland	-0.0015	0.0235	-0.129	5.750	327.6	< 0.001
Russia	-0.0219	0.0298	-0.217	15.460	6677.9	< 0.001
South Africa	0.0312	0.0217	-0.295	6.125	434.4	< 0.001
South Korea	0.0103	0.0242	0.105	20.097	12558.2	< 0.001
Turkey	0.0434	0.0274	-0.082	5.657	304.5	< 0.001
Average	*0.0405*	*0.0237*	*-0.150*	*9.578*		

Of the countries in the dataset BEM2, Russia stands out as an undesirable investment with the lowest mean return of -0.0219% and the highest standard deviation of 0.0298. On the other hand, Brazil exhibits the second highest standard deviation but also has the second highest mean return. The countries with the lowest standard deviations, Philippines, Egypt, and Mexico, all exhibit only moderate mean returns relative to the other countries in the dataset. There are no countries that clearly stand out as good investments in terms of the mean and standard deviation, contrary to the datasets EMU1 and EMU2.

As mentioned in Fabozzi et al. (2006a, 424), it is important to evaluate portfolio strategies under different circumstances. The four datasets provide an opportunity to analyze the performance of portfolios under various market conditions: The average statistics in Table 2, Table 3, Table 4, and Table 5 above show that the parallel datasets EMU1 and BEM1, both spanning from June 2002 to May 2006, exhibit significantly higher mean returns and lower standard deviations than their sequel datasets EMU2 and BEM2 spanning from June 2006 to May 2010.

3.2.2 Serial dependence

For volatility modeling purposes, it is necessary to test for autocorrelation in the squared logarithmic return series in each dataset. If significant autocorrelation is found, the use of volatility modeling to forecast future volatility of the returns is justifiable. For each country in the datasets, the autocorrelation of squared logarithmic returns is computed for seven lags according to Equation 17 and Equation 18. In addition, the Ljung–Box test statistic defined in Equation 19 and the corresponding p-value are calculated for each country for a summary of the total autocorrelation over the seven lags.

Table 6 presents the autocorrelation statistics of the squared returns in the dataset EMU1. For all countries, autocorrelation greater than 0.1 is found on at least one lag. Furthermore, a few countries even exhibit autocorrelation greater than 0.3, Netherlands having the largest autocorrelation in the dataset of 0.413 at lag five.

Table 6 Autocorrelation of squared returns in the dataset EMU1

This table presents the autocorrelation of the squared logarithmic returns for seven lags for each country in the dataset EMU1. The Ljung–Box test statistic and the corresponding p-value denote the significance of the total autocorrelation over the seven lags.

	Lag=1	Lag=2	Lag=3	Lag=4	Lag=5	Lag=6	Lag=7	Ljung-Box	p-value
Austria	0.156	0.098	0.114	0.137	0.079	0.070	0.076	85.0	< 0.001
Belgium	0.309	0.261	0.214	0.272	0.216	0.291	0.229	480.7	< 0.001
Finland	0.044	0.118	0.084	0.117	0.093	0.153	0.068	75.8	< 0.001
France	0.247	0.330	0.297	0.244	0.327	0.395	0.293	686.5	< 0.001
Germany	0.233	0.316	0.333	0.313	0.306	0.327	0.326	689.8	< 0.001
Greece	0.028	0.171	0.099	0.113	0.143	0.082	0.048	84.3	< 0.001
Ireland	0.059	0.150	0.080	0.088	0.096	0.037	0.066	56.9	< 0.001
Italy	0.183	0.269	0.278	0.236	0.301	0.300	0.271	508.1	< 0.001
Netherlands	0.253	0.314	0.370	0.212	0.413	0.317	0.299	725.6	< 0.001
Portugal	0.243	0.166	0.213	0.225	0.137	0.102	0.144	238.9	< 0.001
Spain	0.194	0.347	0.263	0.309	0.203	0.350	0.371	642.4	< 0.001

The Ljung–Box test statistics for the seven lags show notable autocorrelation in the squared return series of all the countries. The low p-values for all the countries indicate that the test statistics are statistically significant. Nevertheless, there are significant differences between the countries: Ireland shows relatively low levels of autocorrelation, below 0.1 for the majority of the lags. In contrast, Germany exhibits very notable serial dependence in the return series with autocorrelation greater than 0.3 at six of the seven lags. According to the statistics in Table 6, the return series clearly exhibit serial dependence, and volatility modeling could be worth implementing to forecast future volatility of the returns.

Table 7 presents the autocorrelation of the squared return series in the dataset EMU2. In general, the autocorrelation present is greater for the countries in the dataset EMU2 than in the dataset EMU1. All countries, with the exception of Finland, exhibit autocorrelation greater than 0.2 for at least one lag, Netherlands even featuring autocorrelation of 0.454 at lag five.

Table 7 Autocorrelation of squared returns in the dataset EMU2

This table presents the autocorrelation of the squared logarithmic returns for seven lags for each country in the dataset EMU2. The Ljung–Box test statistic and the corresponding p-value denote the significance of the total autocorrelation over the seven lags.

	Lag=1	Lag=2	Lag=3	Lag=4	Lag=5	Lag=6	Lag=7	Ljung-Box	p-value
Austria	0.340	0.335	0.355	0.285	0.303	0.314	0.297	730.9	< 0.001
Belgium	0.233	0.241	0.089	0.176	0.145	0.317	0.098	289.1	< 0.001
Finland	0.123	0.130	0.155	0.099	0.172	0.120	0.071	117.7	< 0.001
France	0.197	0.239	0.229	0.232	0.329	0.132	0.163	363.2	< 0.001
Germany	0.140	0.238	0.239	0.151	0.277	0.064	0.219	292.5	< 0.001
Greece	0.151	0.197	0.276	0.239	0.235	0.209	0.223	352.1	< 0.001
Ireland	0.264	0.223	0.124	0.290	0.293	0.391	0.263	540.6	< 0.001
Italy	0.188	0.222	0.250	0.276	0.291	0.176	0.126	363.1	< 0.001
Netherlands	0.244	0.328	0.296	0.254	0.454	0.209	0.229	636.8	< 0.001
Portugal	0.129	0.132	0.127	0.177	0.243	0.126	0.125	176.8	< 0.001
Spain	0.157	0.161	0.168	0.273	0.188	0.116	0.140	227.3	< 0.001

The Ljung–Box statistics tell an interesting story: autocorrelation in the dataset EMU2 appears to be lower for those countries that have a high test statistic in the dataset EMU1. The opposite is true for the countries that display lower Ljung–Box statistics in the dataset EMU1; their autocorrelation appears to be higher in the dataset EMU2. The most profound example is Austria, which has a Ljung–Box statistic of 730.9 in the dataset EMU2 whereas in the dataset EMU1 the test statistic is only 85.0. These types of observations imply that the parameters underlying the data may not be stationary. The low p-values in Table 7 denote that the Ljung–Box test statistics are statistically significant. Overall, the statistics in Table 7 indicate notable serial dependence in the return series. Therefore, volatility forecasting is likely to be worth implementing.

The autocorrelation statistics of the squared return series in the dataset BEM1 are presented in Table 8. The autocorrelation levels for most countries are lower than for countries in the datasets EMU1 and EMU2. The fact that the countries in this dataset are exposed to currency exchange rate fluctuations may have an effect on the autocorrelation levels. However, most countries still exhibit autocorrelation greater than 0.1 for at least one lag.

Table 8 Autocorrelation of squared returns in the dataset BEM1

This table presents the autocorrelation of the squared logarithmic returns for seven lags for each country in the dataset BEM1. The Ljung–Box test statistic and the corresponding p-value denote the significance of the total autocorrelation over the seven lags.

	Lag=1	Lag=2	Lag=3	Lag=4	Lag=5	Lag=6	Lag=7	Ljung-Box	p-value
Brazil	0.192	0.169	0.211	0.280	0.141	0.253	0.139	303.3	< 0.001
China	0.064	0.077	0.095	0.103	0.128	0.053	0.136	70.2	< 0.001
Egypt	0.221	0.092	0.072	0.038	0.137	0.003	-0.011	86.0	< 0.001
India	0.472	0.141	0.058	0.091	0.084	0.029	0.002	272.1	< 0.001
Indonesia	0.061	0.093	0.043	0.010	0.046	0.030	-0.002	18.1	0.012
Mexico	0.134	0.117	0.242	0.079	0.111	0.213	0.086	168.3	< 0.001
Philippines	0.091	-0.001	0.082	0.019	0.078	0.022	0.000	22.8	0.002
Poland	0.003	0.100	0.049	0.058	0.141	0.073	0.072	48.0	< 0.001
Russia	0.161	0.169	0.199	0.108	0.137	0.104	0.034	142.0	< 0.001
South Africa	0.175	0.107	0.131	0.092	0.206	0.185	0.192	188.5	< 0.001
South Korea	0.044	0.173	0.100	0.061	0.203	0.109	0.152	126.8	< 0.001
Turkey	0.256	0.140	0.053	0.092	0.120	0.054	0.063	122.5	< 0.001

The Ljung–Box statistics are, in general, lower in the dataset BEM1 than the statistics in the datasets EMU1 and EMU2. Indonesia even features a low test statistic of 18.1, with autocorrelation levels lower than 0.1 at each lag. However, the p-values for the test statistics are low for all countries and denote statistical significance. Overall, the dataset BEM1 does not appear as attractive for volatility modeling as the datasets EMU1 and EMU2. Based on the statistics in Table 8, volatility modeling for countries such as Indonesia or Philippines may not be worthwhile.

Table 9 presents autocorrelation statistics of the squared returns in the dataset BEM2. Similarly to the relationship between the datasets EMU1 and EMU2, the dataset BEM2 exhibits higher levels of autocorrelation than present for countries in the dataset BEM1. All countries in the dataset BEM2 exhibit autocorrelation levels greater than 0.1 for at least one lag.

Table 9 Autocorrelation of squared returns in the dataset BEM2

This table presents the autocorrelation of the squared logarithmic returns for seven lags for each country in the dataset BEM2. The Ljung–Box test statistic and the corresponding p-value denote the significance of the total autocorrelation over the seven lags.

	Lag=1	Lag=2	Lag=3	Lag=4	Lag=5	Lag=6	Lag=7	Ljung-Box	p-value
Brazil	0.137	0.419	0.219	0.260	0.467	0.260	0.360	751.4	< 0.001
China	0.210	0.335	0.331	0.166	0.206	0.154	0.189	410.3	< 0.001
Egypt	0.149	0.053	0.041	0.043	0.101	0.035	0.030	42.2	< 0.001
India	0.072	0.078	0.071	0.082	0.121	0.088	0.096	56.8	< 0.001
Indonesia	0.174	0.148	0.173	0.187	0.119	0.071	0.062	145.0	< 0.001
Mexico	0.208	0.322	0.314	0.237	0.363	0.259	0.290	606.0	< 0.001
Philippines	0.059	0.149	0.174	0.137	0.087	0.017	0.091	93.9	< 0.001
Poland	0.137	0.123	0.180	0.295	0.259	0.155	0.136	273.3	< 0.001
Russia	0.126	0.162	0.242	0.200	0.135	0.156	0.154	214.6	< 0.001
South Africa	0.242	0.275	0.125	0.277	0.248	0.176	0.230	385.0	< 0.001
South Korea	0.062	0.113	0.065	0.170	0.284	0.191	0.039	174.7	< 0.001
Turkey	0.108	0.131	0.142	0.209	0.111	0.116	0.055	125.8	< 0.001

The Ljung–Box test statistics for most of the countries are higher in the dataset BEM2 than in the dataset BEM1. For example, Mexico has a high test statistic of 606.0 and exhibits autocorrelation greater than 0.2 at each lag whereas in the dataset BEM1 its test statistic is 168.3. The p-values for all the countries are low and indicate statistical significance of the Ljung–Box statistics. Overall, the autocorrelation statistics in Table 9 demonstrate notable serial dependence in the return series. Therefore, volatility modeling may be useful for the return series in the dataset BEM2.

3.3 Portfolio optimization and the investment process

3.3.1 Determining the minimum variance portfolio

To study the performance of an investment portfolio over a large sample of data, computer programming is essential. What is needed is a computer algorithm that will automatically go through the data and allocate wealth across available assets according to optimal portfolio weights. The optimal weights are calculated repeatedly at specific intervals. To accomplish this in the empirical study, the statistical programming language R is utilized. The return series in the datasets can be loaded into the R program, after which an algorithm that deals with portfolio optimization and the investment process is run. The programming code developed for the empirical study is provided in Appendix 1.

The thesis examines the performance of minimum variance portfolios with different estimation models for the covariance matrix, compared against a benchmark strategy. Therefore, the optimization problem needed to solve during the investment process is the one for the minimum variance portfolio, defined in Equation 5.

The minimum variance portfolio optimization problem can be solved by quadratic programming which takes the following form:

$$\begin{gathered} \min_{\boldsymbol{x}} \left(\frac{1}{2} \boldsymbol{x}' \boldsymbol{Q} \boldsymbol{x} + \boldsymbol{c}' \boldsymbol{x} \right) \\ \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \\ \boldsymbol{C}\boldsymbol{x} \geq \boldsymbol{d}. \end{gathered} \tag{29}$$

In this equation, $\boldsymbol{Q}$ denotes an $N \times N$ matrix, $\boldsymbol{c}$ denotes an N-dimensional vector, $\boldsymbol{A}$ denotes an $M \times N$ matrix, $\boldsymbol{b}$ denotes an M-dimensional vector, $\boldsymbol{C}$ denotes a $P \times N$ matrix, and $\boldsymbol{d}$ denotes a P-dimensional vector. When $\boldsymbol{Q}$ is a positive semidefinite matrix, which is the case with covariance matrixes, we are dealing with a convex programming problem. Convex programming problems have the pleasant feature that a local optimum solution is also a global optimum. (Cornuejols & Tütüncü 2007, 121, 138–139.)

