FACULTY
OF MATHEMATICS
AND PHYSICS
Charles University

MASTER THESIS

Jan Vágner

Multivariate volatility forecasts for large portfolios

Department of Probability and Mathematical Statistics

Supervisor of the master thesis: prof. RNDr. Tomáš Cipra DrSc.
Study programme: Financial and Insurance Mathematics

Prague 2023

I declare that I carried out this master thesis independently, and only with the cited sources, literature and other professional sources. It has not been used to obtain another or the same degree.

I understand that my work relates to the rights and obligations under the Act No. 121/2000 Sb., the Copyright Act, as amended, in particular the fact that the Charles University has the right to conclude a license agreement on the use of this work as a school work pursuant to Section 60 subsection 1 of the Copyright Act.

In .... date ....

Author's signature

I would like to thank Prof. Cipra for his willingness to invest his time and knowledge into the preparation of this thesis. He supplied me with a large number of ideas that significantly enhanced the final form of my thesis.

Finally, I would like to dedicate this thesis to all the ones who supported me throughout my entire life.

Title: Multivariate volatility forecasts for large portfolios

Author: Jan Vágner

Department: Department of Probability and Mathematical Statistics

Supervisor: prof. RNDr. Tomáš Cipra DrSc., (Department of Probability and Mathematical Statistics

Abstract: One deals with the estimation and consequent forecast of the integrated covariance matrix in the context of high-frequency stock price data and high dimensionality regarding the number of analyzed assets. We present several methods for the integrated covariance estimation and then use these estimates as a basis for forecasting models. We mainly focus on the multivariate extensions of the HAR model. Finally, in the empirical study, we compare different model-estimator combinations (based on 5-min interval observation and 50 assets) using economic and statistical evaluation. Economic evaluation is based on portfolio optimization, including transaction costs.

Keywords: Integrated covariance estimators, Multivariate HAR model, Multivariate volatility, Portfolio optimization with transaction costs, Realized Covariance

Contents

Introduction

Estimates and forecasts of the second moment represent an important topic in the context of finance and economics. The concept of volatility and covariance (covolatility) of assets plays a key role in portfolio selection, where it measures risk. A wide range of papers was dedicated to the development of different techniques for the estimation of covariance and volatility of assets. Consequently, forecasting models assuming certain time dependence of covariance are of interest. Commonly used models in the context of covariance estimation and forecasting include ARCH and GARCH models that have been extensively studied over the past years. However, there has been a significant shift in the relevant research in the recent years.

Over the last decades, we have observed impactful technological enhancements in the finance. Increasing computational power and better overall connectivity enable us to process and share data faster than before. Possibly the most groundbreaking change in current finance enabled by these technological advancements is the availability of high-frequency intraday data. Generally speaking, there is no clear definition of high-frequency data. However, any data with a time step smaller than one day could usually be considered as highfrequency data. Nevertheless, high-frequency data commonly include time intervals of 30-sec, 1-min, 2-min, 5-min, 15-min, and 30-min. In the recent years, we can also notice the emergence of ultra-high-frequency data. Such data correspond to so-called tick-by-tick observations, which are recorded even a hundred times per second.

In the past, most studies analyzing the covariance structure of portfolios of assets almost exclusively based main findings on the daily price observations (or even less frequently observed data). The ability to observe data more frequently, therefore, brings to the table new possibilities and topics in terms of the development of estimators and forecasting methods. On the contrary, such intraday data also introduce new challenges that have to be taken into the consideration. The prime example of such a challenge is microstructure noise. This term refers to various concepts that influence the price on small time intervals. Interestingly, these effects are more predominant for observations with higher frequency.

In contrast to the studies focusing on the methods developed on the basis of low-frequency data, e.g., Šípka (2022) for daily data, we concentrate on the analysis and comparison of different estimators and forecasting models based on high-frequency data. As the high-frequency data enables us to approximate the continuous price process better, it is reasonable to focus on the estimation of the so-called integrated covariance matrix of a continuous price process itself. Nowadays, this topic is highly discussed in modern finance as it promises more accurate descriptions of the actual covariance structure of the price process.

In this thesis, we focus at first on different estimators of the integrated covariance. Consequently, we use obtained estimates as a basis for forecasting models. We will survey several competing forecasting models which can be combined with different integrated covariance estimators. We will be mainly interested in the generalizations of a univariate HAR model into a multidimensional case since the univariate HAR model shows favorable results for modeling univariate volatilities. We will present several such generalizations, which we later compare in an empirical study. We have not found such a comparison of different combinations of covariance estimators and forecasting models in the relevant literature. Therefore, we view it as an interesting topic to discuss. The comparison itself will be based on both economic and statistical forecast evaluation. The economic comparison is tightly connected with the portfolio optimization problem, including transaction costs. Transaction cost is also a very rarely discussed topic in the relevant literature. The thesis's content is ordered so that we first present the critical theory in the first four chapters. Lastly, the fifth chapter presents an empirical study, where we apply described methods in the theoretical part to real-life data consisting of the stock prices of 50 different companies using 5-min interval data. The data and the corresponding code will be a part of an electronic attachment.

I. Integrated Covariance

In this chapter, we define the integrated covariance matrix, also known as the integrated covolatility matrix. Such matrices represent the analogue of the integrated volatility in the multivariate case. We adopt the notation of formal definition mainly from Bollerslev et al. (2019). In general, we will assume that there are N assets analyzed over time $t \in [0, T]$. However, in order to present a more comprehensive explanation of the corresponding model, in this chapter, we set $t \in [0, 1]$. We start with the definition of a log-price process which we define as $\boldsymbol{X}_t^* = (X_{t,(1)}^*, \ldots, X_{t,(N)}^*)^T$, where $X_{t,(n)}^*$ denotes the log-price of the n-th asset at time t. This process is assumed to be evolving in a continuous time and following Itô (semi-martingale) process:

$$\boldsymbol{dX}_t^* = \boldsymbol{\mu}_t dt + \boldsymbol{\sigma}_t \boldsymbol{dB}_t, \ 0 \leq t \leq 1. \tag{I.1}$$

Vector $\boldsymbol{\mu}_t = (\mu_{t,(1)}, \ldots, \mu_{t,(N)})^T$ denotes an N-dimensional locally bounded drift process, $\boldsymbol{\sigma}_t$ is an $N \times N$ symmetric matrix corresponding to a given covariance process and $\boldsymbol{B}_t = (B_{t,(1)}, \ldots, B_{t,(N)})^T$ is an N-dimensional vector of independent Brownian motions. We also denote the spot covariance matrix of $\boldsymbol{X}_t^*$ as $\boldsymbol{\Sigma}_t = \boldsymbol{\sigma}_t \boldsymbol{\sigma}_t^T$.

The characteristic we will mainly focus on for the rest of this thesis is the integrated covariance. It is formally defined as a $N \times N$ matrix:

$$IC = \int_0^1 \boldsymbol{\Sigma}_s ds. \tag{I.2}$$

Unfortunately, the process $\boldsymbol{X}_t^*$ and the integrated covariance IC are not observable in practice. Therefore, in order to deal with the integrated covariance, we define the difference $\boldsymbol{r}_{t_i}^* = (\boldsymbol{X}_{t_{i+1}}^* - \boldsymbol{X}_{t_i}^*)$ as a continuous log-return vector process (further simply continuous return process) with $\Delta = \max_i(t_{i+1} - t_i) \to 0$. IC can be then in theory estimated by the quadratic variation of the process $\boldsymbol{X}_t^*$. Formally, the quadratic variation of a stochastic process is defined using a continuous return process, $\boldsymbol{r}_t^*$, as:

$$QV = \lim_{\Delta \to 0} \left(\sum_i (\boldsymbol{r}_{t_i}^*)(\boldsymbol{r}_{t_i}^*)^T \right), \quad 0 \leq t_i \leq 1. \tag{I.3}$$

By the theory of the quadratic variation, it holds that $QV \to IC$. Such estimator is unbiased, assuming the high sampling frequency asymptotically free of a measurement error as described in Andersen et al. (2001).

On the other hand, it is important to note that the differences in sampling times are assumed to be infinitesimally small, which is hardly achievable in practice. Therefore, we define the log-price vector process as follows:

$$\boldsymbol{x}_t = \boldsymbol{X}_t^* + \boldsymbol{u}_t,$$

where $\boldsymbol{u}_t$ represents various microstructure noises occurring on the real market (e. g. discreteness, non-trading, bid-ask spread, etc.). Such a process $\boldsymbol{x}_t$ is what we are capable of observing on the market. Therefore, the estimator based on the cross-products of intraday returns obtained using $\boldsymbol{x}_t$, as introduced in equation I.3 for continuous counterpart, becomes very noisy due to the presence of the microstructure noises on the market. Hence, the resulting estimates are contaminated with potentially significant measurement errors, which is highly undesirable. We will devote the following chapter to various estimators of the integrated covariance using discretized data. Some of them specifically focus on the task of dealing with microstructure noises.

Remark. In its more general form, the continuous log-price vector process $\boldsymbol{X}_t^*$ could also include components corresponding to jumps and co-jumps. We will discuss these topics in the following chapter, but it is not our main concern to describe these particular concepts in great detail.

Remark. The best approximation of the continuous log-price process $\boldsymbol{X}_t^*$ is the so-called tick-by-tick (or, in other words, ultra-high-frequency) data. However, this type of data carries several disadvantages. The first one is their quality which is usually quite poor because such datasets frequently include a lot of missing observations. Observations are also usually not equidistantly spread, which can be, in some cases, difficult to handle. Another important property is that microstructure noises are much more persistent when using ultra-high-frequency data. In fact, adjusting the estimators to remain consistent when using ultra-high-frequency data is usually necessary. From a more practical point of view, their noticeable shortcoming is their high acquisition cost. For those reasons, it is more feasible to work with other high-frequency intraday data like 5-min or 2-min data, which does bring fewer challenges, and their performance in the context of various integrated covariance estimators is similar to ultra-high-frequency data.

II. Estimators of Integrated Covariance

In this chapter, we present various realized covariance estimators, or more precisely, estimators of the integrated covariance. In the next chapter, these estimators will be used to obtain forecasts of the realized covariance matrix. We consider only several estimators in this thesis, even though there are many distinct approaches to estimating the integrated covariance. It would simply not be possible to include all estimators, so we rather focus on selected ones that are based on different estimation methods.

On the other hand, we would like to mention some interesting estimators that we will omit later. One such estimator is the composite realized kernel estimator introduced in Lunde et al. (2016). This estimator uses two-dimensional realized kernel estimators (described in the subsection II.3), which are later recombined to construct high-dimensional estimates of realized covariance. Even though this estimator is very data efficient and can be used for a large number of assets, it is not guaranteed that the resulting estimate will be positive semi-definite. Therefore, the estimated matrix has to be projected to the space of positive definite matrices to obtain a proper estimator.

The second appealing approach described in Bollerslev et al. (2019) combines multivariate modulated realized covariance estimators with pre-averaging estimators. These estimators will be later discussed in chapter II.4. The estimator of Bollerslev et al. (2019) accounts for serially correlated microstructure noises and price process jumps, a concept that will be described more broadly in II.5. Another possible advantage of the resulting estimator is its flexibility because it can be extended to an approximative factor model as it is detailed in the Bollerslev et al. (2019), where the principal component analysis (PCA) is used to reduce the dimensionality.

Similarly to the previous chapter, we will assume that there are N investment assets in consideration. From now on, we will work exclusively with discrete observations, in contrast to the first chapter dealing with the continuous vector price process. We specify here the denotation in detail to avoid any future confusion. We will assume that observations are collected over T time periods, and during each time period t, there are several instants i during which we observe the log-prices of assets. First, we introduce the notation for the log-price vector and then for log-returns of assets:

$$\boldsymbol{x}_{t,i} = (x_{t,i,(1)}, \dots, x_{t,i,(N)})^T,$$

$$\boldsymbol{r}_{t,i} = \boldsymbol{x}_{t,i} - \boldsymbol{x}_{t,i-1} = (r_{t,i,(1)}, \dots, r_{t,i,(N)})^T,$$

where $t \in \{1, 2, \dots, T\}$ serves as a time period index (day index) for the interval $[t-1, t)$, index i represents an equidistant partition of particular time period t, and indices $1, 2, \dots, N$ denote individual assets. We will generally assume that each time period t contains $M + 1$ equidistant values of log-prices, $i \in \{0, 1, \dots, M\}$.

This translates to M partitions, $i \in \{1, 2, \ldots, M\}$, for time series of returns in each time period t. Effectively, we have M observations in each time period t.

Estimators described in this chapter will be later used in an empirical study, where we will compare their forecasting capabilities. We use intraday observable vectors of log-returns (briefly returns) $\boldsymbol{r}_{t,i}$ to define individual estimators in the rest of this section.