Since the thesis considers only minimum variance portfolios, vector $\boldsymbol{c}$ in the above Equation 29 consists of only zeros. Therefore, matrix $\boldsymbol{Q}$ is the only parameter that needs to be estimated from the data. This parameter, or the covariance matrix, needs to be estimated again every time the portfolio is considered for rebalancing. In addition, there is a requirement that the entire wealth of the investor is allocated to the assets and thus $\boldsymbol{A}$ is a vector of ones and $\boldsymbol{b}$ is a scalar one in the constraint $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$.

In order to avert irrational portfolio weights, and to make the investment process more realistic, we set a portfolio weight constraint

$$|w_i| \leq \alpha, \tag{30}$$

where w_i denotes the weight for asset i, and α denotes a boundary for each asset weight. In the empirical study, we set $\alpha = 0.25$, so that no asset can get more than a quarter of the total wealth allocated to it, be it a long or a short position. In the quadratic optimization problem defined in Equation 29, the absolute value of the weights can be expressed by first taking the lower bound of the constraint to $\boldsymbol{C}\boldsymbol{x} \geq \boldsymbol{d}$, where $\boldsymbol{C}$ is a diagonal matrix of ones and $\boldsymbol{d}$ is a vector consisting of the parameter $-\alpha$. The upper bound can be expressed so that $\boldsymbol{C}$ is a diagonal matrix of minus ones, and $\boldsymbol{d}$ is a vector consisting of the parameter $-\alpha$. This constraint is applied to all minimum variance portfolios constructed in the empirical study. Moreover, portfolio weight constraints have been utilized in previous studies as well: For example, DeMiguel et al. (2009b) apply a short-sale constraint in part of the tested models.

3.3.2 Estimation window

As Fabozzi et al. (2006a, 424–426) point out, it is important to test portfolio performance on a different part of data than parameter estimation is done with; this means that the data needs to be separated into in-sample and out-of-sample parts. According to Fabozzi et al., if a portfolio strategy is not tested on out-of-sample data but instead on the same data that is used for parameter estimation, it presents a danger of discovering a model that may perform well on the sample data but will do poorly out-of-sample.

In the empirical study, the estimation of the covariance matrix is done on a specific length of data called the estimation window. The programming code presented in Appendix 1 goes through an entire dataset, stopping at specific intervals to re-estimate the covariance matrix and to determine new optimal weights for the portfolio. During this process, the estimation window is always comprised of data that has already been gone through, so-called in-sample data. In contrast, portfolio performance evaluation is done on data that is yet to come, that is, out-of-sample data. As the programming code goes through the dataset, old out-of-sample data becomes in-sample data on which the parameter estimation is performed. Thus, only the most recent information is used for the parameter estimation. This type of approach is used in DeMiguel et al. (2009b, 1927–1928) as well.

At the beginning of the investment period, there are 250 days (approximately one year) of startup data for each of the datasets to form the estimation window. For the datasets EMU1 and BEM1, this extra data consists of the time period from June 2001 to May 2002. For the datasets EMU2 and BEM2, this time period is from June 2005 to May 2006, which overlaps with the datasets EMU1 and BEM1. Thus, old out-of-sample data is used as in-sample data for parameter estimation in the sequel datasets EMU2 and BEM2.

Broadie (1993, 47–52) presents evidence of nonstationary parameters; a return series can have different underlying parameters in one decade and the next decade. In theory, this phenomenon could apply to shorter time periods as well. For this reason, we test three different estimation window lengths in the empirical study: 20 days (approximately one month), 60 days (approximately one quarter of a year), and 250 days (approximately one year).

3.3.3 *Rebalancing and trading costs*

Throughout the investment period, the portfolio will be rebalanced at specific intervals to update old portfolio weights to new ones calculated from data in the new estimation window. Thus, at each point the portfolio is considered for rebalancing, the covariance matrix is estimated from the data and new optimal weights are calculated. Upon rebalancing, required trading takes place to bring the portfolio weights to the new optimal ones.

Obviously, rebalancing has an effect on the amount of trading costs: the more frequently the portfolio is rebalanced, the greater the portfolio turnover, and thus, the larger the total trading costs. While it may be ideal to rebalance the portfolio as often as possible if trading were costless, in the real world we have to deal with friction in the market. A compromise has to be made between up-to-date portfolio weights and losing money in the form of trading costs incurred by rebalancing (Fabozzi et al. 2006a, 91).

The empirical study takes trading costs into account and includes them in the performance of the portfolio. Each time the portfolio is rebalanced, the incurring trading costs are deducted from the investor's wealth. The trading costs for the entire portfolio are determined according to the following formula:

$$TC = \sum_{i=1}^{N} \left| w_{i,new} - w_{i,old} \right| * c, \tag{31}$$

where $w_{i,new}$ denotes the new optimal weight for asset i, $w_{i,old}$ denotes the previous weight for asset i, c denotes the assumed level of transaction costs in basis points, and N denotes the number of available assets.

To investigate the effect that the length of the rebalancing interval has on portfolio performance, we consider three fixed intervals: 5-day (weekly), 20-day (approximately monthly), and 60-day (approximately quarterly) rebalancing. In addition, a special criterion for rebalancing is introduced in the form of threshold rebalancing: the new optimal portfolio weights are calculated each trading day, after which the following inequality determines whether rebalancing the portfolio is worth the reduction in wealth incurred by trading costs:

$$\frac{\sum_{i=1}^{N} \left| w_{i,new} - w_{i,old} \right| * c}{\sqrt{\boldsymbol{w}_{old} \boldsymbol{\Sigma} \mathbf{w}'_{\text{old}}} - \sqrt{\boldsymbol{w}_{new} \boldsymbol{\Sigma} \mathbf{w}'_{\text{new}}}} < 1, \tag{32}$$

where $\boldsymbol{w}_{old}$ denotes a vector of the current portfolio weights, $\boldsymbol{w}_{new}$ denotes a vector of the new optimal weights, and $\boldsymbol{\Sigma}$ denotes the new covariance matrix estimate. The above criterion is artificial in the sense that we have simply selected that the incurred trading costs need to be smaller than the change in estimated portfolio volatility. Nevertheless, it describes a preference fairly well and serves the purpose.

3.4 Estimation models tested

3.4.1 Benchmark strategy

To properly evaluate the performance of different parameter estimation models, a simple benchmark strategy is needed, against which the more complex models are compared. We select the equal-weight strategy, where the investor allocates wealth evenly across the available assets, as the benchmark. In this strategy, the portfolio weight for each asset i is defined as:

$$w_i = \frac{1}{N}, \tag{33}$$

where N denotes the total number of assets available for investment. Thus, w_i is equal for every asset in the portfolio.

The equal-weight strategy serves as a good benchmark because in spite of more complex portfolio selection models, investors still keep using simple allocation rules, often dividing their wealth evenly across available assets (see, for example, Benartzi & Thaler 2001; Huberman & Jiang 2006, 764). Furthermore, DeMiguel et al. (2009b, 1917–1921) compare different models against the equal-weight strategy and advocate its use as a benchmark.

The benchmark strategy is not actually an estimation model since there are no parameters to estimate from data. The only variable that needs to be defined for the equal-weight strategy is the rebalancing interval: Since the prices of assets change over time, the equal-weight strategy needs to be rebalanced in order to keep the assets weighted evenly in the portfolio. Before comparing the more complex models against the equal-weight strategy, we test the strategy with daily, 2-day, 5-day, and 20-day rebalancing intervals. The interval that produces the best result in terms of the average out-of-sample Sharpe ratio across the four datasets is then selected as the rebalancing interval for the benchmark strategy.

3.4.2 Sample covariance matrix model

The sample covariance matrix is the easiest and most straightforward way of estimating the covariance matrix, therefore it serves as a natural starting point. In this approach, the covariance matrix is estimated directly from the logarithmic return series in the estimation window, according to Equation 12. The sample covariance matrix portfolio is also tested in DeMiguel et al. (2009b).

Because each observation in the estimation window is given equal weight in the calculation of the sample covariance matrix, the estimation window length can obviously have a large effect on the produced estimate. Hence, the testing of different estimation window lengths becomes even more important for this estimation model.

The sample covariance matrix is tested with the previously mentioned three estimation window lengths and three fixed rebalancing intervals. In addition, the threshold rebalancing criterion defined in Equation 32 is utilized in the case of the 20-day and 60-day estimation windows. This amounts to a total of 11 sample covariance matrix portfolios evaluated in the empirical study.

3.4.3 Constant-correlation shrinkage model

The constant-correlation shrinkage model for the covariance matrix proposed by Ledoit and Wolf (2004) is the second estimation model tested in the empirical study. The model combines the constant-correlation covariance matrix with the sample covariance matrix, as defined in Equation 14. The optimal shrinkage weight needs to be computed from data in the estimation window and is defined as:

$$\delta^* = \max\left\{0, \min\left\{\frac{\pi - \vartheta}{T\gamma}, 1\right\}\right\}, \tag{34}$$

where

$$\pi = \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\frac{1}{T}\Big((r_{it} - \bar{r}_i)(r_{jt} - \bar{r}_j) - \sigma_{ij}\Big)^2, \tag{35}$$

$$\begin{aligned}\vartheta = {} & \textstyle\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{T}((r_{it} - \bar{r}_i)^2 - \sigma_i^2)^2 + \sum_{i=1}^{N}\sum_{j=1, j\neq i}^{N}\Bigg(\frac{\rho}{2}\Big(\frac{\sigma_j}{\sigma_i}\frac{1}{T}\sum_{t=1}^{T}((r_{it} - \bar{r}_i)^2 - \\ & \sigma_i^2) \times \Big((r_{it} - \bar{r}_i)(r_{jt} - \bar{r}_j) - \sigma_{ij}\Big)\Big) + \Big(\frac{\sigma_i}{\sigma_j}\frac{1}{T}\sum_{t=1}^{T}\Big((r_{jt} - \bar{r}_j)^2 - \sigma_j^2\Big) \times \\ & \Big((r_{it} - \bar{r}_i)(r_{jt} - \bar{r}_j) - \sigma_{ij}\Big)\Big)\Bigg),\end{aligned} \tag{36}$$

and

$$\gamma = \sum_{i=1}^{N} \sum_{j=1}^{N} \left(\rho\sigma_i\sigma_j - \sigma_{ij}\right)^2. \tag{37}$$

In the above equations, r_{it} and r_{jt} denote the logarithmic returns for asset i and asset j at time t, $\bar{r}_i$ and $\bar{r}_j$ denote the arithmetic means of the returns for asset i and asset j, σ_{ij} denotes the covariance of asset i and asset j, σ_i and σ_j denote the standard deviations of asset i and asset j, ρ denotes the average correlation of the assets, T denotes the length of the estimation window, and N denotes the number of the available assets. (Ledoit & Wolf 2004, 112–113, 117–118.)

The shrinkage estimation model is tested with the previously mentioned three estimation window lengths and three fixed rebalancing intervals. In addition, the threshold rebalancing criterion is utilized in the case of the 20-day and 60-day estimation windows. Thus, a total of 11 minimum variance portfolios with shrinkage estimation are evaluated in the empirical study.

3.4.4 DCC-GARCH model

Since there is evidence of autocorrelation in the squared return series of the assets, it is worth attempting to take advantage of this feature by employing a GARCH model to forecast future volatility. We use the GARCH(1,1) model to forecast individual asset volatilities. The GARCH(1,1) model is a special case of the general model presented in Equation 21. It is defined as:

$$\sigma_t^2 = \omega + \alpha r_{t-1}^2 + \beta \sigma_{t-1}^2, \tag{38}$$

where ω, α, and β are parameters that need to be calibrated for the data, r_{t-1}^2 denotes the squared return from the previous time period, and σ_{t-1}^2 denotes the value of the model from the previous period (Hull 2009, 477–481).

To forecast the entire covariance matrix, we need to forecast the correlation structure as well. To achieve this, we utilize the two-step procedure described in Engle and Sheppard (2001, 4–7), constructing a DCC model introduced in Engle (2002, 341–342). The general form of this model is defined in Equation 26 and Equation 27. We employ a DCC(1,1) model and require that $\alpha + \beta = 1$. This modifies the GARCH-like process in the DCC model into an integrated process also introduced in Engle (2002, 341) and thus allows for a simpler and faster iterative search for the parameter in the second step.

The so-called integrated DCC process and the resulting time-varying correlation matrix are defined as follows:

$$\boldsymbol{Q}_t = (1-\lambda)\boldsymbol{\varepsilon}_{t-1}\boldsymbol{\varepsilon}'_{t-1} + \lambda\boldsymbol{Q}_{t-1} \tag{39}$$

$$\boldsymbol{R}_t = \boldsymbol{Q}_t^{*-1}\boldsymbol{Q}_t\boldsymbol{Q}_t^{*-1}, \tag{40}$$

where λ is the parameter calibrated for the data, $\boldsymbol{\varepsilon}_{t-1}$ denotes a vector of standardized residuals of the previous period from the GARCH(1,1) model estimated for the individual volatilities, and $\boldsymbol{Q}_t^*$ is a diagonal matrix that consists of the square roots of the diagonal elements of matrix $\boldsymbol{Q}_t$. After the time-varying correlation matrix $\boldsymbol{R}_t$ is determined, the conditional covariance matrix is formed by multiplying the conditional volatility matrix with the correlation matrix as in Equation 23. (Engle & Sheppard 2001, 5; Engle 2002, 341.)