Remark. Certain approaches presented in this chapter do not require partitions to be equidistantly spaced. In general, partitions can vary for distinct assets. In such cases, we will introduce methods that could be used to handle these specific circumstances. Implicitly, we will not use the assumption of equidistant partitions in such situations.

Remark. (Non-synchronicity) The concept of varying partitions of a given time interval for different assets is often called non-synchronicity. On the other hand, if we observe stock prices of assets in the same time instances (same partitions), we call this data synchronous. In other words, non-synchronicity concerns the possibility that assets can have different trading time windows, whereas synchronicity implies the same trading time windows for all analyzed assets.

II.1 Realized Covariance

Firstly, we introduce the simplest integrated covariance estimator, which was foreshadowed in chapter I. It was initially outlined in its univariate form by Andersen et al. (2001). We will refer to this estimator as the realized covariance[1] estimator defined as:

$$RC_t = \sum_{i=1}^{M} \boldsymbol{r}_{t,i} \boldsymbol{r}_{t,i}^T, \tag{II.1}$$

where M is once again the number of partitions and $\boldsymbol{r}_{t,i}$ is an N-dimensional vector of asset returns in the day t and the i-th partition.

The main advantage of this estimator is its simplicity. Therefore, it seems convenient to use this particular estimator in practice; however, if the number of intraday observations M increases, then RC_t becomes inconsistent due to microstructure noises. Realized covariance neither accounts for jumps of the process. This motivates us to construct different estimators that would overcome the bias in the case of finer partitioning.

II.2 Realized Bi-power Covariance

Next, we present the multivariate realized bi-power estimator, originally proposed in Barndorff-Nielsen and Shephard (2003) assuming a univariate price process

[1]This estimator is also known as realized covolatility or realized covariation.

and later generalized for a multivariate price process in Shephard and Barndorff-Nielsen (2004).

The idea behind this estimator is to capture the jump component of the price process and factor its contribution to the variation of the process. It describes both the continuous part and jump component of quadratic variation. Such an approach is a significant modification in comparison with the previously described realized covariance estimator, where only the continuous part of the quadratic variation was targeted. The bi-power covariance can also be used together with the realized covariance to construct a non-parametric statistic and corresponding statistical test for the occurrence of jumps.

The multivariate realized bi-power estimator $BP_t = (bp_{t,(m,n)})$, for $m, n = 1, 2, \ldots, N$, at time period t is an $N \times N$ symmetric matrix with the (m, n)-th element of the matrix defined as:

$$\begin{aligned} bp_{t,(m,n)} =& \frac{\pi}{8} \sum_{i=2}^{M} \Big(\left| r_{t,i,(m)} + r_{t,i,(n)} \right| \left| r_{t,i-1,(m)} + r_{t,i-1,(n)} \right| - \\ & \left| r_{t,i,(m)} - r_{t,i,(n)} \right| \left| r_{t,i-1,(m)} - r_{t,i-1,(n)} \right| \Big), \end{aligned} \tag{II.2}$$

where $r_{t,i,(n)}$ is the n-th element of vector $\boldsymbol{r}_{t,i}$. One drawback of this estimator is that it does not deal with microstructure noises. However, it was shown in Shephard and Barndorff-Nielsen (2004) that the impact of these noises on the consistency of this estimator is rather negligible.

II.3 Realized Kernel Covariance

Another important class of estimators is the so-called realized kernel covariance. This estimator's multivariate setting was first introduced in Barndorff-Nielsen et al. (2011), where important properties of this approach, including its consistency and guaranteed positive semi-definiteness, were derived. At the same time, it is relatively robust to microstructure noises and non-synchronicity.

Realized kernel estimator is formally defined as:

$$\begin{aligned} RK_t = \Gamma_0 + 2 \sum_{h=1}^{H} f\left(\frac{h-1}{H}\right) \Gamma_h, \\ \text{where } \Gamma_h = \sum_{i=h}^{M} \boldsymbol{r}_{t,i} \boldsymbol{r}_{t,i-h}^T \text{ for} H \geq 0. \end{aligned} \tag{II.3}$$

In this case, H denotes an arbitrarily chosen bandwidth parameter crucial for controlling the microstructure noises and non-synchronicity. In the equation II.3, the function $f : \mathbb{R} \rightarrow \mathbb{R}$ is a non-stochastic weight function conventionally referred to as a kernel function.

Regarding parameter selection, the parameter H has to be increased with expanding portfolio size to maintain consistency of realized kernel covariance.

The recommended choice is to set H approximately to $n^{1/2}$. As for the kernel function f, there are various possible options like Parzen's, Tukey-Hanning's, or modified Tukey-Hanning's kernel, etc. We present and later implement only Parzen's kernel, which can be defined as:

$$f(x) = \begin{cases} 1 - 6x^2 + 6x^3 & k \leq \frac{1}{2} \\ 2(1-x)^3 & \frac{1}{2} \leq k \leq 1 \\ 0 & x > 1. \end{cases} \tag{II.4}$$

II.4 Modulated Realized Covariance

Next, we introduce the modulated realized covariance estimator that was first described in Christensen et al. (2010). This estimator should behave more robustly in the presence of microstructure noises and non-synchronous trading. However, it needs bias correction to remain consistent, especially in larger dimensions (in terms of the number of assets). This multivariate estimator is based on its univariate counterpart introduced in Jacod et al. (2009) as its generalization. Such a univariate estimator is frequently referred to as a pre-averaging estimator in related literature.

The adjective "pre-averaging" relates to the application of the pre-averaging procedure to returns. The motivation for the application of the pre-averaging procedure is to remove the effects of microstructure noises. The general idea behind this method is taken over from the concept of smoothing, where the less dominant effects should be removed from the principal pattern contained in the data. There is a practical approach to pre-averaging presented in Christensen et al. (2010), which we will closely follow.

First, we define a sequence of integers k_M and a number $\theta \in (0, \infty)$ such that:

$$\frac{k_M}{\sqrt{M}} = \theta + o(M^{-1/4}).$$

Then we choose a weight function $g : [0,1] \to \mathbb{R}$, which is continuous, piecewise continuously differentiable with a piecewise Lipschitz derivative g' and $g(0) = g(1) = 0$ fulfilling $\int_0^1 g^2(s)ds > 0$. We also construct following functions related to g:

$$\phi_1(s) = \int_s^1 g'(u)g'(u-s)du, \quad \phi_2(s) = \int_s^1 g(u)g(u-s)du,$$

$$\psi_1 = \phi_1(0), \quad \psi_2 = \phi_2(0),$$

$$\Phi_{1,1} = \int_0^1 \phi_1^2 ds, \quad \Phi_{1,2} = \int_0^1 \phi_1(s)\phi_2(s)ds,$$

$$\Phi_{2,2} = \int_0^1 \phi_2^2(s)ds.$$

Functions $\phi_1,\ \phi_2$ are assumed to be equal to zero outside the $[0,1]$ interval. With all these prerequisites in hand, we can now define the pre-averaged return for given t as:

$$\tilde{\boldsymbol{r}}_{t,i} = \sum_{j=1}^{k_M - 1} \left(g\left(\frac{j}{k_M}\right) \boldsymbol{r}_{t,i+j} \right) \text{ for } i = 0, 1, \ldots, M - k_M + 1. \tag{II.5}$$

In this general setting, we can define the modulated realized covariance as:

$$MRC_t = \frac{M}{M - k_M + 2} \cdot \frac{1}{\psi_2 k_M} \sum_{i=0}^{M-k_M+1} \tilde{\boldsymbol{r}}_{t,i} \tilde{\boldsymbol{r}}_{t,i}^T, \tag{II.6}$$

where the first term $\frac{M}{M-k_M+2}$ is a correction for the true number of summands.

One commonly used choice of the weight function is $g(x) = min(x, 1-x)$. In the related literature Hautsch and Podolskij (2013), it is also recommended to use one particular choice of parameter θ, $\theta = 0.8$. We mentioned at the beginning of this subsection that it is necessary to perform a bias correction for the estimator to remain consistent. Such a bias correction can be constructed as follows:

$$MRC_t = \frac{M}{M - k_M + 2} \cdot \frac{1}{\psi_2 k_M} \sum_{i=0}^{M-k_M+1} \tilde{\boldsymbol{r}}_{t,i} \tilde{\boldsymbol{r}}_{t,i}^T - \frac{\psi_1^{k_M}}{\theta^2 \psi_2^{k_M}} \hat{\Psi}, \tag{II.7}$$

where

$$\psi_1^{k_M} = k_M \sum_{i=1}^{k_M} \left(g\left(\frac{i+1}{k_M}\right) - g\left(\frac{i}{k_M}\right)\right),$$

$$\psi_n^{k_M} = \frac{1}{k_M} \sum_{i=1}^{k_M-1} g^2\left(\frac{i}{k_M}\right), \hat{\Psi} = \frac{1}{2M} \sum_{i=1}^{M} \boldsymbol{r}_{t,i} \boldsymbol{r}_{t,i}^T.$$

However, there is still one obstacle. Modulated realized covariance with bias correction is no longer a positive semi-definite matrix. Thus, it is necessary to perform certain procedures to ensure the estimator's positive semi-definiteness. One practical approach is presented in Shephard and Barndorff-Nielsen (2004). This method requires first performing the spectral decomposition of our estimated matrix. Then we concentrate on the diagonal matrix, which has eigenvalues on its diagonal. We then find and change all the negative eigenvalues to 0. Finally, we reconstruct the covariance matrix estimate using the newly obtained diagonal matrix of eigenvalues (with zeros instead of negative eigenvalues) in the spectral decomposition. In fact, we will later refer to this estimator with bias correction as the modulated realized covariance.

There is also an interesting connection between the modulated realized covariance estimator and the realized kernel estimator. More precisely, the modulated realized estimator can be mapped into the realized kernel estimator. For example, suppose we choose the weight function $g = min(x, 1-x)$ as mentioned earlier. In that case, it can be proven that the corresponding kernel estimator is Parzen's kernel, described in the subsection II.3.

II.5 Realized Threshold Covariance

In this subsection, we introduce the realized threshold covariance estimator, first presented in Mancini and Gobbi (2012). The main difference between this specific estimator and the ones presented above is that it considers potential jumps and co-jumps of the log-price process. Effectively, the jump component represents an idiosyncratic risk, while the co-jump component stands for systematic risk. The idiosyncratic risk can be thought of as a risk that is specific to a single asset. On the other hand, systematic risk is a risk that affects the whole market (or, possibly, a given market sector).

The realized threshold covariance estimator deals with non-synchronicity. By applying a particular threshold, it focuses on removing the effects of jumps. We will only present the final form of the estimator here because its derivation is very technical. One can find the detailed derivation of this specific estimator in Mancini and Gobbi (2012). We define realized threshold covariance $RTC_t = (rtc_{t,(m,n)}),\ m, n = 1, 2, \ldots, N$, as:

$$rtc_{t,(m,n)} = \sum_{i=1}^{M} r_{t,i,(m)} \mathbb{1}_{\{r^2_{t,i,(m)} \leq TR_{M,(m)}\}} r_{t,i,(n)} \mathbb{1}_{\{r^2_{t,i,(n)} \leq TR_{M,(n)}\}}, \tag{II.8}$$

where $TR_{M,(n)}$ is a given threshold value for individual asset n and $\mathbb{1}$ is an indicator function. The threshold parameter does not have to depend on an individual asset, but we present here a parameter choice that does assume its dependence on a specific asset. Paper by Mancini and Gobbi (2012) does not introduce any optimal threshold value and instead presents different possibilities that are compared using simulations. Nevertheless, there was subsequent research done concerning this topic, for example, Jacod and Todorov (2009), where the threshold value is defined as $9\Delta_M^{-1} \cdot bp_{t,(n,n)}$, where Δ_M is a window size (time partitioning lag) and $bp_{t,(n,n)}$ is a bi-power variation of asset n. It is not shown whether such a threshold behaves optimally; however, it represents a more sophisticated way of setting the threshold value.

II.6 Two-time Scale Realized Covariance

In the following subsection, we introduce the two-time scale realized covariance estimator. First, we present a general definition of this estimator, including bias correction, initially defined in Zhang et al. (2005) and Zhang (2011). Later, we describe its robust version that should be more immune against jumps. Both of these methods, in theory, require tick-by-tick data to perform as expected by theory. Nevertheless, we will be later interested in their performance on lower frequency data in comparison with the remaining methods. Both of these estimators are very complex, and they focus on dealing with microstructure noises and non-synchronicity.

First, it is necessary to calculate moving averages of returns based on two different rolling basis (time scales) of log-prices. Therefore, we will be using log-price series to derive this estimator. We follow the notation used at the beginning of this chapter, but we also need to introduce new time series for individual assets. Hence, the log-price time series for one unique asset over a time period t will be denoted as $\boldsymbol{x}_{t,(n)} = (x_{t,0,(n)}, \ldots, x_{t,M,(n)})^T$ for $n = 1, 2, \ldots, N$. Furthermore, we will refer to the short rolling basis J as the fast time scale and wider rolling basis K as the slow time scale. Typically, the fast time scale parameter J equals 1 (1 day). We then define the diagonal elements and non-diagonal elements of the two-time scale realized covariance $TS_t = (ts_{t,(m,n)})$, $m, n = 1, 2, \ldots, N$ separately. Let us start with the diagonal elements that can be for an asset n obtained as:

$$ts_{t,(n,n)} = \left(1 - \frac{\overline{M}_K}{\overline{M}_J}\right)^{-1} \left([\boldsymbol{x}_{t,(n)}, \boldsymbol{x}_{t,(n)}]^{(K)} - \frac{\overline{M}_K}{\overline{M}_J}[\boldsymbol{x}_{t,(n)}, \boldsymbol{x}_{t,(n)}]^{(J)}\right), \tag{II.9}$$

where

$$\overline{M}_K = \frac{M - K + 2}{K}, \quad \overline{M}_J = \frac{M - J + 2}{J},$$

$$[\boldsymbol{x}_{t,(n)}, \boldsymbol{x}_{t,(n)}]^{(l)} = \frac{1}{l} \sum_{i=1}^{M-l+1} \left(\boldsymbol{x}_{t,i+l,(n)} - \boldsymbol{x}_{t,i,(n)}\right)^2 \text{ for } l = J, K.$$

The definition of non-diagonal elements of this estimator requires a few additional steps. Synchronizing data at the beginning is necessary, meaning we first find the time partitions at which all the assets were traded. This is done using the refresh time method initially derived in Harris, F. H. deB. et al. (1995). We present here the refresh time method alongside the definition of non-diagonal elements of the two-time scale realized covariance of two assets m and n.

We need to assume a more general setting than the one described at the beginning of the chapter II to show the most crucial aspects of the refresh time method. We assume that the log-price $\boldsymbol{x}_{t,(m)}$ of the asset m was observed during instants (partitions of t) $0 = \tau_0 \leq \tau_1 \leq \cdots \leq \tau_{p_1} = 1$ and the log-price $\boldsymbol{x}_{t,(n)}$, of the asset n was observed during instants (partitions) $0 = \theta_0 \leq \theta_1 \leq \cdots \leq \theta_{p_2} = 1$. In general, these partitions do not have to be equidistant, and the assets could be observed in different instants of t. We also denote parameter $P = p_1 + p_2$, which will be important shortly. Now, we synchronize the observations using the refresh time method by applying relations:

$$t_i = max\{\tau \in (\tau_1, \ldots, \tau_{p_1}) : \tau \leq v_i\},$$

$$s_i = max\{\theta \in (\theta_1, \ldots, \theta_{p_2}) : \theta \leq v_i\},$$

where v_i can be for example defined using the relation $v_i - v_{i-1} = \Delta v$ for $\forall i$, where Δv represents a constant. This implies that a series of v_i's would be equally spaced out in time. We also assume that the length of this time series is equal to P. We can finally define the non-diagonal elements of TS_t as:

$$ts_{t,(m,n)} = c\left([\boldsymbol{x}_{t,(m)}, \boldsymbol{x}_{t,(n)}]^{(K)} - \frac{\overline{M}_K}{\overline{M}_J}[\boldsymbol{x}_{t,(m)}, \boldsymbol{x}_{t,(n)}]^{(J)}\right), \tag{II.10}$$

where

$$[\boldsymbol{x}_{t,(m)}, \boldsymbol{x}_{t,(n)}]^{(l)} = \frac{1}{l} \sum_{i=1}^{P-l+1} \left(\boldsymbol{x}_{t_{i+l},(m)} - \boldsymbol{x}_{t_i,(m)}\right)^T \left(\boldsymbol{x}_{s_{i+l},(n)} - \boldsymbol{x}_{s_i,(n)}\right)^T \text{ for } l = J, K,$$

c is a constant that must satisfy the relation $c = 1 + o(M_P^{-1/6})$, where M_P is a sampling frequency of the time series $\{v_i\}$. This constant should compensate for bias from the threshold construction.

There also exists a robust modification of the two-time scale realized estimator. It uses a similar approach concerning the refresh time method and the way how it deals with microstructure noises. However, it also targets jumps by adding additional terms and then by rescaling the estimated covariances. We will not present its definition here because it requires a significant generalization of our log-price process. One can study the full theory behind this specific robust estimator in Boudt and Zhang (2010). Unfortunately, both of these estimators are not necessarily positive semi-definite. Therefore, we need to apply a particular procedure to obtain a positive semi-definite estimator. One possibility is to perform the method previously described in subsection II.4.

II.7 Realized Cholesky Covariance

Next, we present estimator introduced in Boudt et al. (2017), which uses different approach in order to obtain an estimate of the integrated covariance. This model also adjusts for non-synchronous data and deals with microstructure noises. An important advantage of this estimator is also its guaranteed positive semi-definiteness by definition.

The realized Cholesky covariance distinguishes significantly from previously introduced estimators. We will briefly describe the method itself, but we will not go into great detail. An in-depth description of the algorithm itself can be found in Boudt et al. (2017).

We assume that the covariance matrix can be decomposed using Cholesky (LDL) decomposition[2] as follows:

$$\Sigma = HGH^T,$$

where

$$H = \begin{pmatrix} 1 & 0 & \dots & 0 \\ h_{2,1} & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ h_{N,1} & h_{N,2} & \dots & 1 \end{pmatrix} \quad \text{and} \quad G = \begin{pmatrix} g_{1,1} & 0 & \dots & 0 \\ 0 & g_{2,2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & g_{N,N} \end{pmatrix}.$$

Next, it is necessary to construct a factor model using synchronized returns. The refresh time method could be used for this purpose. Generally, it is assumed that factors are normally distributed with zero expected value and that their variance is proportional to the time between observations. Lastly, the estimates of specific elements of matrices G and H could be obtained as residual variances of the factors and beta coefficients from the regression of returns on these factors. Thus, it

[2] We neglect the time index for a more comprehensive explanation. In general, it holds $\Sigma_t = H_t G_t H_t^T$, hence H_t and G_t would be dependent on time.

is necessary to estimate quadratic variations of factors and quadratic covariations of factors and returns to obtain estimates of $g_{n,n}$ and $h_{m,n}$. Different estimators of integrated covariance could be used in this context.[3] It is recommended to apply the algorithm iteratively on an expanding dataset of returns.

Finally, we obtain the estimates of H and G that we use to reconstruct the realized Cholesky estimator as:

$$CholC = \hat{H}\hat{G}\hat{H}^T \tag{II.11}$$

As could be seen above, this approach is rather complex and could be very time inefficient. However, based on the simulation results from Boudt et al. (2017), it seems that the precision gain, especially in the environment of non-synchronicity, is not negligible, even when compared with composite estimators.

II.8 POET Estimator

Lastly, we present one representative of a slightly different class of estimators. The majority of estimators presented till now do not consider any specific adjustments for large dimensionality in terms of a number of assets.[4] Effectively, they usually perform quite poorly in cases when the number of assets N is fairly large. However, when N starts to be very large, for example, close to or even greater than the number of observations in a single time period, then serious problems with the overall performance of discussed estimators could occur. Therefore, it is reasonable to introduce some regularization to improve the accuracy of the final estimate. One such technique is presented by Fan et al. (2013), and it is usually referred to as the principal orthogonal complement thresholding (POET).

The POET estimator puts the covariance matrix into the context of an approximative factor model. This setup also enables the covariance matrix to be potentially sparse.[5] We assume a factor model for our returns given by:

$$\boldsymbol{r}_{t,i} = \boldsymbol{B}_t \boldsymbol{f}_{t,i} + \boldsymbol{u}_{t,i}.$$

Then we define a factor model for the covariance matrix using the relation:

$$\Sigma_t = \boldsymbol{B}_t cov(\boldsymbol{f}_{t,i})\boldsymbol{B}_t^T + \Sigma_u,$$

where $\boldsymbol{B}_t = (\boldsymbol{b}_{t,(1)}, \ldots, \boldsymbol{b}_{t,(N)})^T$ and $\boldsymbol{b}_{t,(n)}$ is an S-dimensional vector of factor loadings. Term $\boldsymbol{f}_{t,i}$ is an S-dimensional vector of common factors and Σ_u is an

[3] In Boudt et al. (2017), it is recommended to use the modulated realized covariance because of its robustness to microstructure noises.

[4] There are few exceptions that have been already outlined in this thesis. The realized Cholesky estimator does reduce dimensionality in some sense, and the briefly introduced composite kernel could also be transformed into an approximative factor model similar to the POET estimator.

[5] Sparse in the context of a matrix denotes a situation when the matrix consists of a large number of zero or close to zero (in a numerical sense) elements.

$N \times N$ covariance matrix of N-dimensional vector of idiosyncratic components $\boldsymbol{u}_{t,i}$, $i = 1, 2, \ldots, M$. Factor loadings $\boldsymbol{B}_t$, common factors $\boldsymbol{f}_{t,i}$ and idiosyncratic terms $\boldsymbol{u}_{t,i}$ (Σ_u) are all not observable. The only observable values in this factor model are returns themselves. Thus, we will start by estimating the covariance matrix Σ_t. In the case of the POET estimator, it is obtained using the realized covariance estimator. The factor model is then estimated using principal component analysis (PCA), which significantly reduces the dimensionality.

We present here the crucial steps of the POET estimator. Certain steps are omitted, but one can follow the in-depth definition of this method in Fan et al. (2013). We assume that $\hat{\lambda}_{t,1} \geq \hat{\lambda}_{t,2} \geq \cdots \geq \hat{\lambda}_{t,N}$ are ordered eigenvalues of the realized covariance matrix and vectors $\hat{\boldsymbol{z}}_{t,n}, n = 1, 2, \ldots, N$, are their corresponding eigenvectors. It holds due to the spectral decomposition of the realized covariance that:

$$\widehat{\Sigma}_t = \sum_{n=1}^{K} \hat{\lambda}_{t,n} \hat{z}_{t,n} \hat{z}_{t,n}^T + \widehat{\boldsymbol{R}}_{t,K},$$

where $\widehat{\boldsymbol{R}}_{t,K} = \sum_{n=K+1}^{N} \hat{\lambda}_{t,n} \hat{z}_{t,n} \hat{z}_{t,n}^T$ is the principal orthogonal complement and K is the number of diverging eigenvalues of Σ_t. K is generally unknown and has to be estimated. Bai and Ng (2001) describes a commonly used procedure for this purpose.

After determining the value of K, we apply thresholding to $\widehat{\boldsymbol{R}}_{t,K}$. Various methods can be used to set threshold parameters and thresholds themselves. One such method was presented in subsection II.5. We will not present any additional method in this thesis, and we will simply assume that after applying thresholds to $\widehat{\boldsymbol{R}}_{t,K}$ we obtain $\widehat{\boldsymbol{R}}_{t,K}^{\tau}$. With this fact in mind, we finally receive our estimator of the integrated covariance:

$$POET_t = \sum_{n=1}^{K} \hat{\lambda}_{t,n} \hat{z}_{t,n} \hat{z}_{t,n}^T + \widehat{R}_{t,K}^{\tau}. \tag{II.12}$$

The evident advantage of this estimator is that it does not require any optimization and, thus, is very feasible computationally. Another already mentioned advantage of this estimator is its accuracy, even with a large number of assets in a portfolio. Lastly, this covariance estimator guarantees positive semi-definiteness, which is a desirable property of the resulting integrated covariance matrix.

III. Forecasting Methodology

In this chapter, we introduce several approaches to forecasting integrated covariance. We will be interested exclusively in one-day (one-period) ahead forecasting throughout the rest of this thesis. This chapter will describe all models using the realized covariance estimator. Later, in the chapter V, we will apply selected models to different integrated covariance estimators.

First, we can mention the forecasting method, which is the simplest one possible. We will assume that the realized covariance matrix follows a random walk process as it evolves over time. Therefore, based on this assumption the resulting one-day ahead forecast is simply the estimate for the last observable time period. At first glance, this method seems too simple to yield reasonable results. Nevertheless, given the high-frequency intraday observations, it ultimately generates quite competitive forecasts. Hence, it performs relatively well despite the fact that this approach completely neglects the long-term trends of the realized covariance. We will formally denote it as the RW model.