To determine the parameters for the DCC-GARCH model, maximum likelihood estimation is employed. The programming code for the likelihood estimation process is provided in Appendix 1. Following Engle (2002, 342), we partition the likelihood estimation into two stages:

$$L(\boldsymbol{\theta}, \boldsymbol{\phi}) = L_V(\boldsymbol{\theta}) + L_C(\boldsymbol{\theta}, \boldsymbol{\phi}), \tag{41}$$

where the volatility term is

$$L_V(\boldsymbol{\theta}) = -\frac{1}{2}\sum_t \left(N\log(2\pi) + \log|\boldsymbol{D}_t|^2 + \boldsymbol{r}'_t\boldsymbol{D}_t^{-2}\boldsymbol{r}_t\right), \tag{42}$$

and the correlation part is

$$L_C(\boldsymbol{\theta}, \boldsymbol{\phi}) = -\frac{1}{2}\sum_t \left(\log|\boldsymbol{R}_t| + \boldsymbol{\varepsilon}'_t\boldsymbol{R}_t^{-1}\boldsymbol{\varepsilon}_t - \boldsymbol{\varepsilon}'_t\boldsymbol{\varepsilon}_t\right). \tag{43}$$

The volatility likelihood is the sum of individual GARCH likelihoods and can be jointly maximized by singly maximizing each term in

$$L_V(\boldsymbol{\theta}) = -\frac{1}{2}\sum_t \sum_{i=1}^{N} \left(\log(2\pi) + \log\left(\sigma_{i,t}^2\right) + \frac{r_{i,t}^2}{\sigma_{i,t}^2}\right) \tag{44}$$

(Engle 2002, 342). In the above equations, $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$ denote vectors of the parameters that are estimated in the maximum likelihood process, N denotes the number of available assets, $\boldsymbol{D}_t$ denotes a matrix of conditional volatilities at time t, $\boldsymbol{r}_t$ denotes a vector of returns at time t, $\boldsymbol{\varepsilon}_t$ denotes a vector of standardized residuals at time t, $\sigma_{i,t}^2$ denotes the conditional variance for asset i at time t, and $r_{i,t}^2$ denotes the squared return for asset i at time t (Engle 2002, 341–342). For the starting value of the iterative process, we set $\sigma_{i,1}^2$ equal to the unconditional variance of the return series in the estimation window.

The second step is finding the maximum for L_C, which is for the estimation of the conditional correlation matrix. The optimal parameter vector from the first step,

$$\boldsymbol{\theta}^* = \operatorname{argmax}\{L_V(\boldsymbol{\theta})\}, \tag{45}$$

is given as a constant in the second step:

$$\max_{\boldsymbol{\phi}} L_C(\boldsymbol{\theta}^*, \boldsymbol{\phi}) \tag{46}$$

(Engle 2002, 342). For the starting value of the process, we set $\boldsymbol{Q}_1$ equal to the average of the product of previous residual vectors and the unconditional correlation matrix of the residuals. When the estimation window is sufficiently long, such as 250 days, the selected value becomes less relevant. There appears to be no guideline on selecting a starting value for the process in the literature. The purpose of the above described starting value is to kick off the iterative process so that all weight in the DCC model is not placed on the regressive part of the process.

It is worth mentioning that iterative optimization methods may not be able to find the absolute optimal solutions for the maximum likelihood estimation problems. However, differences between the obtained parameters and the true optimal values are likely to be small. As is usually the case with any estimation procedure, the parameters estimated for this process include estimation error.

Once the parameters have been estimated with the maximum likelihood method, the DCC-GARCH model is utilized to forecast the time-varying correlation matrix for the next period, $t+1$. In the same manner, the individual asset variances are forecasted for $t+1$. These forecasts are then combined to form the covariance matrix for $t+1$:

$$\boldsymbol{\Sigma}_{t+1} = \boldsymbol{D}_{t+1}\boldsymbol{R}_{t+1}\boldsymbol{D}_{t+1}. \tag{47}$$

This is the covariance matrix that is utilized in determining the minimum variance portfolio.

Since the data we are dealing with are daily return series from different countries and the closing prices for different stock markets are recorded at different times of the day, a worldwide market movement can show up in one country's closing price a day later than in another country's closing price. This might affect the performance of the DCC-

GARCH model. Ideally, the closing prices should be recorded at exactly the same time of the day. This is approximately true in the case of the datasets EMU1 and EMU2 as the countries included in these datasets are situated in Europe and their stock markets close at approximately the same hour. However, the situation is different for the datasets BEM1 and BEM2 where, for example, South Korea's stock market closes at around the time Brazil's stock market opens. Thus, trading in Brazil's stock market and its price movements show up only in the next day's closing prices of South Korea's stock market, assuming that movements in these two markets are dependent. On the other hand, the DCC-GARCH model could be thought of as taking this phenomenon into account and implicitly modeling the information lag between the stock markets. For simplification purposes, this matter is not considered relevant in the thesis. However, it would be ideal to investigate the performance of the model in one stock market only.

Since the DCC-GARCH approach is different from the sample covariance matrix and shrinkage portfolios in that it attempts to take advantage of short-term forecasts made for the next trading day, the parameters for the DCC-GARCH model are estimated daily. For this reason, threshold rebalancing is the only form of rebalancing examined for DCC-GARCH portfolios in the empirical study. Furthermore, the shortest estimation window (20 days) is not considered in the case of DCC-GARCH portfolios since it would not permit the maximum likelihood method for the parameter estimation to be accurate. Therefore, a total of two DCC-GARCH portfolios are evaluated in each data-set.

3.5 Performance evaluation

The performance evaluation of the minimum variance portfolios is done by comparing their out-of-sample Sharpe ratios to that of the benchmark strategy. The Sharpe ratio was originally introduced in Sharpe (1966, 122–125) as reward-to-variability ratio; it is defined as

$$SR_p = \frac{\mu_p - r_f}{\sigma_p} \tag{48}$$

where μ_p denotes the mean return of the portfolio, r_f denotes a riskless rate of return, and σ_p denotes the standard deviation of the portfolio. For simplification purposes, we do not consider a riskless rate in the empirical study. Hence, r_f in Equation 48 is assumed to equal zero. Thus, Sharpe ratios reported in the empirical results are simply ratios of the mean and standard deviation of portfolio returns. Furthermore, the reported Sharpe ratios are annualized for more sensible values. The annualization is done by multiplying the daily ratio of the mean return and standard deviation by the square root of the approximate amount of trading days per year, or 250. This modified Sharpe ratio is formally defined as

$$SR_{Mod} = \sqrt{250}\,\frac{\mu_p}{\sigma_p}. \tag{49}$$

In the empirical results, a reference in the text for Sharpe ratio means the modified ratio defined in Equation 49.

There are different opinions on the use of the Sharpe ratio as a portfolio performance measure. For example, Biglova, Ortobelli, Rachev and Stoyanov (2004) compare different risk measures and present that the Sharpe ratio is not an accurate performance measure when portfolio returns are not normally distributed. On the other hand, Eling and Schuhmacher (2007) examine how the Sharpe ratio works in evaluating the performance of hedge funds, which are known to exhibit returns that deviate notably from the normal distribution. They find that the Sharpe ratio produces virtually identical rankings for the funds compared to several more sophisticated performance measures.

While there may not be a clear consensus on the goodness of the Sharpe ratio as a performance measure, it appears to be somewhat of a standard measure for evaluating portfolio performance. The Sharpe ratio is also utilized in DeMiguel et al. (2009b) to compare portfolios against the benchmark strategy. Hence, it is reasonable to use the Sharpe ratio as a performance measure for better comparability of results with previous research. It is worth mentioning that the Sharpe ratios in the empirical study of the thesis include the effect of trading costs, whereas the Sharpe ratios reported in DeMiguel et al. (2009b, 1931) do not.

Furthermore, we test for the statistical significance of the differences between the Sharpe ratios of the benchmark and minimum variance portfolios. The statistical test aims to find out whether the Sharpe ratio of the minimum variance portfolio is truly different from that of the benchmark portfolio. To achieve this, the HAC inference method discussed in Ledoit and Wolf (2008, 852–853) is employed by utilizing the R code provided by the authors. DeMiguel et al. (2009b, 1927–1931) also test the statistical significance of the difference between Sharpe ratios but they use an older method by Jobson and Korkie (1981, 891–895), which is shown in Ledoit and Wolf (2008, 857–858) to produce smaller p-values in general. Ledoit and Wolf (2008, 858) point out that the older method is not good because of the heavy tails and time series characteristics present in financial returns. In contrast, they state that the HAC inference method works well with large samples. For comparison, we also compute p-values according to Jobson and Korkie (1981, 891–895), using the apparently corrected formula presented in DeMiguel et al. (2009b, 1928).

In addition to the Sharpe ratio, we also report and compare the turnover amounts of the portfolios to the turnover of the benchmark strategy. This provides insight into the effects of transaction costs on the portfolio performance. The total turnover of the portfolio is calculated in the following manner:

$$\tau = \sum_{t=1}^{T} |\boldsymbol{w}_t - \boldsymbol{w}_{t-1}| \boldsymbol{\iota} \tag{50}$$

where $\boldsymbol{w}_t$ denotes a vector of portfolio weights at time t, $\boldsymbol{w}_{t-1}$ denotes the portfolio weights at time $t-1$, $\boldsymbol{\iota}$ denotes a vertical vector of ones, and T denotes the number of trading days in the investment period.

4 EMPIRICAL RESULTS

The programming code presented in Appendix 1 was run for all combinations of the covariance matrix estimation models, estimation window lengths, and rebalancing intervals. Thus, time series of portfolio returns were obtained. In this section, we evaluate the performance of these portfolios; their out-of-sample Sharpe ratios and turnover amounts are compared against the benchmark strategy.

Before examining the performance of the minimum variance portfolios, we determined the rebalancing interval that produced the best results for the equal-weight strategy. The turnover of the equal-weight strategy is very low and different rebalancing intervals did not have a large effect on its performance. Across all four datasets, the portfolio rebalanced every five days (weekly) produced the highest average Sharpe ratio of 0.623. The differences were small: the average Sharpe ratio was 0.622 for the portfolio rebalanced every two days, 0.616 for the daily rebalanced portfolio, and 0.615 for the monthly rebalanced portfolio. Based on these results, we select the weekly rebalanced equal-weight portfolio as the benchmark strategy.

The p-values calculated for the differences between the Sharpe ratios of the benchmark and minimum variance portfolios using the HAC inference method discussed in Ledoit and Wolf (2008, 852–853) and the method presented in Jobson and Korkie (1981, 891–895) were not significantly different. Therefore, only the p-values obtained using the HAC inference method are reported.

4.1 Performance of the portfolios

4.1.1 Sample covariance matrix portfolios

The out-of-sample Sharpe ratios of the minimum variance portfolios that employ the sample covariance matrix in portfolio optimization are reported in Table 10. There is a total of 11 minimum variance portfolios, each with a different combination of estimation window length and rebalancing interval. The Sharpe ratios of the minimum variance portfolios are compared to those of the benchmark strategy and p-values for the difference are in parentheses below each Sharpe ratio.

Table 10 Out-of-sample Sharpe ratios of the sample covariance matrix portfolios

The table presents the out-of-sample Sharpe ratios for the benchmark strategy and for each combination of estimation window length and rebalancing frequency for minimum variance portfolios constructed using the sample covariance matrix estimator in portfolio optimization. In parentheses below each Sharpe ratio is the p-value for the difference between the above Sharpe ratio and the Sharpe ratio of the benchmark strategy. The highest Sharpe ratio of the minimum variance portfolios in each dataset is in bold text.

Strategy	**EMU1**	**EMU2**	**BEM1**	**BEM2**
Benchmark				
1/N rebalanced weekly	0.691	-0.214	1.397	0.619
Sample covariance matrix portfolios				
Estimation window = 250 days				
Rebalanced quarterly	1.402	-0.022	**1.406**	0.472
	(0.03)	(0.30)	(0.97)	(0.52)
Rebalanced monthly	1.374	-0.060	1.393	0.480
	(0.05)	(0.45)	(0.99)	(0.56)
Rebalanced weekly	1.358	-0.110	1.358	0.390
	(0.06)	(0.62)	(0.88)	(0.33)
Estimation window = 60 days				
Rebalanced quarterly	1.496	**0.316**	1.187	0.448
	(0.01)	(0.01)	(0.45)	(0.51)
Rebalanced monthly	1.343	0.312	1.333	0.360
	(0.07)	(0.03)	(0.82)	(0.34)
Rebalanced weekly	1.258	0.030	1.274	0.057
	(0.13)	(0.32)	(0.68)	(0.03)
Threshold rebalancing	**1.518**	-0.039	1.310	0.311
	(0.02)	(0.46)	(0.75)	(0.29)
Estimation window = 20 days				
Rebalanced quarterly	1.511	0.276	0.527	**0.531**
	(0.02)	(0.02)	(0.01)	(0.76)
Rebalanced monthly	1.227	-0.045	0.936	0.250
	(0.16)	(0.49)	(0.20)	(0.21)
Rebalanced weekly	0.708	-0.222	0.600	-0.837
	(0.96)	(0.98)	(0.03)	(0.00)
Threshold rebalancing	1.362	-0.096	1.296	0.269
	(0.07)	(0.65)	(0.78)	(0.19)

The most striking observation from Table 10 is that in the datasets EMU1 and EMU2, all but one of the minimum variance portfolios have a higher Sharpe ratio than that of the benchmark strategy. The highest Sharpe ratio in the dataset EMU1, 1.518, is achieved with the 60-day estimation window, using the threshold rebalancing criterion.

It has a p-value of 0.02, meaning that the difference to the benchmark's Sharpe ratio is statistically significant. The highest Sharpe ratio in the dataset EMU2 is 0.316, obtained by the quarterly rebalanced portfolio with the 60-day estimation window. It has a p-value of 0.01, denoting statistical significance. This portfolio has a Sharpe ratio of 1.496 in the dataset EMU1, also with a p-value of 0.01. Thus, it performs better than the benchmark strategy in both datasets, and in both cases the difference is statistically very significant.