III.1 HAR Model

We presented the simplest approach possible to forecasting. However, this approach seems, in theory, very naive. We will now introduce another competing method that incorporates lagged values into the model and, thus, enables our model to capture additional information about the behavior of the covariance matrix in a longer time horizon. First, we present the univariate version of the model and, later, its generalization to a multidimensional case.

III.1.1 Univariate HAR Model

The heterogeneous autoregressive model (HAR) of realized volatility was introduced by Corsi (2009). As we are working with a one-dimensional model, we will use the univariate counterpart of the realized covariance, which is usually referred to as the realized volatility, RV_t. The property of asset's (logarithmic) price volatility that the model accounts for is often referred to as heterogeneity. Several papers were published concerning different aspects of this hypothesis, and extensive studies empirically demonstrated that the heterogeneous property is worth dealing with. There are distinct entities contributing to the heterogeneity of volatility, e.g., temporal horizons, different risk profiles of agents, distinct ways of information processing, etc. To deal with all the mentioned sources of heterogeneity (and others) at once is nearly impossible. Corsi (2009) with his HAR model concentrates on the heterogeneity arising from the time horizon differences. He explains that the heterogeneity from time horizon differences originates from various market participants. These participants could be roughly divided into three distinct groups. The first group consists of short-term traders who frequently change their positions, usually in only a few minutes or hours (sometimes in just split seconds). In the second group, there are medium-term agents who change their positions at approximately weekly intervals. At last, long-term investors do not change their positions very often, and they sometimes do not change their holdings for years. These distinct groups eventually act differently in the given market environment. Hence, they create discrepancies in the volatility of asset prices. The HAR model aims to capture heterogeneity originating from such differences. In other words, the model should examine not only the current behavior of the volatility but also its long-term trends.

From a mathematical perspective, the univariate HAR model is basically an AR-type model with volatilities considered over variety of time horizons. Thus, it does not technically belong to the class of long-memory models. Nevertheless, it is shown by Corsi (2009) that thanks to the aggregation of volatilities, it can capture the long-memory property. In other words, it can approximate the heterogeneous property of volatility despite remaining a fairly simple model. The HAR model of realized volatility of a single asset n is defined as:

$$RV_{t+1} = \alpha + \beta^{(d)} RV_t + \beta^{(w)} RV_t^{(w)} + \beta^{(m)} RV_t^{(m)} + \varepsilon_{t+1}, \tag{III.1}$$

where RV_{t+1} and RV_t are realized volatilities at time $t+1$ and t. Thus, in the setup of our model, the term RV_t is a 1-day lagged realized volatility. The remaining realized volatilities in the model correspond to aggregated lagged values of the realized volatility over a week (5 days), $RV_t^{(w)}$, and over a month (20 or 22 days), $RV_t^{(m)}$. For example, the weekly aggregated lagged realized volatility is defined as:

$$RV_t^{(w)} = \frac{1}{5}\left(RV_t + RV_{t-1} + RV_{t-2} + RV_{t-3} + RV_{t-4}\right).$$

Coefficient α is an intercept, and $\beta^{(d)}$, $\beta^{(w)}$, $\beta^{(m)}$ are coefficients corresponding to lagged and aggregated values of realized volatility. Finally, ε_{t+1} is the non-observable volatility fluctuation. This model can also be modified to contain additional lagged independent variables, e.g., lagged realized volatilities aggregated over two weeks, etc. The structure of the model is very advantageous concerning forecasting as it essentially does not require any major operations to obtain forecasts. We can simply obtain the forecast for time period $t+1$ based on the values known at time t just from the model definition. We will denote this forecast as $\widehat{RV}_{t+1|t}$.

There are also several modifications of the univariate HAR model. One that is certainly worth noting is an approach introduced by Hillebrand and Medeiros (2010). They focused on combining the univariate HAR model with neural networks. To be more precise, they used bootstrap aggregation (bagging) in connection with the HAR model to target the noisiness in the data. This approach, based on their findings, improved the accuracy of (out-of-sample) forecasts.

III.1.2 Multivariate HAR Model

We established the model for the univariate case. However, our ultimate goal is to forecast the multivariate realized covariance matrix. Therefore, a natural question of generalization of described HAR model arises. Numerous approaches extending HAR model into the multidimensional case were introduced in the past. We will discuss the notable ones in the following paragraphs.

We need to consider several things as we generalize the HAR model to multiple dimensions. In the univariate case, we always assume non-negativeness of the variance. Similarly, in the multivariate setting, we assume that the covariance matrix is symmetric and positive semi-definite. However, such a condition on the covariance matrix is much more restrictive in the multivariate case and, therefore, more complicated to fulfill than the one in the one-dimensional case. Hence, it is not reasonable to construct a multivariate HAR model in the form of III.1 for each individual pairwise covariance term in the covariance matrix because such an approach would plausibly result in violation of the positive semi-definiteness of the covariance matrix. Bauer and Vorkink (2011) introduced different approach. First, they apply a matrix logarithm to a covariance matrix. Then they vectorize obtained matrix and define the multivariate HAR model on the log-scale, which will be shown in this thesis later. We will follow their approach and formally define mentioned transformation before continuing with the multivariate model itself.

Now, we introduce several functions that will be used to transform the realized covariance. Generally speaking, let us have a square matrix $\boldsymbol{A}$. We define matrix exponential function, *expm*, as:

$$\boldsymbol{B} = expm(\boldsymbol{A}) = \sum_{k=0}^{\infty} \frac{1}{k!}\boldsymbol{A}^k.$$

The reasoning behind such a transformation is that the matrix $\boldsymbol{B}$ would satisfy the positive semi-definiteness (even positive definiteness) condition if $\boldsymbol{A}$ is a symmetric square matrix. We also define the inverse function of *expm* as *logm*, $\boldsymbol{A} = logm(\boldsymbol{B})$. Properties of both *logm* and *expm* functions are discussed in the attachment A.3. Next, we will introduce the *vech* operator of a matrix, $\boldsymbol{a} = vech(\boldsymbol{A})$, which stacks the elements on and below the diagonal of $\boldsymbol{A}$ to a vector $\boldsymbol{a}$. Assuming that the matrix $\boldsymbol{A}$ has the dimensions $N \times N$, then the resulting vector would be of size $P = \frac{1}{2}N(N+1)$. There also exists a unique inverse operator *invvech*, which forms an $N \times N$ symmetric matrix, $\boldsymbol{A} = invvech(\boldsymbol{a})$.

After introducing all the relevant functions and operators, we can proceed to the definition of the multivariate HAR model. In the case of the univariate HAR model, we first constructed lagged and aggregated values of realized volatilities RV_t, RV_t^w, RV_t^m. The construction of the multivariate model follows the same principle as the one in the one-dimensional case. The one-weak lagged and aggregated realized covariance, $RC_t^{(w)}$, is defined as:

$$RC_t^{(w)} = \frac{1}{5}\left(RC_t + RC_{t-1} + RC_{t-2} + RC_{t-3} + RC_{t-4}\right).$$

Similarly, we can define the aggregated realized covariance matrix $RC_t^{(m)}$ over a month. The next step is to apply *logm* transformation to lagged and aggregated matrices as follows:

$$\boldsymbol{A}_{t+1} = logm(RC_{t+1}), \boldsymbol{A}_t = logm(RC_t), \boldsymbol{A}_t^{(w)} = logm(RC_t^{(w)}),$$
$$\boldsymbol{A}_t^{(m)} = logm(RC_t^{(m)}).$$

After these matrix logarithmic transformations we can apply *vech* operator to obtain the corresponding vectors $\boldsymbol{a}_{t+1}, \boldsymbol{a}_t, \boldsymbol{a}_t^{(w)}, \boldsymbol{a}_t^{(m)}$. The size of these vectors is $P = \frac{1}{2}N(N+1)$, where N is the number of assets in the considered portfolio.

III.1.2.1 MHAR-A

We will now consider several extensions of the univariate HAR model of Corsi (2009). The first possibility is to use the previously described transformation of the realized covariance and its lagged (aggregated) values and then construct separate linear models, where each element of $\boldsymbol{a}_{t+1}$ is dependent purely on the corresponding lagged values. We define this model as follows:

$$a_{t+1,(p)} = \alpha_p + \beta_p^{(d)} a_{t,(p)} + \beta_p^{(w)} a_{t,(p)}^{(w)} + \beta_p^{(m)} a_{t,(p)}^{(m)} + \varepsilon_{t+1,(p)}, p = 1, 2, \ldots, P, \quad \text{(III.2)}$$

where $a_{t,(p)}$ corresponds to the p-th element of the vector $\boldsymbol{a}_t$, similarly for the remaining terms $a_{t+1,(p)}, a_{t,(p)}^{(w)}, a_{t,(p)}^{(m)}, \varepsilon_{t+1,(p)}$ and their corresponding P-dimensional vectors $\boldsymbol{a}_{t+1}, \boldsymbol{a}_t^{(w)}, \boldsymbol{a}_t^{(m)}, \boldsymbol{\varepsilon}_t$. Such a model is still relatively simple as it requires estimating $3P$ coefficients for lagged values and P coefficients for the intercept term. However, there may be some information loss due to separate estimations. On the other hand, this model does not require too many parameters compared to other HAR extensions that will be covered below. We will denote this model MHAR-A, and we will more broadly discuss its performance in chapter V, where we compare it with different models.

III.1.2.2 MHAR-B

Another viable method similar to the presented one was developed by Chiriac and Voev (2011). We present here their original approach based on Cholesky decomposition. Nevertheless, they suggested implementing it using the logarithmic transformation (as it is applied in chapter V). Originally, one used Cholesky decomposition similar to the one introduced in the subsection II.7. However,Chiriac and Voev (2011) applied the classical Cholesky decomposition, not the LDL decomposition. Assuming that one makes use of the realized covariance matrix RC_t, we can construct its Cholesky decomposition as $RC_t = \boldsymbol{U}_t^T \boldsymbol{U}_t$, where $\boldsymbol{U}_t$ is the upper triangular matrix. Chiriac and Voev (2011) then define an $P = \frac{1}{2}N(N+1)$-dimensional vector $\boldsymbol{y}_t = vech(\boldsymbol{U}_t)$ and construct the corresponding multivariate HAR model as:

$$\boldsymbol{y}_{t+1} = \boldsymbol{\alpha} + \beta^{(d)} \boldsymbol{y}_t + \beta^{(w)} \boldsymbol{y}_t^{(w)} + \beta^{(m)} \boldsymbol{y}_t^{(m)} + \boldsymbol{\epsilon_{t+1}}, \quad \text{(III.3)}$$

where again $\boldsymbol{y}_t^{(w)}$ stands for aggregated lagged values over a week period (similarly for monthly aggregation $\boldsymbol{y}_t^{(m)}$), $\boldsymbol{\alpha}$ is a P-dimensional parameter and $\beta^{(k)}$, $k =$

d, w, m, are scalar coefficient corresponding to each lag. It is clear from the model structure that it provides only limited flexibility and possibly omits important relations. Nevertheless, it is applicable for high-dimensional cases because it has a very reasonable number of unknown parameters and offers an interesting alternative to the MHAR-A model. As it was previously mentioned, in this thesis we apply alternative MHAR-B model defined as:

$$\boldsymbol{a}_{t+1} = \boldsymbol{\alpha} + \beta^{(d)}\boldsymbol{a}_t + \beta^{(w)}\boldsymbol{a}_t^{(w)} + \beta^{(m)}\boldsymbol{a}_t^{(m)} + \boldsymbol{\varepsilon_{t+1}}, \tag{III.4}$$

where the *logm* transformation is used instead of the Cholesky decomposition.

III.1.2.3 MHAR-C

The last extension of the univariate HAR model to multiple dimensions we present here was introduced by Bauer and Vorkink (2011). It still follows the same structure of lagged values as the previously discussed models, but it incorporates cross-section parameters. We can construct the desired multivariate HAR model of Bauer and Vorkink (2011) as:

$$\boldsymbol{a}_{t+1} = \boldsymbol{\alpha} + \boldsymbol{B}^{(d)}\boldsymbol{a}_t + \boldsymbol{B}^{(w)}\boldsymbol{a}_t^{(w)} + \boldsymbol{B}^{(m)}\boldsymbol{a}_t^{(m)} + \boldsymbol{\varepsilon_{t+1}}, \tag{III.5}$$

where $\boldsymbol{a}_t, \boldsymbol{a}_t^{(w)}, \boldsymbol{a}_t^{(m)}$ are vectors of lagged and aggregated values of dimension P and $\boldsymbol{B}, \boldsymbol{B}^{(w)}, \boldsymbol{B}^{(m)}$ are matrices of unknown coefficients of the dimension $P \times P$. It can be seen that such a model has a very large number of parameters that have to be estimated. To be more precise, we have P parameters for the intercept term and $3P^2$ coefficients corresponding to the remaining explanatory variables. Therefore, this model is barely usable in practice as the enormous number of unknown parameters does not make it possible to apply such an approach for a greater number of assets N. Hence, Bauer and Vorkink (2011) continue to develop their model by using principal component analysis for $\boldsymbol{a}_t$ (and $\boldsymbol{a}_t^{(j)}, j = w, m$) aiming at a variable reduction. This procedure significantly reduces the number of parameters because it targets the correlated cross-section coefficients in the model, which is due to the covariance modeling very large. The number of parameters in the final model then depends on the selected number of principal components. For example, if we use two principal components in our model, the final model would end up with $7P$ parameters. Model of Bauer and Vorkink (2011) incorporating principal component analysis will be denoted MHAR-C in this thesis.