In the datasets BEM1 and BEM2, the minimum variance portfolios perform, in general, slightly worse than the benchmark strategy. The benchmark strategy has a higher Sharpe ratio than any of the minimum variance portfolios in the dataset BEM2 and a higher Sharpe ratio than all but one of the minimum variance portfolios in the dataset BEM1. The highest Sharpe ratio in the dataset BEM1, 1.406, is achieved by the quarterly rebalanced portfolio with the estimation window of 250 days. Being very close to the benchmark's Sharpe ratio, it has a p-value of 0.97 signifying that the difference is statistically not significant. In the dataset BEM2, the best performing minimum variance portfolio is the quarterly rebalanced portfolio with the 20-day estimation window: it has a Sharpe ratio of 0.531 with a p-value of 0.76.

An interesting observation from the results in Table 10 is that the weekly rebalanced minimum variance portfolios produce the lowest Sharpe ratios for all estimation window lengths in most of the datasets. This is a sign of the fact that trading costs are eating up the profit due to frequent rebalancing. Furthermore, in the datasets EMU1 and EMU2, the quarterly rebalanced portfolios always produced higher Sharpe ratios than the monthly rebalanced portfolios. However, in the datasets BEM1 and BEM2, the results are mixed favoring both monthly and quarterly rebalancing. Threshold rebalancing proved to be a functional criterion for some portfolios but did not produce superior Sharpe ratios in general.

The results in Table 10 do not provide a clear answer to which estimation window length produces the best results. The dataset EMU2 appears to favor the 60-day estimation window, while the 250-day window produces good results in the dataset BEM1. Moreover, the results are mixed in the datasets EMU1 and BEM2, with each estimation window length producing varying results with different rebalancing intervals.

It should be noted that the datasets BEM1 and BEM2 are assumed a trading cost of 40 basis points, whereas the datasets EMU1 and EMU2 are assumed a trading cost of 20 basis points. To further examine the effects that portfolio rebalancing and trading costs have on the performance, the turnover amounts of the portfolios are presented in Table 11.

Table 11 Turnover amounts of the sample covariance matrix portfolios

The table presents the total turnover amounts for the benchmark strategy and for each combination of estimation window length and rebalancing frequency for minimum variance portfolios constructed using the sample covariance matrix in portfolio optimization. The turnover amount is computed according to Equation 50.

Strategy	**EMU1**	**EMU2**	**BEM1**	**BEM2**
Benchmark				
1/N rebalanced weekly	3.4	3.9	6.0	6.1
Sample covariance matrix portfolios				
Estimation window = 250 days				
Rebalanced quarterly	8.4	11.8	7.9	9.6
Rebalanced monthly	11.9	17.3	10.7	14.2
Rebalanced weekly	19.2	39.4	18.8	24.9
Estimation window = 60 days				
Rebalanced quarterly	20.4	22.3	23.4	22.4
Rebalanced monthly	32.4	39.0	36.6	37.9
Rebalanced weekly	58.5	68.9	68.0	77.4
Threshold rebalancing	5.5	14.7	4.3	4.9
Estimation window = 20 days				
Rebalanced quarterly	27.9	28.0	28.8	32.8
Rebalanced monthly	67.6	75.1	85.3	90.8
Rebalanced weekly	147.6	160.2	195.1	202.0
Threshold rebalancing	16.3	57.2	27.8	32.7

It is important to notice that the portfolios obtaining high Sharpe ratios in all the datasets are those that have a low turnover, i.e. are rebalanced less frequently. In the dataset EMU1, the highest Sharpe ratio is obtained by the minimum variance portfolio with the 60-day estimation window and threshold rebalancing; it has a turnover of 5.5 times the total wealth, not much higher than the benchmark strategy's turnover of 3.4. In the datasets EMU2, BEM1, and BEM2, the highest Sharpe ratios are obtained with quarterly rebalanced portfolios that have turnovers ranging from 7.9 to 32.8.

The reason why weekly rebalanced portfolios performed poorly is visible in Table 11: the turnover amounts for those portfolios are exceptionally high. A portfolio that turns over its wealth 50 or even 200 times during the four-year investment period cannot be successful when trading costs are present.

Another important thing to notice from Table 11 is that the turnover amounts are larger the shorter the estimation window is. For example, the monthly rebalanced portfolios with the 250-day estimation window have turnover amounts ranging between 10.7 and 17.3. However, for monthly rebalanced portfolios with the 20-day estimation window, the corresponding range is between 67.6 and 90.8. These portfolios are rebalanced at exactly the same frequency, yet turnover is about six times higher for the portfolios with the shorter estimation window. This is implies that the shorter estimation window must be producing more unstable portfolio weights, requiring more trading to take place upon rebalancing.

To sum up, the minimum variance portfolios clearly outperform the benchmark strategy in the datasets EMU1 and EMU2. In the datasets BEM1 and BEM2, the benchmark strategy performs better than the minimum variance portfolios but the differences are not significant in most cases. The reason behind this performance variation between the datasets could be at least partly attributed to the higher transaction costs assumed for the datasets BEM1 and BEM2, since the turnover amounts of the minimum variance portfolios are generally much higher than those of the benchmark strategy.

4.1.2 Constant-correlation shrinkage portfolios

The out-of-sample Sharpe ratios of the minimum variance portfolios that are constructed utilizing the constant-correlation shrinkage model introduced in Ledoit and Wolf (2004) are presented in Table 12. The p-values for the differences between the Sharpe ratios of the benchmark strategy and the minimum variance portfolios are reported in parentheses below each Sharpe ratio. Similar to the sample covariance matrix portfolios, the shrinkage portfolios obtain higher Sharpe ratios than the benchmark strategy for all but one combination of estimation window length and rebalancing frequency in the datasets EMU1 and EMU2.

Table 12 Out-of-sample Sharpe ratios of the constant-correlation shrinkage portfolios

The table presents the out-of-sample Sharpe ratios for the benchmark strategy and for each combination of estimation window length and rebalancing frequency for minimum variance portfolios constructed using the constant-correlation shrinkage estimator in portfolio optimization. In parentheses below each Sharpe ratio is the p-value for the difference between the above Sharpe ratio and the Sharpe ratio of the benchmark strategy. The highest Sharpe ratio of the minimum variance portfolios in each dataset is in bold text.

Strategy	**EMU1**	**EMU2**	**BEM1**	**BEM2**
Benchmark				
1/N rebalanced weekly	0.691	-0.214	1.397	0.619
Constant-correlation shrinkage portfolios				
Estimation window = 250 days				
Rebalanced quarterly	1.345	0.162	1.392	0.588
	(0.05)	(0.12)	(0.99)	(0.88)
Rebalanced monthly	1.303	0.153	1.377	0.559
	(0.07)	(0.15)	(0.94)	(0.78)
Rebalanced weekly	1.301	0.128	1.362	0.515
	(0.08)	(0.18)	(0.89)	(0.63)
Estimation window = 60 days				
Rebalanced quarterly	1.314	0.430	1.135	0.655
	(0.05)	(0.02)	(0.30)	(0.87)
Rebalanced monthly	1.161	**0.436**	1.245	0.466
	(0.17)	(0.02)	(0.55)	(0.53)
Rebalanced weekly	1.067	0.283	1.287	0.268
	(0.29)	(0.08)	(0.68)	(0.15)
Threshold rebalancing	**1.404**	0.118	**1.430**	**0.733**
	(0.05)	(0.22)	(0.88)	(0.58)
Estimation window = 20 days				
Rebalanced quarterly	1.072	0.420	0.978	0.581
	(0.20)	(0.01)	(0.10)	(0.87)
Rebalanced monthly	0.974	0.195	1.172	0.205
	(0.41)	(0.13)	(0.39)	(0.11)
Rebalanced weekly	0.676	0.027	1.022	-0.229
	(0.97)	(0.41)	(0.15)	(0.00)
Threshold rebalancing	1.352	0.299	1.202	0.580
	(0.06)	(0.08)	(0.48)	(0.87)

The highest Sharpe ratio in the dataset EMU1, 1.404, is obtained with the 60-day estimation window and threshold rebalancing. It has a p-value of 0.05, signaling that the difference to the Sharpe ratio of the benchmark strategy is statistically significant. In the dataset EMU2, the highest Sharpe ratio is 0.436 which is achieved by the monthly rebalanced portfolio with the 60-day estimation window. The difference to the Sharpe ratio of the benchmark portfolio is statistically significant with a p-value of 0.02. As is the case with the sample covariance matrix portfolios, the shrinkage portfolios perform, in general, slightly worse than the benchmark strategy in the datasets BEM1 and BEM2. However, for most minimum variance portfolios, the difference to the benchmark's Sharpe ratio is small and the corresponding p-values denote statistical insignificance.

It is interesting to note that the Sharpe ratios in Table 12 are lower than those of the sample covariance matrix portfolios presented in Table 10 for all combinations of estimation window length and rebalancing frequency in the dataset EMU1. Conversely, exactly the opposite is true in the dataset EMU2: the shrinkage portfolios outperform the sample covariance matrix portfolios with every combination. In the dataset BEM1, the results are more mixed. In the dataset BEM2, the shrinkage portfolios perform better than the sample covariance matrix portfolios with most combinations of estimation window length and rebalancing frequency.

Weekly rebalancing produced the lowest Sharpe ratios for the shrinkage portfolios for all estimation window lengths in most of the datasets. This is again a sign of the fact that frequent rebalancing causes the total trading costs to go up, leading to poor performance. Furthermore, in the datasets EMU1 and BEM2, quarterly rebalancing produces better results than monthly rebalancing for all estimation window lengths. In the datasets EMU2 and BEM1, the results are more mixed with the monthly and quarterly rebalancing intervals.

As is evident from Table 12, threshold rebalancing produced very good results for the shrinkage portfolios: In three of the datasets, the highest Sharpe ratios are achieved with the threshold rebalancing criterion. Figure 4 illustrates threshold rebalancing in action. It produces a portfolio that is rebalanced relatively infrequently, only when keeping the current portfolio weights would result in a significantly higher estimated volatility than the new optimal weights.

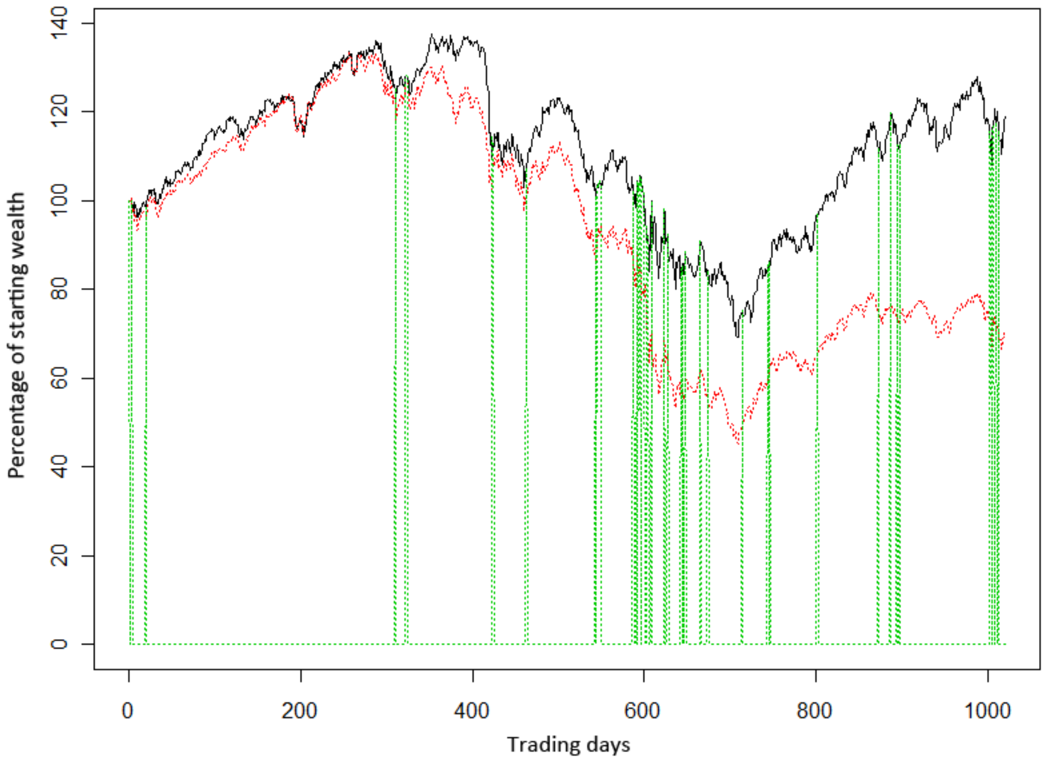

Figure 4 Performance of the threshold rebalancing criterion

The figure plots the development of wealth during the investment period for the constant-correlation shrinkage portfolio with the 20-day estimation window and threshold rebalancing (solid line) and the benchmark portfolio (dashed line) in the dataset EMU2. The vertical lines mark the days when the shrinkage portfolio is rebalanced.

Under the threshold rebalancing, the portfolio can stay without rebalancing for very long periods of time, such as the shrinkage portfolio during the first 300 trading days in Figure 4, where the vertical lines signal the rebalancing days of the portfolio. This resembles a buy-and-hold strategy with portfolio composition being monitored constantly and the investor ready to take action in case new information causes a need for updating.

According to Table 12, the portfolio with the 60-day estimation window and threshold rebalancing obtains a higher Sharpe ratio than the benchmark strategy in all four datasets. However, the Sharpe ratio difference is statistically significant in only one of the datasets. The 20-day estimation window appears to produce the worst Sharpe ratios in general, possibly due to unstable portfolio weights leading to increased trading costs. To get a better idea of the amount of trading, the turnover amounts of the shrinkage portfolios are reported in Table 13.

Table 13 Turnover amounts of the constant-correlation shrinkage portfolios

The table presents the total turnover amounts for the benchmark strategy and for each combination of estimation window length and rebalancing frequency for minimum variance portfolios constructed using the constant-correlation shrinkage estimator in portfolio optimization. The turnover amount is computed according to Equation 50.

Strategy	**EMU1**	**EMU2**	**BEM1**	**BEM2**
Benchmark				
1/N rebalanced weekly	3.4	3.9	6.0	6.1
Constant-correlation shrinkage portfolios				
Estimation window = 250 days				
Rebalanced quarterly	7.7	9.3	6.8	8.7
Rebalanced monthly	10.3	14.6	8.9	11.9
Rebalanced weekly	16.0	23.4	15.3	20.0
Estimation window = 60 days				
Rebalanced quarterly	16.8	18.9	15.0	16.7
Rebalanced monthly	27.4	31.5	22.3	26.0
Rebalanced weekly	45.4	52.9	40.6	48.6
Threshold rebalancing	2.1	11.4	1.1	4.2
Estimation window = 20 days				
Rebalanced quarterly	19.5	22.2	18.0	20.6
Rebalanced monthly	49.7	55.1	45.7	53.4
Rebalanced weekly	100.1	107.6	90.6	106.3
Threshold rebalancing	8.4	33.3	7.2	14.5

The shrinkage portfolios have, in general, lower turnover amounts than the sample covariance matrix portfolios. This implies that the shrinkage estimation produces more stable portfolio weights, which, in turn, results in lower trading costs. The threshold rebalancing portfolio with the 60-day estimation window that obtained the highest Sharpe ratios in three of the datasets exhibits very low turnover: In the datasets EMU1 and BEM1, its turnover amounts were 2.1 and 1.1 times the total wealth, respectively. This implies that the portfolio was rebalanced very few times during the investment periods.

Table 13 reveals that the turnover amounts of the weekly rebalanced portfolios are very high, ranging from 90.6 to 107.6 for portfolios with the 20-day estimation window. This explains their poor performance as the total trading costs for the portfolios are enormous due to the high turnover amounts. It is important to notice that the turnover amounts are larger for portfolios with shorter estimation windows in every dataset. For example, the monthly rebalanced portfolios with the 250-day estimation window exhibit turnover amounts ranging between 8.9 and 14.6. For monthly rebalanced portfolios with the 20-day estimation window, the range is between 45.7 and 55.1. The portfolios are rebalanced with the exact same frequency, however, the turnover is about four times higher for the portfolios with the shorter estimation window. This implies that the shorter estimation window produces more unstable portfolio weights, which results in more trading to rebalance the portfolios to up-to-date optimal weights.

In general, the constant-correlation shrinkage portfolios exhibit slightly more stable performance compared to the sample covariance matrix portfolios. This can be attributed at least partly to the lower turnover amounts of the shrinkage portfolios. To sum up the performance relative to the benchmark strategy, the shrinkage portfolios clearly outperform the benchmark in the datasets EMU1 and EMU2, whereas in the datasets BEM1 and BEM2, the benchmark strategy performed slightly better than most of the minimum variance portfolios. In this regard, the shrinkage portfolios and the sample covariance matrix portfolios feature similar performance.

4.1.3 DCC-GARCH portfolios

In Table 14, the out-of-sample Sharpe ratios of the DCC-GARCH portfolios and the p-values for the Sharpe ratio differences compared to the benchmark are reported. Since the DCC-GARCH model utilized in this thesis attempts to forecast the future covariance matrix for the next trading day, its performance can be assumed to be based on very frequent updating. Under the threshold rebalancing, a new optimal composition for the portfolio is calculated every trading day, but the portfolio is not rebalanced to the new weights very frequently because of the trading costs that would be incurred. For this reason, we conducted an additional performance evaluation with trading costs set to zero and the portfolio updated daily, to find out if the DCC-GARCH model can provide superior performance when it is allowed to work without limitations. The performance of this additional portfolio, free of trading costs, is also reported in Table 14.

Table 14 Out-of-sample Sharpe ratios of the DCC-GARCH portfolios

The table presents the out-of-sample Sharpe ratios for the benchmark strategy and for minimum variance portfolios constructed using the DCC-GARCH model in portfolio optimization. In parentheses below each Sharpe ratio is the p-value for the difference between the above Sharpe ratio and the Sharpe ratio of the benchmark strategy. In addition, the Sharpe ratios for daily rebalanced DCC-GARCH portfolios without trading costs are also reported.

Strategy	**EMU1**	**EMU2**	**BEM1**	**BEM2**
Benchmark				
1/N rebalanced weekly	0.691	-0.214	1.397	0.619
DCC-GARCH portfolios				
Estimation window = 250 days				
Threshold rebalancing	1.446	0.072	0.992	0.373
	(0.02)	(0.20)	(0.15)	(0.36)
Estimation window = 60 days				
Threshold rebalancing	0.695	-0.339	0.902	0.107
	(0.99)	(0.59)	(0.09)	(0.06)
Without trading costs				
Estimation window = 250 days				
Rebalanced daily	1.481	0.177	1.606	0.275

From Table 14, it is evident that the estimation window length has a crucial role in the performance of the DCC-GARCH portfolios: the 250-day estimation window produced better results than the 60-day window in all four datasets. As was the case with most sample covariance matrix and shrinkage portfolios, the DCC-GARCH portfolio with the 250-day estimation window obtained a higher Sharpe ratio than the benchmark strategy in the datasets EMU1 and EMU2, but performed worse than the benchmark in the datasets BEM1 and BEM2. The difference to the Sharpe ratio of the benchmark strategy is statistically significant only for the portfolio with the 250-day estimation window in the dataset EMU1.

The DCC-GARCH portfolio without trading costs performs better than the portfolios with trading costs in three of the four datasets. However, compared to the portfolio with the 250-day estimation window, the difference in performance is noticeable only in the dataset BEM1, where moving to daily updating without trading costs increased the Sharpe ratio to 1.606. Overall, it appears that the DCC-GARCH model did not manage to forecast the future covariance matrix any better than the sample covariance matrix or the shrinkage estimator. In order to get a better idea of the stability of the portfolio weights and the effect of trading costs, the turnover amounts of the DCC-GARCH portfolios are reported in Table 15.

Table 15 Turnover amounts of the DCC-GARCH portfolios

The table presents the total turnover amounts for the benchmark strategy and for minimum variance portfolios constructed using the DCC-GARCH model in portfolio optimization. The turnover amount is computed according to Equation 50.

Strategy	**EMU1**	**EMU2**	**BEM1**	**BEM2**
Benchmark				
1/N rebalanced weekly	3.4	3.9	6.0	6.1
DCC-GARCH portfolios				
Estimation window = 250 days				
Threshold rebalancing	13.7	66.3	7.8	31.3
Estimation window = 60 days				
Threshold rebalancing	57.5	178.0	33.6	78.3

The turnover amounts in Table 15 reveal that the DCC-GARCH model produced seemingly unstable portfolio weights, especially for the portfolio with the 60-day estimation window. The turnover for this portfolio was at worst over 40 times larger than that of the benchmark strategy, despite the fact that the threshold rebalancing criterion was employed. In general, the turnover amounts of the DCC-GARCH portfolios are much larger compared to the corresponding amounts of the sample covariance matrix and shrinkage portfolios with threshold rebalancing. This is a sign of very unstable portfolio weights, which can be somewhat expected from a model that attempts to forecast the covariance matrix for the next trading day.

The large differences in the performance and turnover amounts between the DCC-GARCH portfolios with 250-day and 60-day estimation windows are likely due to variation in the accuracy of the maximum likelihood method used to estimate the parameters for the DCC-GARCH model. The maximum likelihood estimation can be expected to produce more accurate estimates when the estimation window is longer.

Overall, the DCC-GARCH portfolios did not perform better than the sample covariance matrix or shrinkage portfolios. In fact, they performed slightly worse overall than the minimum variance portfolios based on either of the other two estimation models. Nevertheless, the DCC-GARCH portfolio with the 250-day estimation window still managed to outperform the benchmark strategy in the datasets EMU1 and EMU2, similarly to the sample covariance matrix and shrinkage portfolios.

4.2 Additional considerations

4.2.1 Average performance across the datasets

For all estimation models tested, most of the minimum variance portfolios clearly outperformed the benchmark strategy in the datasets EMU1 and EMU2. However, in the datasets BEM1 and BEM2, the benchmark strategy produced slightly better results. When searching for a portfolio strategy that performs better than the benchmark in all types of market conditions, it is reasonable to compare the average performance of the portfolios across the four datasets. To achieve this, we calculate and compare the average Sharpe ratios of the minimum variance portfolios for all tested combinations of estimation window length and rebalancing frequency. A summary of the best sample covariance matrix, shrinkage, and DCC-GARCH portfolios in terms of the average out-of-sample Sharpe ratio is presented in Table 16.

Table 16 Average Sharpe ratios and turnover amounts of selected portfolios

The table presents the arithmetic means of Sharpe ratios and turnover amounts for the benchmark strategy and the best performing minimum variance portfolios across all four datasets.

Strategy		Avg. Sharpe ratio	Avg. Turnover
Benchmark			
	1/N rebalanced weekly	0.623	4.9
Constant-correlation shrinkage portfolios			
60-day window	With threshold rebalancing	0.921	4.7
60-day window	With quarterly rebalancing	0.883	16.9
250-day window	With quarterly rebalancing	0.872	8.1
Sample covariance matrix portfolios			
60-day window	With quarterly rebalancing	0.862	22.1
60-day window	With monthly rebalancing	0.837	36.5
250-day window	With quarterly rebalancing	0.815	9.4
DCC-GARCH portfolios			
250-day window	With threshold rebalancing	0.721	29.8

As Table 16 shows, the three best performing portfolios overall are shrinkage portfolios rebalanced infrequently with medium or long estimation window lengths. The shrinkage portfolio with the 60-day estimation window and threshold rebalancing has the highest average out-of-sample Sharpe ratio of 0.921 which is significantly higher than the benchmark strategy's average Sharpe ratio of 0.623. This shrinkage portfolio has an average turnover of 4.7 times the total wealth which is slightly smaller than that of the benchmark strategy. Moreover, all three shrinkage portfolios in Table 16 outperform the best sample covariance matrix portfolios in terms of the average out-of-sample Sharpe ratio. The average Sharpe ratio of the DCC-GARCH portfolio is lower than those of the best shrinkage and sample covariance matrix portfolios, but still higher than the average Sharpe ratio of the benchmark strategy.

Overall, the 60-day estimation window appears to be the best out of the three estimation window lengths tested for the sample covariance matrix and shrinkage portfolios. Based on this observation, it is possible that the estimation window of 250 days is subject to more error due to nonstationarity in the parameters, as described in Broadie (1993, 21–22). The 20-day estimation window, on the other hand, may be too short for parameter estimation and is likely to produce unstable portfolio weights.

From Table 16, it is evident that the average turnover amounts of the minimum variance portfolios are higher than the average turnover of the benchmark strategy, with the exception of the shrinkage portfolio with threshold rebalancing. Moreover, the shrinkage portfolios feature lower average turnover amounts than the sample covariance matrix portfolios, which is likely part of the reason for their superior performance.

When evaluating the minimum variance portfolios across all four datasets, Table 16 provides evidence that it is possible to outperform the benchmark strategy with all three estimation models tested if the estimation window length and rebalancing interval are correctly set. It appears that determining the appropriate estimation window length and rebalancing frequency is more important than choosing the best covariance matrix estimation model.

4.2.2 Standard deviations of the portfolios

So far the main interest has been in the Sharpe ratios of the portfolios, but it should be emphasized that the minimum variance portfolio optimization aims to construct a portfolio with the lowest possible standard deviation. In other words, the optimization algorithm is not concerned with the expected return that plays an important role in the Sharpe ratio. Furthermore, it is worth mentioning that the Sharpe ratio is not a consistent measure if the mean return of the portfolio has a negative value. In this case, higher standard deviation actually results in improved Sharpe ratios, which is exactly the opposite of what the investor prefers.

While the Sharpe ratios of most minimum variance portfolios in the datasets BEM1 and BEM2 were lower than those of the benchmark strategy, a closer examination of the standard deviations reveals that the minimum variance optimization did in fact produce portfolios with lower standard deviations than the benchmark strategy. For example, in the dataset BEM1, the daily standard deviation of the benchmark strategy is 0.0105. For comparison, the standard deviations for most minimum variance portfolios are between 0.0092 and 0.0100.

In the dataset BEM2, the benchmark strategy has a standard deviation of 0.0161, whereas most of the minimum variance portfolios exhibit standard deviations between 0.0135 and 0.0140. In contrast, these portfolios feature significantly lower mean returns than the benchmark strategy and thus had their Sharpe ratio downgraded. Considering these observations, the minimum variance portfolio optimization actually performed as it is supposed to, producing lower standard deviations than the benchmark strategy for most portfolios in all four datasets.

4.3 Optimal length of the estimation window

Throughout the empirical results, estimation window length has been a key factor affecting portfolio performance. It seems that the estimation window length is more important than selecting the best parameter estimation model. The estimation window length appears to be so important that it is worth examining more carefully, separate from portfolio weights and trading costs. For this reason, we conduct an additional ex post study that aims to find out what the optimal estimation window length is for each dataset and how different lengths affect the accuracy of covariance matrix estimation.

To investigate the performance of the estimation window, we develop a mathematical model to examine the average difference between estimated sample covariance matrixes and the realized covariance matrixes that the estimates attempt to forecast. We define function d to measure this difference as follows:

$$d(W,F) = \frac{1}{T-W} \sum_{t=1}^{T-F} \boldsymbol{\iota}' \left| \boldsymbol{\Sigma}_{t+1,t+F} - \boldsymbol{\Sigma}_{t-W,t} \right| \boldsymbol{\iota}, \tag{51}$$

where W denotes the length of the estimation window, F denotes the length of the future period that is forecasted, T denotes the number of trading days in the dataset, $\boldsymbol{\iota}$ denotes a vertical vector of ones, and $\boldsymbol{\Sigma}_{a,b}$ denotes the sample covariance matrix estimated from the data spanning the time period from a through b. By feeding the function $d(W,F)$ all discrete estimation window lengths between 5 and 250, we can plot the values of the function to better understand how the length of the window affects the accuracy of the forecast. Figure 5 illustrates the values of the function for $F = 5$, that is, forecasting the covariance matrix for the next five days.