Bauer and Vorkink (2011) also include other explanatory variables to the model, e.g., T-bill interest rate, etc. They specifically use variables that are believed to forecast stock returns and volatility well according to economic theory. They then construct a factor model including both the lagged values and the economic-based explanatory variables. We will omit these additional variables and exclusively focus on the lagged and aggregated values in the model.

III.1.2.4 Different HAR Generalizations

Several other methods generalizing the univariate HAR model of Corsi (2009) were introduced in the past. Čech and Baruník (2017) have generalized the

HAR model by Chiriac and Voev (2011) based on Cholesky decomposition to the so-called GHAR model (Generalized HAR model). They have used seemingly unrelated regression (SUR) to extract information from the correlation of error terms to improve forecasting ability. This approach enables a certain common structure of the realized covariance in the framework of the model in some sense. However, as the error terms of the HAR model are generally heteroscedastic, this model could possibly struggle with efficiency. Another very complex approach is introduced by Hong and Hwang (2022) assuming exponentially decaying coefficients in the multivariate HAR model.

III.1.2.5 Bias Correction

The last thing that we need to consider regarding the multivariate HAR model is the bias of described methods. Due to the logarithmic transformation, we can expect particular bias in our predictions implied by Jensen's inequality. Both Chiriac and Voev (2011) and Bauer and Vorkink (2011) discuss the application of specific bias corrections. The correction described by Bauer and Vorkink (2011) is not suggested to correct the forecast. Chiriac and Voev (2011) extends this correction for forecast correction purposes. This approach is based on the mean or median of the fraction:

$$\xi_{t,(n)} = \frac{\sqrt{RC_{t,(n,n)}}}{\sqrt{\widehat{RC}_{t,(n,n)}}}, \tag{III.6}$$

where $RC_{t,(n,n)}$ corresponds to the n-th diagonal element of the matrix RC_t. They also critically discuss the potential use of such a bias correction. Their study does not find empirical justification for applying such correction unless the bias is fairly large. They measure the bias using the mean of the fraction series $\xi_{t,(n)}$ and its overall deviation from 1 for each n (and different ranges of values used in the calculation of mean). In practice, this correction based on their findings could, in fact, have even negative effects on the resulting forecasts when the bias is relatively small. Therefore, we decided not to apply the bias correction in our case.

III.2 Alternative Models

There are also various alternatives to model the dynamics of realized covariance matrices. We will shortly describe two different approaches in this subsection. Although non of these methods will be later used in the empirical application, we decided to present these methods as their role in the context of the realized covariance forecasting is rather important.

The first potential competitor of the HAR model is the Wishart autoregressive (WAR) model introduced by Gourieroux et al. (2009). The advantage of the WAR model is that it naturally possesses positive definiteness and symmetric properties. Common estimation procedures of the WAR model include a method of moments and maximum likelihood estimation. Possibly major disadvantage of the WAR model in comparison with the HAR model is that it cannot capture the long-memory dependence. We have already discussed properties of volatility observed in practice at the beginning of this chapter, including the importance of the long-memory property.

The second popular model that can be considered in the context of modeling the realized covariance matrix is the vector autoregressive fractionally integrated moving average (VARFIMA) model. This model was introduced by Chiriac and Voev (2011). Similarly to their multivariate HAR model, they define the VARFIMA model using Cholesky factorization and model the vectorized upper triangular matrix $\boldsymbol{U}_t$. The VARFIMA is also capable of incorporating the long-memory property of realized covariance and, therefore, it is a direct competitor to the multivariate HAR models. Chiriac and Voev (2011) in their empirical study found that the VARFIMA model slightly outperformed their multivariate HAR model. However, it should be noted that their multivariate HAR model significantly differs from the one introduced by Bauer and Vorkink (2011) or the univariate HAR models applied to individual elements of the realized covariance. Hence, the comparison of all these forecasting methods presents an interesting topic to be analyzed.

IV. Portfolio Optimization and Forecast Evaluation

In the previous chapter, we presented various forecasting methods of the realized covariance (and different estimators of the integrated covariance). Now, we will be interested in the application and evaluation of resulting forecasts. One such application is portfolio optimization. In this chapter, we will describe the portfolio optimization problem, including the transaction cost constraint. The optimization results will also be used to compare previously described estimators and forecasting methods. As a part of this chapter, we also present a statistical evaluation method that we will use to compare different models in chapter V.

IV.1 Portfolio Optimization Including Transaction Costs

We present a portfolio optimization problem with constraint limiting transaction costs. We will be purely interested in the global minimum variance portfolio (GMVP) because it does not require expected return specification. It was shown in various studies that volatility has a significantly more important role than the expected return in the presence of high-frequency data. Hence, we can solely focus on the integrated covariance estimators that are our main topic throughout the thesis and omit the estimation of the expected return.

There are two main constraints in the most common formulation of the GMVP optimization problem. The first one limits the overall value of usable funds, and the second one describes whether short sales are allowed or not. In general, we will assume that short sales are allowed. However, the possibility of short positions in the optimization problem also has an impact on the second constraint, which is usually mathematically formulated as $\sum_{n=1}^{N} w_{(n)} \leq W$, where $w_{(n)}$ is weight corresponding to asset n (or in other words the portion of wealth that is invested into the asset n) and W is a total disposable amount of wealth. For the rest of this thesis, we will assume that the total disposable wealth of an investor is equal to 1. Under the classical formulation of GMVP with short sales allowed, this condition is not very strict. It is described by Mansini et al. (2015) and Fan et al. (2012) that imposing the so-called booksize constraint (sometimes also called gross exposure constraint):

$$\sum_{n=1}^{N} |w_{(n)}| \leq s, \tag{IV.1}$$

in combination with the classical sum of weights enables us to limit the short positions. The parameter s is then called the gross exposure parameter. It corresponds to the total allowed exposure. One can control the portion of short positions in the portfolio by correspondingly regulating the value of the gross exposure parameter s. Classical Markovitz portfolio would correspond to the case when $s = \infty$. On the other hand, the choice $s = 1$ implies that the short-sales are not allowed.

Secondly, we will focus on the transaction costs. There are three main categories of transaction costs related to an individual trade. First, there is a commission, which is basically a fee that the broker charges for trade execution. The second one is a bid/ask spread that refers to the difference between the buying price and the price if we immediately sell the same stock. Lastly, there is a market impact that is linked to the trading of multiple stocks on the same asset. By executing orders on various stocks of the same company, the investor changes the price and, therefore, adds extra cost to the trade. Sometimes, the fourth type of transaction cost is specified, which refers to the lost profit from the unrealized trading opportunity. In literature, this fourth type of transaction cost is usually referred to as an opportunity cost. However, we do not specifically target lastly mentioned transaction cost.

A rational investor would like to decrease the transaction costs as much as possible. And as each trade is accommodated with given transaction costs, it is reasonable to limit the frequency of portfolio changes. Such a property is ensured by imposing a so-called turnover constraint:

$$\sum_{n=1}^{N} |w_{t,(n)} - w_{t-1,(n)}| \leq L \sum_{n=1}^{N} |w_{t-1,(n)}|, \tag{IV.2}$$

where $w_{t,(n)}$ is a position in the asset n at time t and L is a turnover constraint parameter. Such a condition limits the changes in the portfolio and, therefore, limits the transaction costs. The choice of the turnover constraint parameter L is quite ambiguous as it is usually set to the value between 0 and 1 depending on the desired transaction costs restriction.

Combining all the described constraints with the classical GMVP and also introducing time dependence of weights into the optimization problem, we formulate the optimization problem using a one-day ahead forecast of the covariance matrix $\widehat{\boldsymbol{\Omega}}_{t+1|t}$:

$$\begin{aligned} \min_{\boldsymbol{w}_{t+1}} \quad & \boldsymbol{w}_{t+1}^T \widehat{\boldsymbol{\Omega}}_{t+1|t} \boldsymbol{w}_{t+1} \\ s.t. \quad & \sum_{n=1}^{N} w_{t+1,(n)} \leq 1 \\ & \sum_{n=1}^{N} |w_{t+1,(n)}| \leq s, \\ & \sum_{n=1}^{N} |w_{t+1,(n)} - w_{t,(n)}| \leq L \sum_{n=1}^{N} |w_{t,(n)}|, \\ & \boldsymbol{w}_{t+1} \in \mathbb{R}^N, \end{aligned} \tag{IV.3}$$

where $\boldsymbol{w}_{t+1} = (w_{t+1,(1)}, w_{t+1,(2)}, \ldots, w_{t+1,(N)})^T$, $\boldsymbol{w}_t = (w_{t,(1)}, w_{t,(2)}, \ldots, w_{t,(N)})^T$, are an N-dimensional vectors of weights at time $t+1$, t respectively. $\widehat{\boldsymbol{\Omega}}_{t+1|t}$ is once again one-day ahead forecast of the integrated covariance matrix. Parameter s denotes the gross exposure parameter, and L is the turnover constraint parameter.

We denote the resulting weights for which the minimum is attained as $\widehat{\boldsymbol{w}}_{t+1}$. In general, this optimization problem can be extended for k-days ahead forecasts of the covariance $\widehat{\boldsymbol{\Omega}}_{t+k|t}$ and corresponding portfolio weights $w_{t+k,(n)}, n = 1, 2, \dots, N$.

The turnover constraint depends on portfolio weights from a previous period, as could be seen in the inequality IV.2. In chapter V, we will be mainly interested in portfolio optimization starting with some initial weights. It is necessary to determine these weights beforehand because we assume that we do not have any particular positions in any of the assets from the considered portfolio. It is a common practice to set initial weights to the weights of the equally weighted portfolio $\boldsymbol{w}^{(eq)} = (w_{(1)}^{(eq)}, w_{(2)}^{(eq)}, \dots, w_{(N)}^{(eq)})^T$, $w_{(n)}^{(eq)} = \frac{W}{N}$ for $\forall n$. Therefore, we can reformulate the optimization problem described in IV.3 using the equal weights that substitute the role of weights from the previous period:

$$
\begin{aligned}
\min_{\boldsymbol{w}_{t+1}} \quad & \boldsymbol{w}_{t+1}^T \widehat{\boldsymbol{\Omega}}_{t+1|t} \boldsymbol{w}_{t+1} \\
s.t. \quad & \sum_{n=1}^{N} w_{t+1,(n)} \leq 1 \\
& \sum_{n=1}^{N} |w_{t+1,(n)}| \leq s, \\
& \sum_{n=1}^{N} |w_{t+1,(n)} - w_{(n)}^{(eq)}| \leq L \sum_{n=1}^{N} |w_{(n)}^{(eq)}|, \\
& \boldsymbol{w}_{t+1} \in \mathbb{R}^N,
\end{aligned}
\tag{IV.4}
$$

where again $\boldsymbol{w}_{t+1}$ is an N-dimensional vector of weights at time $t+1$, $\widehat{\boldsymbol{\Omega}}_{t+1|t}$ is one-day ahead forecast of the integrated covariance matrix estimator and $\boldsymbol{w}^{(eq)}$ is a vector of weights for equally weighted portfolio.

IV.2 Forecast Evaluation

At last, we present various methods, which will be later in the chapter V used to compare forecasts. We will mainly focus on the evaluation based on the optimization problem formulated in the previous subsection and the statistical evaluation.

IV.2.1 Economic Forecast Evaluation

The most straightforward way of comparing forecasts based on the results of GMVP is to compare the attained minima for each model-estimator combination directly. This method seems very tempting. However, drawing any reasonable conclusions from this approach is difficult as it does not provide common ground for the model-estimator combinations.