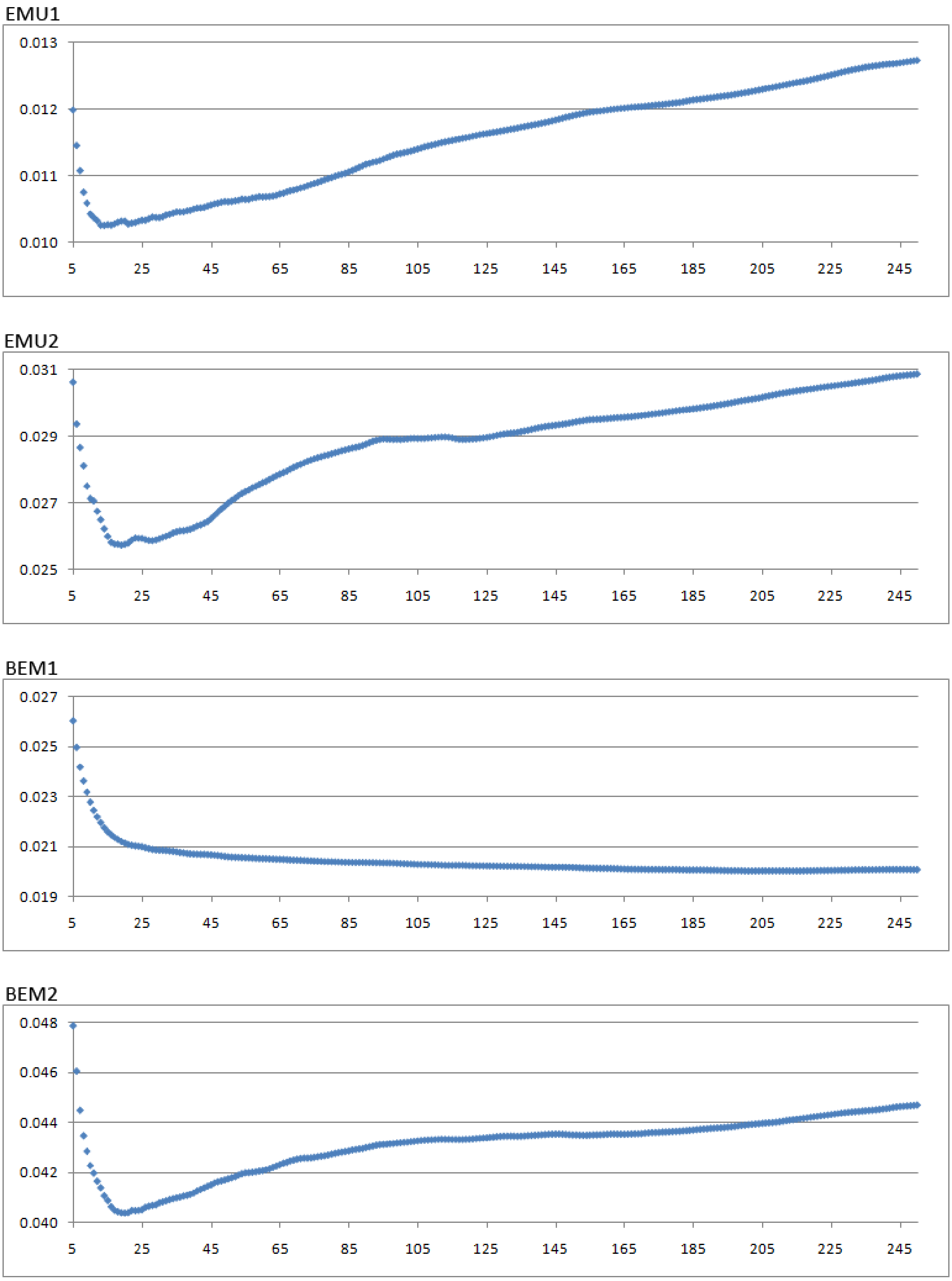

Figure 5 Optimal estimation window length for forecasting one week ahead

The values of the function $d(W, F)$ are plotted for $F = 5$, for each dataset. In essence, the figure displays the average error magnitudes in forecasting the covariance matrix for the next week (5 days) throughout the datasets.

In Figure 5, the values of the function $d(W,F)$ are on the vertical axis and the values of the estimation window length W are on the horizontal axis. The lower the value of the function, the smaller the average difference between the forecasting and the realized covariance matrixes. Thus, a low value indicates that the estimation window length produces a good forecast, on average, relative to window lengths with higher function values. A function value of zero would mean perfect forecasts throughout the dataset.

It is important to notice that the datasets have different optimal estimation window lengths. What is striking is that for the datasets EMU1, EMU2, and BEM2, the curves in Figure 5 are upward sloping. This means, in general, that the shorter the estimation window is, the more accurate are the forecasts. The optimal length of the estimation window for the dataset EMU1 is 14 days. For the dataset EMU2, the optimal window length is 19 days, and for the dataset BEM2 it is 20 days. Longer estimation windows than these provide worse forecasts of the future covariance matrixes for the next five days throughout the datasets. It is interesting that for the datasets EMU1 and EMU2, a 250-day estimation window produces worse forecasts than a 5-day estimation window that is likely to contain a large amount of estimation error due to its lack of observations. This implies that the covariances are not stationary.

The dataset BEM1 features a different curve than the other datasets in Figure 5. The curve is slowly declining, meaning that a longer estimation window produces a more accurate forecast. This is also evident in the empirical results for the minimum variance portfolios, where the 250-day estimation window produced the best results, on average, in the dataset BEM1. The exact optimal window length for the dataset BEM1 is 202 days but the differences in function values are very small around the optimal length.

It is important to notice that the function values are of different magnitude for each dataset. For example, the dataset EMU1 has the lowest values, ranging between 0.010 and 0.013. For the dataset EMU2, the function values are over twice as large, ranging between 0.025 and 0.031. This implies that the future covariances are easier to estimate from historical data in the dataset EMU1 than in the dataset EMU2, and the same applies between the datasets BEM1 and BEM2. Furthermore, the datasets BEM1 and BEM2 feature higher function values than their parallel datasets EMU1 and EMU2, however, they include one more asset, which increases the values.

Following the rebalancing intervals in the empirical study, we also examine the values of $d(W,F)$ for the future periods of 20 days and 60 days. In Figure 6, the values of the function $d(W,F)$ are calculated for $F = 20$, which gives insight on what the optimal estimation window length is for monthly rebalancing.

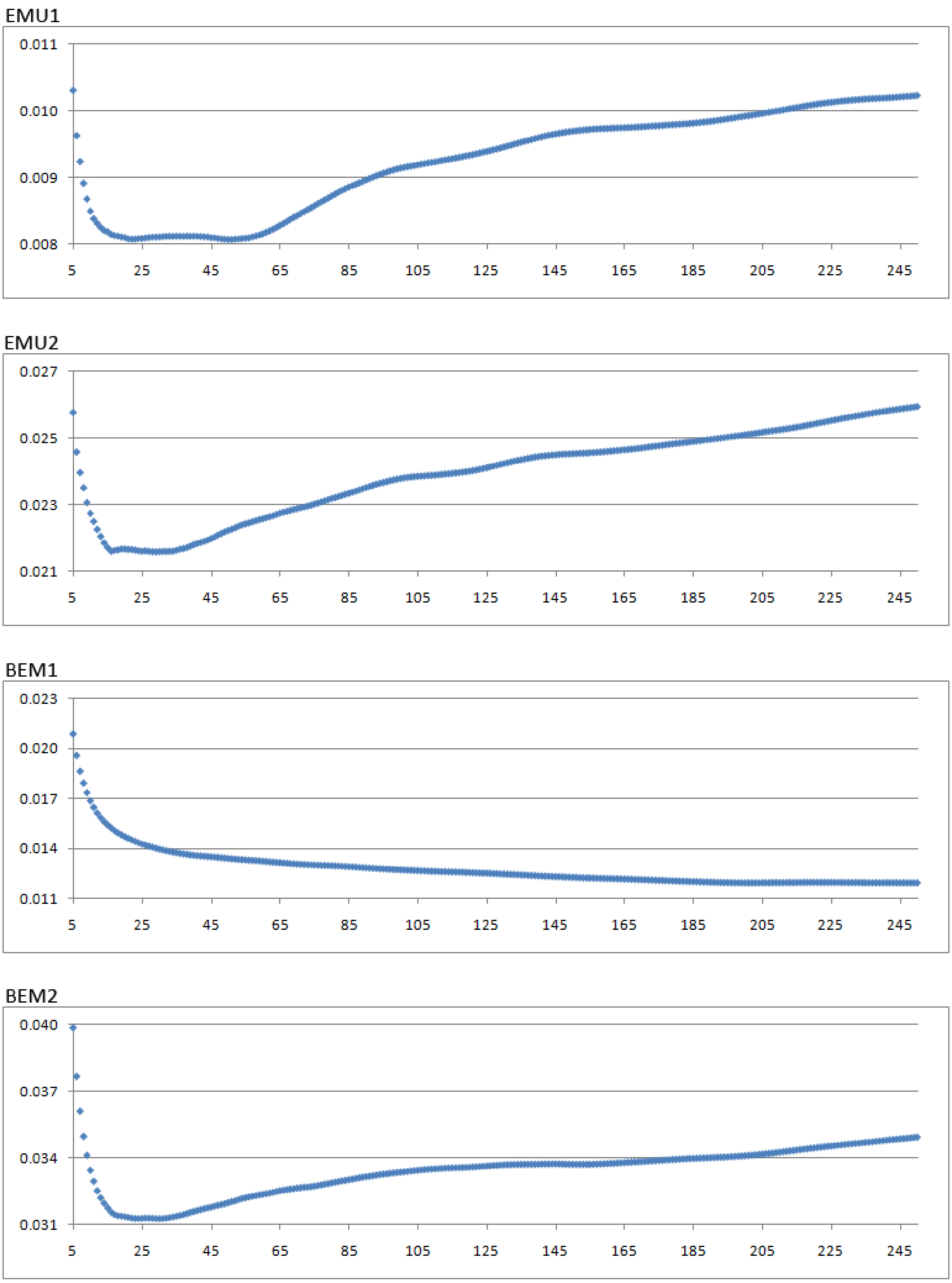

Figure 6 Optimal estimation window length for forecasting one month ahead

The values of the function $d(W, F)$ are plotted for $F = 20$, for each dataset. In essence, the figure displays the average error magnitudes in forecasting the future covariance matrix for the next month (20 days) throughout the datasets.

There is a similar trend in the graphs in Figure 6 as in the ones in Figure 5; the curve is upward sloping for the datasets EMU1, EMU2, and BEM2, but downward sloping for the dataset BEM1. The optimal length of the estimation window for the dataset EMU1 is 51 days, although the differences in function values for the lengths between 15 and 55 days are very small. A longer estimation window than that produces less accurate forecasts for the future period of one month. Moreover, windows shorter than 15 days produce rapidly worse forecasts. A similar trend is observed for the dataset EMU2, with the optimal estimation window length being 29 days. An estimation window of 250 days produces approximately equally accurate forecasts as an estimation window of 5 days for the datasets EMU1 and EMU2, which is striking and a sign of nonstationary parameters.

For the datasets BEM1 and BEM2, the curves in Figure 6 are more steadily sloping than the curves for the datasets EMU1 and EMU2. For the dataset BEM1, an estimation window of 201 days produces the most accurate forecasts, which is approximately the same optimal length as for forecasting one week ahead. For the dataset BEM2, the optimal estimation window length is 30 days and the accuracy of forecasts declines slowly as the length is increased.

The values of the function $d(W,F)$ for each dataset are lower for $F = 20$ than for $F = 5$, meaning that the covariance matrix can be forecasted more accurately for the future period of one month than for the future period of one week. Furthermore, the function values in Figure 6 are higher for the dataset EMU2 than for the dataset EMU1, as was the case in Figure 5. This means that the covariance matrixes are more difficult to forecast in the dataset EMU2 than in the dataset EMU1, and the same applies between the datasets BEM2 and BEM1.

Overall, the optimal estimation window lengths are slightly longer when forecasting the covariance matrix for the next 20 days, than when forecasting for the next 5 days. In Figure 7, the function $d(W,F)$ is plotted for $F = 60$, to examine the optimal estimation window lengths for the longest rebalancing interval in the empirical study, or 60 days.

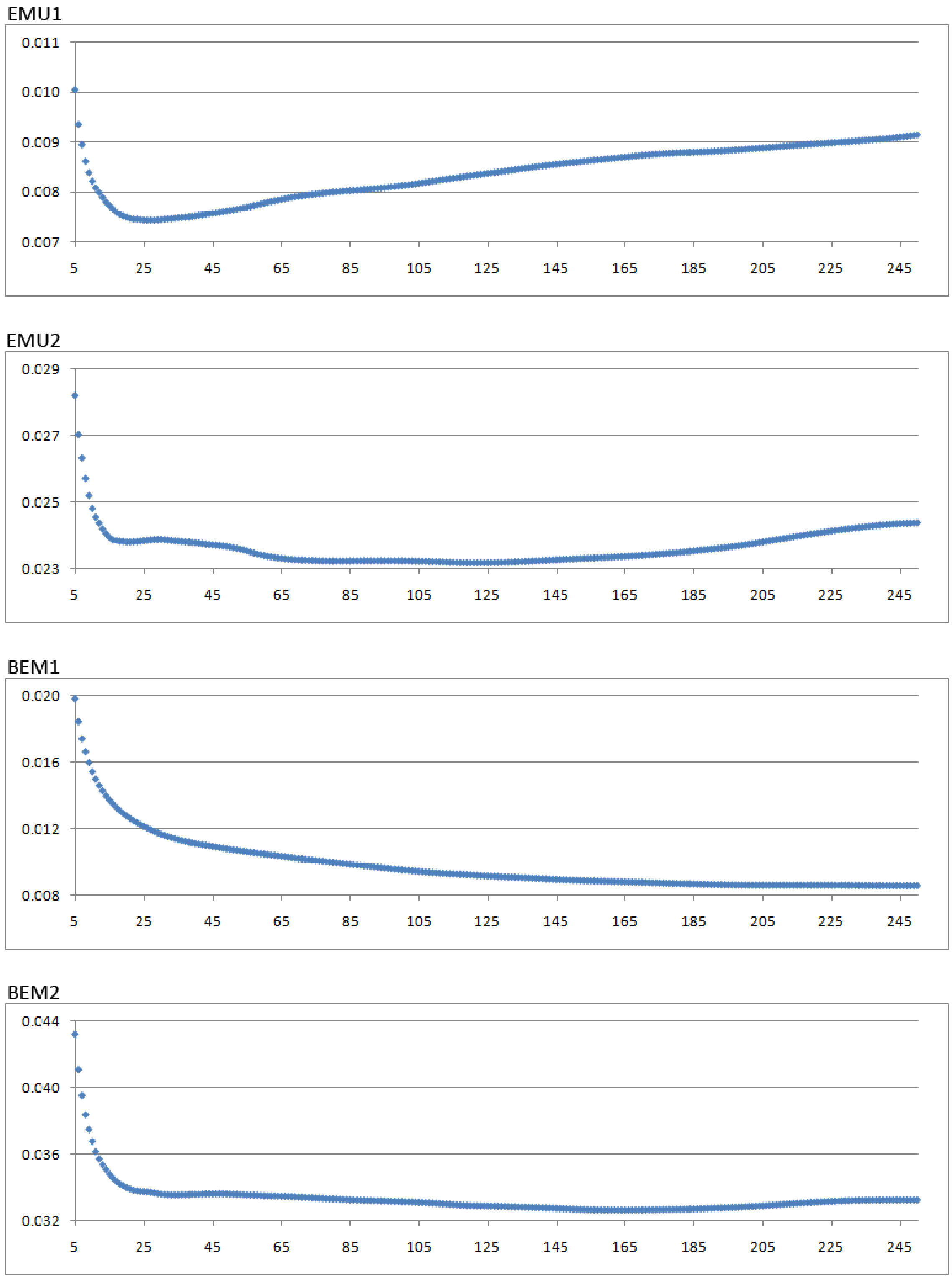

Figure 7 Optimal estimation window length for forecasting one quarter ahead

The values of the function $d(W, F)$ are plotted for $F = 60$, for each dataset. In essence, the figure displays the average error magnitudes in forecasting the future covariance matrix for the next quarter (60 days) throughout the datasets.

From Figure 7, it is evident that the trends of the curves observed in Figure 5 and Figure 6 are not as strong anymore: the curves are flatter for most of the datasets. However, the dataset EMU1 still features an upward sloping curve, and surprisingly, the optimal length of the estimation window is 28 days, forecasting for the future period of 60 days. The dataset EMU2 features a somewhat flat curve, with an optimal window length of 123 days and not much difference in the forecast accuracy around the optimal length. The dataset BEM1 still has a downward sloping curve and the optimal estimation window length is 246 days. The dataset BEM2 features a relatively flat curve with an optimal length of 162 days. Overall, the optimal estimation window length appears to be longer when forecasting the covariance matrix for the next 60 days, except for the dataset EMU1, than when forecasting for the next 20 days.

The function values for the optimal window lengths for $F = 60$ are slightly higher for the datasets EMU2 and BEM2 than they are for $F = 20$, meaning that for these datasets, it is more difficult to forecast the covariance matrix for the future period of one quarter than it is for the future period of one month. On the other hand, the opposite is true for the datasets EMU1 and BEM1; for these datasets it is easier to forecast the covariance matrix for the longer future period. The function values in Figure 7 reveal that, as is the case with the shorter forecast periods, the covariance matrix appears to be more difficult to forecast in the dataset EMU2 than in the dataset EMU1, and the same applies between the datasets BEM2 and BEM1.

The most important finding of this additional study is that the optimal estimation window length is different for each dataset and rebalancing interval. What is remarkable is that a shorter estimation window performs in most cases better than an extensively long estimation window. This is a surprising result and is in contrast with previous studies such as DeMiguel et al. (2009b, 1942–1943) presenting that a longer estimation window should produce better results. On the other hand, our results support the view presented in Broadie (1993, 21–22) that a long estimation window is exposed to error caused by nonstationary parameters.

5 CONCLUSION

The thesis set out to investigate whether it is, contrary to most previous research, possible to outperform a simple benchmark strategy with mean-variance portfolio optimization when covariance matrix estimation and portfolio rebalancing are given appropriate attention. Furthermore, we pursued to find out what are the most important factors affecting portfolio performance. In a broader context, the thesis aimed to shed light on the debated topic: the feasibility of mean-variance optimization in practice. The thesis improved on some of the shortcomings in previous research by using daily return series for parameter estimation instead of monthly returns, and taking transaction costs into account in evaluating portfolio performance across several datasets. Furthermore, we delved deeper into the effects of estimation window length and rebalancing intervals on portfolio performance.

In the empirical part of the thesis, we evaluated the performance of minimum variance portfolios against a simple benchmark strategy that allocates wealth evenly across assets. For the minimum variance portfolios, we employed three covariance matrix estimation models: the sample covariance matrix estimator, the constant-correlation shrinkage estimator of Ledoit and Wolf (2004), and the multivariate DCC-GARCH model introduced in Engle (2002). These models were applied with combinations of different estimation window lengths and rebalancing intervals, including a special threshold rebalancing criterion.

The data used in the empirical study were daily returns of international stock market indexes, divided into four datasets to examine portfolio performance in different market conditions. The performance evaluation of the portfolios was done by comparing their out-of-sample Sharpe ratios to that of the benchmark strategy. Furthermore, the total turnover amounts of the portfolios during the investment period were observed.

The results indicate that with appropriate combinations of estimation window length and rebalancing interval, it is possible to outperform the simple equal-weight strategy with minimum variance portfolios. However, the results varied among the datasets: In two of the datasets, the minimum variance portfolios clearly outperformed the benchmark and the differences in performance were statistically significant in several cases. In the two other datasets, the benchmark strategy performed better than most minimum variance portfolios. However, the differences in performance were not statistically significant in general.

The shrinkage estimator portfolios performed, on average, better than the sample covariance matrix portfolios. They also featured lower turnover amounts, implying that their asset weights were more stable, resulting in less trading costs. The DCC-GARCH portfolios did not live up to expectations and actually performed slightly worse than the sample covariance and shrinkage portfolios. The DCC-GARCH portfolios had high turnover amounts even with threshold rebalancing, signifying that the model produced very unstable weights.

The empirical results indicate that the rebalancing interval is an important factor affecting portfolio performance: minimum variance portfolios rebalanced most often, that is weekly, produced the worst results in most cases. In contrast, the best performing portfolios were rebalanced relatively infrequently. This shows the importance of transaction costs in portfolio performance. The threshold rebalancing criterion worked fairly well, especially for the shrinkage portfolios. Utilizing the criterion in rebalancing, we obtained one portfolio that outperformed the benchmark in all four datasets.

Likewise, the importance of estimation window length was evident in the results. A short estimation window increased the turnover amounts of the portfolios tremendously, implying that shorter windows produce more unstable portfolio weights. In general, the medium and long estimation window lengths were associated with the best portfolio performance.

When the average performance of the portfolios across all datasets was compared, the best minimum variance portfolios clearly outperformed the benchmark strategy. These average performance results demonstrate that it is possible to outperform the benchmark with all the covariance matrix estimation models tested, as long as the estimation window length and rebalancing interval are carefully selected. Strikingly, it appears that defining the estimation window length and rebalancing frequency is more important than selecting the best estimation model.

Since the estimation window length appears to be crucial for portfolio performance, we conducted an additional ex post study to examine the effects of estimation window length separate from portfolio weights and transaction costs. To achieve this, we developed a mathematical model to measure the difference between an estimated sample covariance matrix and the realized covariance matrix that the estimate attempts to forecast. We measured this difference throughout the datasets to define the optimal estimation window length for each dataset and rebalancing interval.

Interestingly, it was found that the optimal estimation window length is different for each dataset and rebalancing interval. A striking result from this additional study is that in most cases, a shorter estimation window produces more accurate estimates of the future. Our results support the view presented in Broadie (1993) that a longer estimation window is subject to error due to nonstationarity in the parameters. Furthermore, the covariance matrix appears to be clearly more difficult to forecast in some datasets than in other datasets.

Overall, our results provide interesting insights into the factors affecting portfolio performance and the efficiency of mean-variance optimization. In general, our results are in contrast with previous studies, such as DeMiguel et al. (2009b), advocating the impracticality of mean-variance optimization. Furthermore, contrary to most previous research, our performance results are inclusive of transaction costs, which strengthens the importance of the results. We presume our different results are due to more attention given to the estimation window length and rebalancing frequency, and the use of more information-rich daily return data.

In conclusion, the thesis has shown that despite the critique mean-variance optimization has received, it still seems an attractive choice over a simple allocation strategy. However, it should be noted that the varying characteristics of available assets and different market conditions imply that a universal model that performs best in all situations might not exist. What works in certain market conditions might be off track in other circumstances. Therefore, the empirical results should not be thought of as robust or universally applying.

Several intriguing topics for further research arise from the results. For example, the optimal length of the estimation window can be examined in more detail and be incorporated into the parameter estimation process. Furthermore, the contrast between a short window producing the most accurate forecasts and the fact that it also produces unstable portfolio weights needs to be solved by applying more sophisticated rebalancing methods. Finally, the robustness of different parameter estimation models, estimation window lengths, and rebalancing intervals could be examined in more detail when various market conditions prevail.

REFERENCES

Becker, Franziska – Gürtler, Marc – Hibbeln, Martin (2009) Markowitz versus Michaud: Portfolio optimization strategies reconsidered. Working paper available at SSRN: http://ssrn.com/abstract=1397488.

Behr, Andreas – Poetter, Ulrich (2009) Modeling marginal distributions of ten European stock market index returns. *International Research Journal of Finance and Economics*, Vol. 28, 104–119.

Benartzi, Shlomo – Thaler, Richard H. (2001) Naive diversification strategies in defined contribution saving plans. *The American Economic Review*, Vol. 91 (1), 79–98.

Best, Michael J. – Grauer, Robert R. (1991) On the sensitivity of mean-variance-efficient portfolios to changes in asset means: Some analytical and computational results. *The Review of Financial Studies*, Vol. 4 (2), 315–342.

Biglova, Almira – Ortobelli, Sergio – Rachev, Svetlozar – Stoyanov, Stoyan (2004) Different approaches to risk estimation in portfolio theory. *The Journal of Portfolio Management*, Vol. 31 (1), 103–112.

Bloomfield, Ted – Leftwich, Richard – Long, John B., Jr. (1977) Portfolio strategies and performance. *Journal of Financial Economics*, Vol. 5 (2), 201–218.

Bollerslev, Tim (1986) Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, Vol. 31 (3), 307–327.

Bollerslev, Tim (1990) Modelling the coherence in short-run nominal exchange rates: A multivariate generalized ARCH model. *The Review of Economics and Statistics*, Vol. 72 (3), 498–505.

Broadie, Mark (1993) Computing efficient frontiers using estimated parameters. *Annals of Operations Research*, Vol. 45 (1–4), 21–58.

Byrne, John Aidan (2010) Traders thrive on life in the fast lane. *Institutional Investor*, Vol. 35 (9), 68–69.

Chopra, Vijay K. – Ziemba, William T. (1993) The effect of errors in means, variances, and covariances on optimal portfolio choice. *Journal of Portfolio Management*, Vol. 19 (2), 6–11.

Cornuejols, Gerard – Tütüncü, Reha (2007) *Optimization methods in finance*. Cambridge University Press. Cambridge, UK.

DeMiguel, Victor – Garlappi, Lorenzo – Nogales, Francisco J. – Uppal, Raman (2009a) A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms. *Management Science*, Vol. 55 (5), 798–812.

DeMiguel, Victor – Garlappi, Lorenzo – Uppal, Raman (2009b) Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy? *The Review of Financial Studies*, Vol. 22 (5), 1915–1953.

Ding, Letian – Fei, Peng (2008) Optimal portfolio choice with dynamic asymmetric correlations and transaction constraints. Working paper available at SSRN: http://ssrn.com/abstract=1346164.

Domowitz, Ian – Glen, Jack – Madhavan, Ananth (2001) Liquidity, volatility and equity trading costs across countries and over time. *International Finance*, Vol. 4 (2), 221–255.

Eling, Martin – Schuhmacher, Frank (2007) Does the choice of performance measure influence the evaluation of hedge funds? *Journal of Banking & Finance*, Vol. 31 (9), 2632–2647.

Elton, Edwin J. – Gruber, Martin J. (1997) Modern portfolio theory, 1950 to date. *Journal of Banking & Finance*, Vol. 21 (11–12), 1743–1759.

Engle, Robert F. (1982) Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, Vol. 50 (4), 987–1007.

Engle, Robert (2002) Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business & Economic Statistics*, Vol. 20 (3), 339–350.

Engle, Robert F. – Sheppard, Kevin (2001) Theoretical and empirical properties of dynamic conditional correlation multivariate GARCH. NYU Working paper No. FIN-01-027.

Evans, John L. – Archer, Stephen H. (1968) Diversification and the reduction of dispersion: An empirical analysis. *The Journal of Finance*, Vol. 23 (5), 761–767.