Second, a more sophisticated comparison method is presented in Bollerslev et al. (2019). They use optimal weights $\widehat{\boldsymbol{w}}_{t+1}$ that minimize the optimization problem (in our case IV.4) for each model-estimator combination and then compare their performance by calculating:

$$\widehat{\boldsymbol{w}}_{t+1}^T RC_{t+1} \widehat{\boldsymbol{w}}_{t+1}, \tag{IV.5}$$

where RC_{t+1} is the realized covariance for the time period $t+1$ (based on 5-min intraday data). Hence, the RC_{t+1} is constructed from the data that were not used during the estimation of the forecasting models as we evaluate the out-of-sample forecasts. The realized covariance matrix RC_{t+1} is taken as a proxy for the integrated covariance at time $t+1$. This approach enables us to compare the obtained values reasonably as they are all connected through the realized covariance matrix for the time period $t+1$.

IV.2.2 Statistical Forecast Evaluation

Alternatively, several different error measures could be used to compare the precision of forecasts. We will present the Frobenius norm solely as it seems to be the most common choice among various already referenced papers concerned with models of the realized covariance. As we are performing (out-of-sample) forecast, we will be interested in the precision of the one-day ahead forecast $\widehat{\boldsymbol{\Omega}}_{t+1|t}$. We will compare this forecast with the integrated covariance proxy calculated based on the data that were observed during day $t+1$, $\widehat{\boldsymbol{\Omega}}_{t+1}$. However, it is important that data for the day $t+1$ were not used to fit the model because, once again, we are interested in the forecast evaluation. We will select two different choices of the integrated covariance proxy initially suggested by Čech and Baruník (2017). The first choice of proxy will be simply the realized covariance (see II.1), and the second one will be the realized kernel covariance (see II.3). We then calculate the Frobenius norm of the matrix $\boldsymbol{C}_{t+1} = \widehat{\boldsymbol{\Omega}}_{t+1|t} - \widehat{\boldsymbol{\Omega}}_{t+1} = (c_{t+1,(m,n)})_{m,n=1,2,\ldots,N}$ as:

$$\|C_{t+1}\|_F = \sqrt{\sum_{m=1}^{N} \sum_{n=1}^{N} |c_{t+1,(m,n)}|^2}. \tag{IV.6}$$

Contrary to the economic forecast evaluation methods, this approach has a formal statistical justification as it coincides with the RMSE (root-mean-square error) evaluation of the forecast. On the other hand, this statistical comparison has one possible disadvantage, which is actually shared with the economic evaluation. The choice of the covariance proxy is obviously not unique, but some proxy has to be selected as the integrated covariance is latent. Using two specific choices may not result in a fair evaluation because the proxies themself can have poor precision, and they will likely favor the forecasts based on the corresponding estimator. Nevertheless, the solution to this issue surely does not consist in extending the number of proxies just by itself. It can be seen that both economic and statistical comparison methods face similar problems regarding the choice of proxy. However, two different choices of covariance proxy in the statistical evaluation should be adequate for our purposes.

V. Empirical Study

In this chapter, we will focus on the application of presented models and estimators on real-life data. We will first introduce the data and procedures used during the calculation itself. In the second subsection, we will show the results and compare different models used to forecast different integrated covariance estimators.

V.1 Data and Technical Details

Our main goal is to estimate the integrated covariance matrix of vector time series and then construct a one-day ahead forecast based on the models presented in the chapter III. We consider 5-min intraday data of stock prices for 50 different companies. We decided to use 5-minute intervals as they provide the properties of high-frequency data and also tend to be a bit less affected by the microstructure noises. We do not observe the before-hours, after-hours, or overnight trading data. This can be quite limiting for certain integrated covariance estimators, especially those developed for the tick-by-tick data preferably observed continuously, including overnight observations (this concerns mainly the two-time scale realized covariance estimator and its robust modification).

Our data consist of 50 randomly selected companies from the S&P500 index. We decided to analyze larger companies as their intraday prices are usually more consolidated. Nevertheless, we still had to deal with missing observations in our case too. We will work with intraday observations from 12 August 2022 till 30 December 2022, a total of 98 business days. We used Python programming language to obtain the intraday observations as it offers an open-source library *yfinance*, which connects users to Yahoo! Finance APIs. Users can then access different trading data, including the intraday data for different time intervals (1-min, 2-min, etc.). The only problem is the maximum time horizon one can obtain. The number of accessible business days differs for particular time intervals; however, for the 5-min time interval, it is capped at 60 days. This means that in a given time, we can access intraday data only for the last 60 business days. Therefore, it is necessary to repeat the procedure and extend the data set over time to obtain more observations. Subsequently, we also use a Python environment to perform cleaning of obtained data (in our case, replacing missing values with linearly interpolated values) and for the calculation of logarithmic returns. The remaining calculations were performed in the R software environment because it has more tools for intraday data analysis. A prime example of such a tool is a library called *highfrequency*, which contains various functions for calculating different integrated covariance estimators. This package also includes a function that transforms a non-positive semi-definite matrix into a positive semi-definite matrix. We mention this fact because this function often does not perform as expected, and the results of the function are numerically unstable. Hence, we would recommend using different algorithms to ensure the symmetric positive semi-definiteness of covariance matrices since the positive semi-definiteness is pivotal not only from the theoretical point of view but also because of the logarithmic matrix transformation in the model development.[1] Models described in the chapter III were handled manually using common R functions like *lm*.

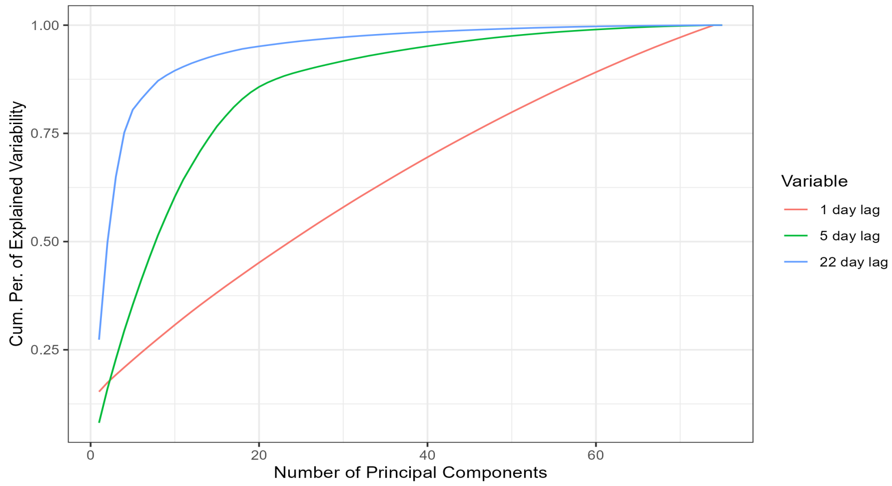

Figure V.1: Graph shows the cumulative explained variability of the principal components in the MHAR-C model combined with the realized covariance estimator for 50 assets. Each curve corresponds to the explained variability of the first principal components for different lagged (aggregated) values of the realized covariance estimator.

Moreover, we present selected company tickers with corresponding names and sectors of each analyzed company in the attachment A.1. Data, together with the briefly described code, will be attached as the electronic attachment.

Remark. For the construction of the MHAR-C model, we use only the first two principal components. If we omitted the principal component analysis in this particular model, we would end up with 4 878 150 parameters for the 50×50 covariance matrix. Estimating such a large number of parameters is completely unacceptable because it would result in serious overfitting. Therefore, the principal component analysis is needed as described in chapter III. Using the first two principal components, we would end up with 8 925 (using three principal components, we would end up with 12 750). Although this number of parameters is still large, we were able to reduce it dramatically.

Remark. The first two (three) principal components in the case of 50 analyzed assets explained roughly $17\%(19\%), 16\%(23\%)$ and $50\%(65\%)$ for 1-day lagged, 5-days lagged aggregated, and 22-days lagged aggregated values, respectively. These percentages are significantly larger if we consider a smaller portfolio of five assets. For five randomly selected assets, the explained variability using the first two principal components is $36\%, 52\%,$ and 83% for 1-day lagged, 5-days lagged aggregated, and 22-days lagged aggregated values, respectively. Such results mostly coincide with the results obtained by Bauer and Vorkink (2011). We could consider more components in the MHAR-C model to increase the explained

[1]One can use for example function *nearest_spd* from library *pracma*.

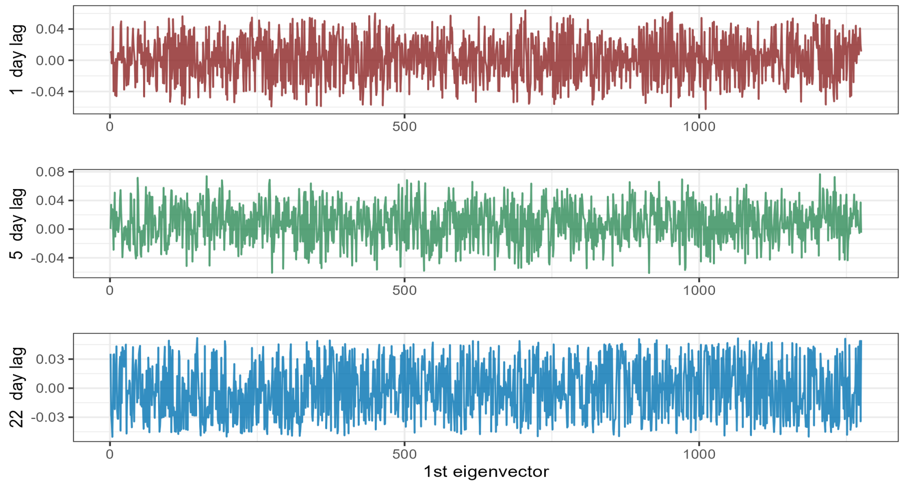

Figure V.2: Graphs show elements of the first eigenvectors arising from the principal component analysis for each lagged variable of the realized covariance matrix in the context of 50 assets.

variability, but such a procedure would necessarily drastically enlarge the number of parameters in the model. Therefore, we decided to stay with only two components as it was suggested by Bauer and Vorkink (2011).

V.2 Comparison of Forecasts

In the following part, we will analyze and compare different approaches to integrated covariance modeling. We first calculate different estimators of the integrated covariance described in chapter II based on the given data. To avoid any misunderstanding, we summarize all the used estimators together with the corresponding notation in the form of the following list:

- Realized Covariance (RC),
- Realized Bi-power Covariance (BP),
- Realized Kernel Covariance (RK),
- Modulated Realized Covariance (MRC),
- Realized Threshold Covariance (RTC),
- Two-Time Scale Realized Covariance (TS),
- Robust Two-Time Scale Realized Covariance (RTS),
- Realized Cholesky Covariance ($CholC$),
- POET Estimator ($POET$).

We will obtain one estimate of the integrated covariance (with dimensions 50×50) for each day and each choice of estimator. This means that we will have nine different estimates of the integrated covariance for every single day from the 98 observed business days sample. Subsequently, we separate the last observation period from the rest because we want to have one out-of-sample observation that we will use for both economic forecast evaluation using the GMVP and statistical forecast evaluation using the Frobenius norm as described in chapter IV.

After removing the last estimated integrated covariance matrices, we are left with estimates over 97 days. We will apply four different forecasting methods to those slightly truncated datasets. A list of the utilized forecasting models can be seen below:

- Random walk assumption-based model (RW),
- The univariate model of Corsi (2009) applied to each element of the logm-transformed estimated integrated covariance matrices (MHAR-A),
- The multivariate HAR model of Chiriac and Voev (2011) (MHAR-B),
- The multivariate HAR model of Bauer and Vorkink (2011) (MHAR-C).

Using the described models, we perform a 1-day ahead forecast for each model in combination with each integrated covariance estimator (each model-estimator combination). We will use the results arising from different model-estimator combinations and compare them using different evaluation techniques described in the chapter IV.

V.2.1 Economic Comparison

We continue by comparing the model-estimator combinations using the economic approach described in the previous chapter IV. We will solve several portfolio optimization tasks described in IV.4 with different choices of the gross exposure parameter s. In this subsection, we present the results for the three most common choices of parameter s. Specifically, s equal 1, 2, and 3, where the case $s = 1$ implies that the short-sales are not allowed. The other two cases restrict the portion of shorted stocks in the portfolio. We also present the results of portfolio optimization for $s = 1.5, 2.5, \infty$ in the attachment A.2. The turnover parameter L will be generally set to a $\frac{1}{2}$ as we want to restrict the transaction costs only moderately. Our portfolio is assumed to be composed of 50 individual assets with a covariance structure given by integrated covariance estimators.