Fabozzi, Frank J. – Gupta, Francis – Markowitz, Harry M. (2002) The legacy of modern portfolio theory. *Journal of Investing*, Vol. 11 (3), 7–22.

Fabozzi, Frank J. – Focardi, Sergio M. – Kolm, Petter N. (2006a) *Financial modeling of the equity market: From CAPM to cointegration*. John Wiley & Sons, Inc. Hoboken, NJ.

Fabozzi, Frank J. – Focardi, Sergio – Jonas, Caroline (2006b) *Survey on trends in equity portfolio modeling 2006*. The Intertek Group.

Fama, Eugene F. (1965) The behavior of stock-market prices. *The Journal of Business*, Vol. 38 (1), 34–105.

Fama, Eugene F. (1970) Multiperiod consumption-investment decisions. *The American Economic Review*, Vol. 60 (1), 163–174.

Garlappi, Lorenzo – Uppal, Raman – Wang, Tan (2007) Portfolio selection with parameter and model uncertainty: A multi-prior approach. *The Review of Financial Studies*, Vol. 20 (1), 41–81.

Huberman, Gur – Jiang, Wei (2006) Offering versus choice in 401(k) plans: Equity exposure and number of funds. *The Journal of Finance*, Vol. 61 (2), 763–801.

Hull, John C. (2009) *Options, futures, and other derivatives*. 7th ed. Pearson Education, Inc. Upper Saddle River, NJ.

Jagannathan, Ravi – Ma, Tongshu (2003) Risk reduction in large portfolios: Why imposing the wrong constraints helps. *The Journal of Finance*, Vol. 58 (4), 1651–1683.

Jarque, Carlos M. – Bera, Anil K. (1987) A test for normality of observations and regression residuals. *International Statistical Review*, Vol. 55 (2), 163–172.

Jobson, J. D. – Korkie, Bob M. (1981) Performance hypothesis testing with the Sharpe and Treynor measures. *The Journal of Finance,* Vol. 36 (4), 889–908.

Jorion, Philippe (1985) International portfolio diversification with estimation risk. *Journal of Business*, Vol. 58 (3), 259–278.

Jorion, Philippe (1986) Bayes-Stein estimation for portfolio analysis. *Journal of Financial and Quantitative Analysis*, Vol. 21 (3), 279–292.

Jorion, Philippe (1991) Bayesian and CAPM estimators of the means: Implications for portfolio selection. Journal of Banking and Finance, Vol. 15 (3), 717–727.

Kan, Raymond – Zhou, Guofu (2007) Optimal portfolio choice with parameter uncertainty. *Journal of Financial and Quantitative Analysis*, Vol. 42 (3), 621–656.

Ledoit, Olivier – Wolf, Michael (2003) Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *Journal of Empirical Finance*, Vol. 10 (5), 603–621.

Ledoit, Olivier – Wolf, Michael (2004) Honey, I shrunk the sample covariance matrix. *The Journal of Portfolio Management*, Vol. 30 (4), 110–119.

Ledoit, Olivier – Wolf, Michael (2008) Robust performance hypothesis testing with the Sharpe ratio. *Journal of Empirical Finance*, Vol. 15 (5), 850–859.

Ljung, G. M. – Box, G. E. P. (1978) On a measure of lack of fit in time series models. *Biometrika*, Vol. 65 (2), 297–303.

Markowitz, Harry (1952) Portfolio selection. *The Journal of Finance*, Vol. 7 (1), 77–91.

Michaud, Richard O. (1989) The Markowitz optimization enigma: Is 'optimized' optimal? *Financial Analysts Journal*, Vol. 45 (1), 31–42.

MSCI Index Calculation Methodologies (2010) MSCI Data Operations & Technology.

Pástor, Lubos (2000) Portfolio selection and asset pricing models. *The Journal of Finance*, Vol. 55 (1), 179–223.

Peiró, Amado (1994) The distribution of stock returns: International evidence. *Applied Financial Economics*, Vol. 4 (6), 431–439.

Rasmussen, Mikkel (2003) *Quantitative portfolio optimisation, asset allocation and risk management*. Palgrave Macmillan. Hampshire.

Roy, A. D. (1952) Safety first and the holding of assets. *Econometrica*, Vol. 20 (3), 431–449.

Scherer, Bernd (2006) A note on the out-of-sample performance of resample efficiency. *Journal of Asset Management*, Vol. 7 (3–4), 170–178.

Scherer, Bernd (2007) Can robust portfolio optimisation help to build better portfolios? *Journal of Asset Management*, Vol. 7 (6), 374–387.

Scott, Robert C. – Horvath, Philip A. (1980) On the direction of preference for moments of higher order than the variance. *The Journal of Finance*, Vol. 35 (4), 915–919.

Sharpe, William F. (1966) Mutual fund performance. *The Journal of Business*, Vol. 39 (1), 119–138.

Sun, Walter – Fan, Ayres – Chen, Li-Wei – Schouwenaars, Tom – Albota, Marius A. (2006) Optimal rebalancing for institutional portfolios. *The Journal of Portfolio Management*, Vol. 32 (2), 33–43.

Tsay, Ruey S. (2010) *Analysis of financial time series*. 3rd ed. John Wiley & Sons, Inc. Hoboken, NJ.

APPENDIX 1 PROGRAMMING CODE FOR THE EMPIRICAL STUDY

The code in the R programming language for the investment process and parameter estimation is presented below. The main file, *invest.R*, calls the parameter estimation files *DCCestimation.R* and *shrinkestimation.R*, and the file that performs the portfolio optimization, *solveweights.R*. The code has been modified for easier readability.

invest.R

```
#---DEFINE VARIABLES---#
dataset<-"EMU1"         #Name of the data set {EMU1,EMU2,BEM1,BEM2}
X<-100                  #Defines starting wealth
tc<-20                  #Set transaction costs in basis points
beg<-251    #Defines the beginning of trading period (always 251)
strategy<-"sample"      #Set strategy {ew,sample,shrink,dcc}
window<-250             #Set estimation window length (in days)
update<-20              #Set updating frequency (in days)
rangerebal<-0           #To use threshold rebalancing, set to 1

#---IMPORT DATA---#
P<-read.csv(paste(dataset,".csv",sep=""), sep=";",
          dec=",",header=FALSE)
rows=dim(P)[1]
cols=dim(P)[2]

RE=rbind(rep(0,cols),data.frame(diff(log(data.matrix(P)))))
#Creates log-returns matrix

v=(rep(0,cols))#Denotes the amount of each asset held in the portfolio
turnover=0
Xseries<-numeric(0)
if(rangerebal==1) {update=1}

#---START ITERATION---#
for(i in beg:rows){
X=sum(v*P[i,])
if(X==0){X=100}
w=v*P[i,]/X

if (any(seq(beg,rows,update)==i)==TRUE) {
#Determines if it is time to rebalance
oldw=as.numeric(w)
w=rep(1/cols,cols)

#Call the covariance matrix estimation procedure
if(strategy=="sample") {C=cov(RE[(i-window+1):i,])}
if(strategy=="dcc") {source("DCCestimation.R")}
if(strategy=="shrink") {source("shrinkestimation.R")}
if(strategy!="ew") {source("solveweights.R")}
```

```
if (i==beg || rangerebal!=1 || ((sum(abs(w-
           oldw))*tc/10000)/(sqrt(sum(as.numeric(t(oldw))%*%Dmat%*%(as
           .numeric(oldw))))-
           sqrt(sum(as.numeric(t(w))%*%Dmat%*%(as.numeric(w))))))<1) {

turnover=turnover+sum(abs(w-oldw))
X=X-(sum(abs(w-oldw))*X*tc/10000)              #Trading costs
v=w*X/P[i,]
}
}
Xseries=rbind(Xseries,X)
}

#---END ITERATION---#
```

DCCestimation.R

```
data=RE[(i-window+1):i,]

#---LogLikelihood-function for GARCH(1,1)---#
Lv <-function(theta) {
h<-numeric(window)
h[1]=var(data2)
for (t in 2:window) {
h[t]=theta[1]+theta[2]*((data2[t-1])^2)+theta[3]*h[t-1]
}
sum(log(h)+(data2^2)/h)
}

#---Estimate GARCH(1,1) parameters for each individual asset---#
theta_s<-matrix(0,cols,3)
startvalues=c(0.0001,0.1,0.8)
for (k in 1:cols) {
data2=RE[(i-window+1):i,k]
theta_s[k,]=constrOptim(c(0.0001,0.1,0.8),Lv,grad=NULL,ui=rbind(diag(1
           ,3),diag(-1,3)),ci=c(rep(1e-07,3),rep(-1,3)))$par
}

#---Determine time-varying variances---#
h<-matrix(0,window,cols)
h[1,]=diag(var(data))
for (t in 2:window) {
h[t,]=theta_s[,1]+theta_s[,2]*((data[t-1,])^2)+theta_s[,3]*h[t-1,]
}
#Determine one-period forecast
hextra=theta_s[,1]+theta_s[,2]*((data[t,])^2)+theta_s[,3]*h[t,]

#---Determine time-varying residuals---#
D=sqrt(h)
epsilon<-matrix(0,window,cols)
for (f in 1:dim(data)[1]) {
for (g in 1:cols) {
epsilon[f,g]=data[f,g]/D[f,g]
}
}
```

```
#---Step 2:Estimate parameters for time-varying correlation matrix---#
Lc <-function(lamda) {
Q<-matrix(0,window,cols*cols)
R<-matrix(0,window,cols*cols)
term1<-numeric(window)
term2<-numeric(window)
term1[1]=1
term2[1]=0
Q[1,]=0.5*as.vector((epsilon[t-1,]%*%t(epsilon[t-
          1,])))+0.5*as.vector(cor(epsilon))
for (t in 2:window) {
Q[t,]=(1-lamda)*as.vector((epsilon[t-1,]%*%t(epsilon[t-
          1,])))+lamda*Q[t-1,]
R[t,]=as.vector(solve(diag(sqrt(diag(matrix(Q[t,],cols,cols)))))%*%mat
          rix(Q[t,],cols,cols)%*%solve(diag(sqrt(diag(matrix(Q[t,],co
          ls,cols))))))

term1[t]=det(matrix(R[t,],cols,cols))
term2[t]=t(epsilon[t,])%*%solve(matrix(R[t,],cols,cols),tol=1e-
          20)%*%epsilon[t,]
}
sum(log(term1)+term2)
}

lamda_s=nlminb(0.5,Lc,lower=0.01,upper=1)$par

#---Determine time-varying correlation matrix---#
Q<-matrix(0,window+1,cols*cols)
R<-matrix(0,window+1,cols*cols)
Q[1,]=0.5*as.vector((epsilon[t-1,]%*%t(epsilon[t-
          1,])))+0.5*as.vector(cor(epsilon))

for (t in 2:(window+1)) {
Q[t,]=(1-lamda_s)*as.vector((epsilon[t-1,]%*%t(epsilon[t-
          1,])))+lamda_s*Q[t-1,]
R[t,]=as.vector(solve(diag(sqrt(diag(matrix(Q[t,],cols,cols)))))%*%mat
          rix(Q[t,],cols,cols)%*%solve(diag(sqrt(diag(matrix(Q[t,],co
          ls,cols))))))
}

#---Define the covariance matrix forecast for the next trading day---#
C=sqrt(diag(as.numeric(hextra)))%*%matrix(R[t,],cols,cols)%*%sqrt(diag
          (as.numeric(hextra)))
```

shrinkestimation.R

```
data=RE[(i-window+1):i,]

p=(2/((cols-1)*cols))*(sum(cor(data)-diag(1,cols))/2)
#Calculate average correlation

ECC=matrix(p,cols,cols)+diag(1-p,cols)
ECC2=diag(sqrt(diag(cov(data))))%*%ECC%*%diag(sqrt(diag(cov(data))))
#Define CC-matrix
```

```
#---Calculate parameters for optimal shrinkage weight---#

piimatrix=matrix(0,cols,cols)
for (k in 1:cols) {
for (l in 1:cols) {
piimatrix[k,l]=(1/window)*sum(((data[,k]-mean(data[,k]))*(data[,l]-
          mean(data[,l]))-cov(data)[k,l])^2)
}
}

vetamatrix=matrix(0,cols,cols)
for (k in 1:cols) {
for (l in 1:cols) {
vetamatrix[k,l]=(p/2)*sum((sd(data[,l])/sd(data[,k]))*(1/window)*tcros
          sprod(sum((data[,k]-mean(data[,k]))^2-
          cov(data)[k,k]),((data[,k]-mean(data[,k]))*(data[,l]-
          mean(data[,l]))-
          cov(data)[k,l]))+(sd(data[,k])/sd(data[,l]))*(1/window)*tcr
          ossprod(sum((data[,l]-mean(data[,l]))^2-
          cov(data)[l,l]),((data[,k]-mean(data[,k]))*(data[,l]-
          mean(data[,l]))-cov(data)[k,l])))
}
}

difmatrix=matrix(0,cols,cols)
for (k in 1:cols) {
for (l in 1:cols) {
difmatrix[k,l]=(ECC2[k,l]-cov(data)[k,l])^2
}
}

#---Calculate optimal weight---#
z=max(0,min(((sum(piimatrix)-sum(diag(piimatrix))-
          sum(vetamatrix)+sum(diag(vetamatrix)))/sum(difmatrix))/wind
          ow,1))

C=z*ECC2+(1-z)*cov(data)  #Define the shrinkage covariance matrix
```

solveweights.R

```
library(quadprog)
# Define optimization parameters for quadratic optimization
Dmat<-C
dvec<-rep(0,cols)
Amat<-t(as.matrix(rbind(rep(1,cols),diag(1,cols),diag(-1,cols))))
bvec<-c(1,rep(-0.25,cols),rep(-0.25,cols))
meq=1

# Solve for the optimal weights
w=solve.QP(Dmat, dvec, Amat, bvec, meq)$solution
```