Let us begin by presenting the attained minima of the objective function, $\boldsymbol{w}_{t+1}^T \widehat{\boldsymbol{\Omega}}_{t+1|t} \boldsymbol{w}_{t+1}$. As we said in the previous chapter, this approach is not ideal for comparing different methods. However, we want to somewhat present the portfolio optimization results to outline the context of our framework. It could be seen in the table V.1 that for the majority of estimators, the results are not very dependent on the choice of gross exposure parameter s. The increasing value of this parameter in the majority of cases does not affect the final results. Such results are quite unexpected because we would expect that setting the parameter s to a higher value should improve the final result of minimization. To be more precise, the increasing value of s enables more short positions in a portfolio and, hence, extends the space of parameters, which should rationally result in the same or improved results in terms of minimization. However, this is not the case because the turnover constraint limits the transaction costs and, therefore, imposes additional limitations on resulting optimal weights. Overall, the best results in terms of the attained minima were obtained using the *POET* estimator, which attained the lowest values among the considered model-estimator combinations. However, the results of the attained minima are not necessarily conclusive, as this approach does not take into account any proxy of the integrated covariance.

	POET	*RTS*	*TS*	*RTC*	*MRC*	*CholC*	*BP*	*RK*	*RC*
					$s = 1$				
RW	0.18	11.73	25.11	39.68	35.34	15.55	11.81	103.56	57.00
MHAR-A	1.08	5.64	10.78	74.74	44.76	19.49	49.34	21.76	156.13
MHAR-B	0.54	26.51	40.67	66.95	66.59	53.33	46.65	33.55	84.42
MHAR-C	0.97	28.11	23.82	85.10	43.71	49.70	55.90	18.26	151.64
					$s = 2$				
RW	0.18	11.73	25.11	39.68	35.34	15.55	11.81	103.56	57.00
MHAR-A	1.08	5.64	10.78	74.74	31.21	22.03	49.34	21.76	156.13
MHAR-B	0.54	26.51	40.67	66.95	66.59	53.33	46.65	41.01	84.42
MHAR-C	0.97	27.69	23.82	85.10	43.71	49.70	55.90	18.26	151.64
					$s = 3$				
RW	0.18	11.73	25.11	39.68	35.34	15.55	11.81	103.56	57.00
MHAR-A	1.08	5.64	10.78	74.74	31.21	22.03	49.34	21.76	156.13
MHAR-B	0.54	26.51	40.67	66.95	66.59	53.33	46.65	41.01	84.42
MHAR-C	0.97	27.69	23.82	85.10	43.71	49.70	55.90	18.26	151.64

Table V.1: The attained minima in the portfolio optimization problem described in the chapter IV for three different choices of gross exposure parameter s $(1, 2, 3)$. All the presented results are displayed on the 10^7 scale.

Next, we will use the optimal weights $\widehat{\boldsymbol{w}}_{t+1}$ obtained as the solution of the portfolio optimization for each model-estimator combination and calculate the following:

$$\widehat{\boldsymbol{w}}_{t+1}^T RC_{t+1} \widehat{\boldsymbol{w}}_{t+1},$$

where $\widehat{\boldsymbol{w}}_{t+1}$ are the optimal weights in which portfolios attend the minima presented in the table V.1 (as theoretically described in the chapter IV). The realized covariance RC_{t+1} serves as a proxy of the integrated covariance at time $t+1$. It is much clearer how to compare different methods using this approach because all the resulting values are now connected by the realized covariance matrix from the time period $t+1$ (in our case, day 98). Hence we have a common link between the evaluation of all the selected methods. Therefore, comparing such results is much more reasonable than purely the attained minima of the portfolio optimization.

Results for each model-estimator combination are presented in the table V.2.

	EqW	*POET*	*RTS*	*TS*	*RTC*	*MRC*	*CholC*	*BP*	*RK*	*RC*
					$s = 1$					
RW	139.84	24.10	22.07	21.94	20.65	20.36	19.67	25.74	22.73	24.45
MHAR-A	139.84	17.56	28.50	14.33	18.52	17.27	14.34	21.62	13.32	20.62
MHAR-B	139.84	19.30	18.88	21.99	18.51	18.70	19.42	20.13	17.57	18.00
MHAR-C	139.84	21.12	21.80	16.53	20.20	19.09	19.57	22.57	13.73	18.80
					$s = 2$					
RW	139.84	24.10	22.07	21.94	20.65	20.36	19.67	25.74	22.73	24.45
MHAR-A	139.84	17.56	28.50	14.33	18.52	12.94	14.21	21.62	13.32	20.62
MHAR-B	139.84	19.30	18.88	21.99	18.51	18.70	19.42	20.13	18.34	18.00
MHAR-C	139.84	21.12	20.88	16.53	20.20	19.09	19.57	22.57	13.73	18.80
					$s = 3$					
RW	139.84	24.10	22.07	21.94	20.65	20.36	19.67	25.74	22.73	24.45
MHAR-A	139.84	17.56	28.50	14.33	18.52	12.94	14.21	21.62	13.32	20.62
MHAR-B	139.84	19.30	18.88	21.99	18.51	18.70	19.42	20.13	18.34	18.00
MHAR-C	139.84	21.12	20.88	16.53	20.20	19.09	19.57	22.57	13.73	18.80

Table V.2: Table contains the evaluation of model-estimator combinations based on the calculation of $\widehat{\boldsymbol{w}}_{t+1}^T RC_{t+1} \widehat{\boldsymbol{w}}_{t+1}$, where RC_{t+1} is the realized covariance for the out-of-sample period $t+1$ (day 98) and $\widehat{\boldsymbol{w}}_{t+1}$ are the optimal weights obtained as a result of the optimization problem described in IV.4 (the portfolio weights corresponding to the attained minima presented in the table V.1). The table shows the performance based on the described evaluation method for three selected choices of gross exposure parameter s $(1, 2, 3)$. The *EqW* column corresponds to the portfolio with equal weights. All the presented results are displayed on the 10^6 scale.

The results presented in the table V.2 are obviously not very dependent on the choice of gross exposure parameter s. Such results could be expected as the weights used for calculating the criterion in the table V.2 are the same as the ones used in the table V.1 of attained minima. We will discuss the individual performance of estimators and forecasting models first. Afterward, we will be interested in the comparison of model-estimator combinations using economic comparison.

V.2.1.1 Economic Comparison of Estimators

We begin with the comparison of individual estimators. The best way to compare the presented estimators independently on the selected forecasting models is by using the random walk assumption-based model (RW). Such a choice of forecasting model does not greatly influence the evaluation; hence, we analyze the forecasting ability of each estimator solely. It can be seen that all the estimators are capable of outperforming the results of an equally weighted portfolio (EqW). Such results implicate rather significant predictive ability of individual estimators, at least in the context of one-day ahead forecasts. Table V.2 shows that the results are not affected by the choice of s. Therefore, we can compare the results of individual estimators irrespective of the value of parameter s.

The realized Cholesky covariance ($CholC$) estimator performs the best in comparison with all the presented estimators based on the evaluation using the weights from the portfolio optimization. The second-best estimator is the modulated realized covariance (MRC), and the third-best is the realized threshold covariance estimator (RTC). Based on these results, it seems that more robust estimators with respect to the microstructure noises and non-synchronicity perform better than simpler estimators like the realized covariance (RC) or the realized bi-power covariance (BP). Interestingly enough, the two-time scale realized covariance (TS) and its robust modification (RTS) perform relatively well despite the fact that they were developed for the tick-by-tick data. These estimators were able to outperform all the simpler estimators, e.g., the realized covariance (RC) or the realized bi-power covariance (BP), but also some of the more complex estimators like the $POET$ estimator.

On the other hand, the worst-performing estimator is the realized bi-power covariance (BP), which performs even worse than the realized covariance (RC) estimator. However, both of these estimators are not significantly outperformed by their counterparts. The results for all estimators are very competitive, but it seems that more complex estimators like the $CholC$, MRC, or RTC do achieve better results when comparing simply the predictive ability of individual estimators.

V.2.1.2 Economic Comparison of Forecasting Models

Comparing individual estimators is still quite straightforward by using the RW model. On the other hand, discussing the economic evaluation for forecasting models is much more difficult because the results are substantially dependent on the choice of the estimator. However, there are a few interesting points that we would like to point out.

First, we can observe in V.2 that MHAR models are able to outperform the RW model for the majority of estimators. Only exceptions could be seen when using the two-time scale realized covariance (TS) and its robust version (RTS). However, all the estimators are able to predict relatively well on their own (in combination with the RW model). This statement is supported by the fact that the RW model is capable of significantly outperforming the equally weighted portfolio and realizing fairly competitive results compared to MHAR models. The MHAR models are beneficial in most cases. On the other hand, combining certain estimators with specific models sometimes results in inferior performance, which will be more broadly discussed in the next subsection. One such example is the robust two-time scale realized covariance in combination with the MHAR-A model.

Secondly, it seems that the MHAR-A model performs more consistently for various choices of estimators than the remaining forecasting models. The MHAR-A model outperforms its competitors for five different choices of covariance estimators and, therefore, achieves the best results based on the economic evaluation. Both MHAR-B and MHAR-C models are also able to outperform the RW model for almost any choice of the covariance estimator. However, the results of these models are surely a bit more volatile than the ones of the MHAR-A model. Overall, it seems that there is a significant contribution of MHAR models when it comes to covariance forecasting.

V.2.1.3 Economic Comparison of Model-Estimator Combinations

Next, we will discuss in detail the results of the model-estimator combinations. An interesting phenomenon that we touched on earlier is that the results are significantly influenced by the choice of the estimator, as some forecasting models perform better in combination with certain specific estimators. For example, MHAR-B outperforms the other MHAR models in combination with the robust two-time scale realized covariance (RTS) or the realized threshold covariance (RTC). The MHAR-C model achieves the best results compared to the remaining models for the realized kernel covariance (RK) and the realized covariance (RK). However, in both of those cases, it still does not outperform the MHAR-A model.

The RW model performs comparatively well in combination with the realized threshold covariance (RTC) and the realized Cholesky covariance ($CholC$). It can also outperform some MHAR models in combination with the two-time scale realized covariance (TS) and the robust two-time scale realized covariance (RTS). In general, the random walk assumption-based model RW performs better in comparison with the remaining forecasting models when combined with more complex estimators.

The MHAR-A model in combination with the modulated realized covariance (MRC) performs the best across all the model-estimator combinations. This combination significantly outperforms the MRC in combination with competing forecasting models. However, when the gross exposure parameter is equal to one, it does not achieve better results than, for example, the realized kernel estimator (RK). A slightly worse overall performance could be observed for the combination of the realized kernel estimator (RK) and the MHAR-A model, followed by the RK estimator combined with the MHAR-C model. Very impressive results are also achieved using the $CholC$ and the TS estimators combined with the MHAR-A model. According to the economic evaluation, all of these models perform significantly better than the competing model-estimator combinations.

It is difficult to decide which model-estimator combination is the best based on the presented economic evaluation as the results depend on many factors, like the choice of gross exposure parameter s. However, the trend occurring in the evaluation of the model-estimator combinations seems to be as follows. The RW model usually performs well compared to the remaining models in combination with the more complex estimators (the robust two-time scale realized covariance, RTS, or the realized Cholesky covariance, $CholC$). The simpler the estimator gets, the better performance we achieve using the more complex forecasting models compared to the RW model, e.g., the realized covariance (RC) or the realized kernel estimator (RK). This trend is more predominant when we compare the RW model with the MHAR-B and MHAR-C models. The MHAR-A model works well with more complex estimators like the realized Cholesky covariance ($CholC$).

It seems that the MHAR-A model combined with more complex estimators that consider the impact of microstructure noises (MRC or $CholC$) achieves the best results in comparison with the remaining models. On the other hand, some estimators do not significantly benefit from the use of MHAR models, e.g., the robust two-time scale realized covariance (RTS). In such cases, the RW model seems to achieve very competitive results.

To summarize our results, it can be seen that all three MHAR models benefit from the use of certain specific estimators. The MHAR-B model outperforms the competing MHAR models when combined with the robust two-time scale realized covariance (RTS) or the realized threshold covariance (RTC). The MHAR-C model benefits substantially from the combination with the realized covariance estimator (RC) or the realized kernel estimator (RK). However, in the case of both of these estimators, it does not achieve the best overall results. The MHAR-A model outperforms the remaining models in combination with the majority of estimators. The most significant differences in the performance could be observed for the modulated realized covariance (MRC) and the realized Cholesky covariance ($CholC$) too.

In general, the MHAR models outperform the RW model in most cases when comparing single estimator choices. The best performance for a single model-estimator combination is achieved by the MHAR-A model with the MRC estimator. For s equal to one, the best-performing model-estimator combination is the MHAR-A model with the realized kernel estimator (RK). However, the MHAR-C model, combined with the RK estimator, is also very close in terms of the economic evaluation to the MHAR-A model with the RK estimator.

To be fair, this evaluation can be quite misleading. If we were to compare our models only based on economic evaluation, we could omit the quality of the prediction from a different perspective, like the error measures. For those reasons, we will present the statistical evaluation of the selected models in the next subsection.

V.2.2 Statistical Comparison

In the following subsection, we will be interested in the statistical evaluation of presented model-estimator combinations using the Frobenius norm as described in the chapter IV. We again present our results in the form of a table, where we show the resulting values of the Frobenius norm for different models and integrated covariance estimators.

First, we will look at the results where the realized covariance estimator is used as a proxy for the integrated covariance. In the table V.3, we show the resulting values of the Frobenius norm for each forecast. It is quite likely that this method favors the realized covariance estimator because this specific estimator is also used as the proxy. Therefore, it is unsurprising that the realized covariance (RC) achieves the best results.

Both MHAR-A and MHAR-C models achieve the smallest values of the Frobenius norm in connection with the RC estimator. Nevertheless, we will mainly focus on comparing the remaining estimators and models. The random walk assumption-based forecasts perform relatively well compared to the more sophisticated MHAR models, even in terms of statistical evaluation. In the previous subsection, we saw that the RW model achieved competitive results regarding economic evaluation. However, we expected a more significant difference when using the comparison based on the Frobenius norm. In some cases, the RW forecasts perform basically the same as the remaining models, e.g., the $POET$ estimator.

The estimators developed for the tick-by-tick data (the two-time scale realized covariance, TS, and the robust two-time scale realized covariance, RTS) are able to achieve very competitive overall results in comparison with the remaining estimators. For example, they outperform the $POET$ estimator, which achieved the worst results across all the forecasting models except the RW model. The realized kernel estimator (RK) combined with the RW model achieves the worst results. This indicates that the predictive ability of the individual estimator is rather poor. However, the combination of this particular estimator with MHAR models produces very competitive results. For example, the MHAR-A model with the RK estimator performs second-best in terms of the statistical evaluation out of all the model-estimator combinations.

It is again fairly difficult to decide which estimator is the overall best performing according to the Frobenius norm based on the realized covariance proxy as we see different best estimator choices for different forecasting models. Nevertheless, we can observe that the best-performing model-estimator combination from the economic comparison (the MHAR-A model with MRC estimator) performs quite well when we compare it with the remaining estimators using the Frobenius norm. Overall similar results are obtained for forecasts based on the realized kernel estimator (RK) and the modulated realized covariance (MRC), which achieved the second-smallest values of the norm (after the realized covariance, RC).

Regarding the forecasting models, it seems that the best overall results were obtained for the MHAR-A model. In most cases, this model performs the best or very close to the best-performing model in terms of the value of the Frobenius norm. Contrary to economic evaluation, the realized bi-power covariance achieves very competitive results, especially for the MHAR-A and the MHAR-C models.

	POET	*RTS*	*TS*	*RTC*	*MRC*	*CholC*	*BP*	*RK*	*RC*
RW	8.562	6.845	6.070	6.707	5.892	6.390	6.207	14.151	7.551
MHAR-A	8.570	7.618	5.799	5.407	5.154	5.249	5.266	5.090	3.858
MHAR-B	8.582	7.687	7.473	5.921	6.107	6.829	6.313	6.801	4.303
MHAR-C	8.592	7.237	7.149	5.548	6.124	6.520	5.192	6.488	3.958

Table V.3: Table containing the resulting values of the Frobenius norm of the matrix obtained as the difference between the forecast based on given model-estimator combination and the realized covariance used as the proxy for the integrated covariance of the out-of-sample period $t + 1$ (day 98).

Next, we present a comparison based on the Frobenius norm with the realized kernel covariance (RK) selected as the proxy for the integrated covariance, shown in the table V.4.

The realized covariance estimator (RC), in combination with various multivariate HAR models, quite convincingly outperforms the remaining model-estimator combinations. The realized kernel estimator (RK) achieves similar values as the remaining models even though the choice of proxy should be in its favor. Outside these two estimators, we see quite comparable results to the previous statistical evaluation. The $POET$ estimator seems to perform the worst out of the considered estimators.

The MHAR-A model again performs the best out of the selected models across all the different estimators. The MHAR-C model performs almost identically to the MHAR-A model when combined with the realized threshold covariance (RTC) or the realized covariance (RC) estimators. Therefore, our results based on the realized kernel covariance proxy seem consistent with those obtained for the realized covariance as the integrated covariance proxy.

	POET	*RTS*	*TS*	*RTC*	*MRC*	*CholC*	*BP*	*RK*	*RC*
RW	9.521	7.983	7.191	7.992	7.072	7.473	7.899	15.108	9.182
MHAR-A	9.519	8.667	7.053	6.760	6.501	6.528	6.458	6.192	5.115
MHAR-B	9.534	8.645	8.442	7.141	7.150	7.818	7.440	7.823	5.789
MHAR-C	9.538	8.219	8.175	6.713	7.199	7.535	6.438	7.537	5.139

Table V.4: Table containing the resulting values of the Frobenius norm of the matrix obtained as the difference between the forecast based on given model-estimator combination and the realized kernel covariance used as the proxy for the integrated covariance of the out-of-sample period $t + 1$ (day 98).

The Frobenius norm is the only truly formal method to evaluate the quality of forecasts presented in this thesis. Hence, it gives us the best idea about the performance of each model-estimator combination. Based on the results obtained using two different proxies for the integrated covariance in the out-of-sample period, it seems that the best-performing combination is to use the realized covariance estimator to estimate the integrated covariance and then apply the MHAR-A model to forecast the covariance. The choice of MHAR-A could be potentially justified by the fact that this model mostly performs better than the remaining forecasting models and that it is simpler than the model MHAR-C.

On the other hand, all the multivariate HAR models perform better than the RW model in combination with RC, RK, or BP. Interestingly, we observe a very similar trend as in the economic evaluation. The RW model performs relatively better in combination with more complex estimation methods like RTS, TS, or $POET$ compared to the remaining models (especially MHAR-B and MHAR-C). It is also worth noting that in the context of both statistical evaluations (evaluation with both proxies), it seems that the multivariate HAR model is most advantageous in comparison with the RW model in combination with the RC and the RK estimators. For these two estimators, there seems to be a significant gap in the performance of the RW and MHAR models. Therefore, using one of these two estimators in combination with MHAR models as a primary option is quite rational based on the statistical evaluation.

Conclusion

In this thesis, we presented a concise theory behind the integrated covariance of the multivariate price process. Later we introduced several methods that can be used to estimate the integrated covariance. Our first goal was to combine these estimators with different forecasting models that were rigorously described in chapter III. Next, we presented different approaches to the evaluation of resulting forecasts. One of them is the portfolio optimization problem also involving transaction costs. In the empirical study, we applied the described methods to real-life data of 50 companies from S&P500 and analyzed the results by comparing different model-estimator combinations.

In the empirical study, we briefly discussed technical details regarding the acquisition of data and the calculation of forecasts. We also discussed the most challenging part of the implementation, which was tightly connected to the numerical stability of the matrices. Simply put, satisfying the positive semi-definiteness of both forecasts and estimates of the integrated covariance for each day was fairly difficult.

Next, we primarily focused on the comparison of various models and estimators of the integrated covariance. We tried to view them separately by individually analyzing the most persistent properties of selected estimators and models. On the contrary, we also discussed in detail the properties of model-estimator combinations as one entity. We found that more complex estimators seem to work better with the simpler MHAR-A model or even with the random walk assumption-based model (RW). In contrast, simpler estimators like the realized covariance do benefit from the use of the multivariate HAR models and, according to statistical evaluation, perform the best among the considered model-estimator combinations. According to the economic forecast evaluation, few of the more complex estimators, e.g., the modulated realized covariance or the realized Cholesky covariance, perform well in combination with the MHAR-A model. On the contrary, such estimators do not work as great with the remaining multivariate HAR models. The statistical evaluation seems to support such results as the forecasting precision in the context of both proxies of the integrated covariance proves to be better with the MHAR-A model than in combination with other MHAR models. Therefore, there seems to be a noticeable contribution to the forecasting performance of estimators that consider microstructure noises when combined with the MHAR-A or random walk assumption-based models.

Based on our findings, using the multivariate HAR models (especially the MHAR-A and MHAR-C models) combined with simpler estimators like the realized covariance or even the realized kernel estimator seems rational. Nevertheless, if we decide to use more sophisticated approaches regarding the integrated covariance estimation (e.g., the realized Cholesky covariance), it might be beneficial to use a simpler MHAR-A model or even the random walk assumption-based model as they performs more consistently with the more complex estimators.

References

Andersen, T. G., Bollerslev, T., Diebold, F. X., & Ebens, H. (2001). The distribution of realized stock return volatility. *Journal of Financial Economics*, *61*(1), 43-76. doi: 10.1016/S0304-405X(01)00055-1

Bai, J., & Ng, S. (2001). Determining the number of factors in approximate factor models. *Econometrica*, *70*(1), 191–221. doi: 10.1111/1468-0262.00273

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., & Shephard, N. (2011). Multivariate realised kernels: Consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading. *Journal of Econometrics*, *162*(2), 149-169. doi: 10.1016/j.jeconom.2010.07.009

Barndorff-Nielsen, O. E., & Shephard, N. (2003). *Econometrics of testing for jumps in financial economics using bipower variation* (Economics Series Working Papers No. 2004-FE-01). University of Oxford, Department of Economics. doi: 10.1093/jjfinec/nbi022

Bauer, G. H., & Vorkink, K. (2011). Forecasting multivariate realized stock market volatility. *Journal of Econometrics*, *160*(1), 93-101. doi: 10.1016/j.jeconom.2010.03.021

Bollerslev, T., Meddahi, N., & Nyawa, S. (2019). High-dimensional multivariate realized volatility estimation. *Journal of Econometrics*, *212*(1), 116-136. doi: 10.1016/j.jeconom.2019.04.023

Boudt, K., Laurent, S., Lunde, A., Quaedvlieg, R., & Sauri, O. (2017). Positive semidefinite integrated covariance estimation, factorizations and asynchronicity. *Journal of Econometrics*, *196*(2), 347-367. doi: 10.1016/j.jeconom.2016.09.016

Boudt, K., & Zhang, J. (2010). Jump robust two time scale covariance estimation and realized volatility budgets. *Quantitative Finance*, *15*(6), 1041–1054. doi: 10.2139/ssrn.1699730

Chiriac, R., & Voev, V. (2011). Modelling and forecasting multivariate realized volatility. *Journal of Applied Econometrics*, *26*(6), 922-947. doi: 10.1002/jae.1152

Chiu, T. Y. M., Leonard, T., & Tsui, K.-W. (1996). The matrix-logarithmic covariance model. *Journal of the American Statistical Association*, *91*(433), 198–210. doi: 10.1080/01621459.1996.10476677

Christensen, K., Kinnebrock, S., & Podolskij, M. (2010). Pre-averaging estimators of the ex-post covariance matrix in noisy diffusion models with non-synchronous data. *Journal of Econometrics*, *159*(1), 116-133. doi: 10.1016/j.jeconom.2010.05.001

Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, *7*(2), 174-196. doi: 10.1093/jjfinec/nbp001

Čech, F., & Baruník, J. (2017). On the modelling and forecasting of multivariate realized volatility. *Journal of Forecasting*, *36*, 181–206. doi: 10.1002/for.2423

Fan, J., Li, Y., & Yu, K. (2012). Vast volatility matrix estimation using high-frequency data for portfolio selection. *Journal of the American Statistical Association*, *107*(497), 412-428. doi: 10.1080/01621459.2012.656041

Fan, J., Liao, Y., & Mincheva, M. (2013). Large covariance estimation by thresholding principal orthogonal complements. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, *75*(4), 603–680. doi: 10.1111/rssb.12016

Gourieroux, C., Jasiak, J., & Sufana, R. (2009). The wishart autoregressive process of multivariate stochastic volatility. *Journal of Econometrics*, *150*(2), 167-181. doi: 10.1016/j.jeconom.2008.12.016

Harris, F. H. deB., McInish, T. H., Shoesmith, G. L., & Wood, R. A. (1995). Cointegration, error correction, and price discovery on informationally linked security markets. *The Journal of Financial and Quantitative Analysis*, *30*(4), 563-579. doi: 10.2307/2331277

Hautsch, N., & Podolskij, M. (2013). Pre-averaging based estimation of quadratic variation in the presence of noise and jumps: Theory, implementation, and empirical evidence. *Journal of Business and Economic Statistics*, *31*, 165-183. doi: 10.1080/07350015.2012.754313

Hillebrand, E., & Medeiros, M. (2010). The benefits of bagging for forecast models of realized volatility. *Econometric Reviews*, *29*, 571–593. doi: 10.1080/07474938.2010.481554

Hong, W.-T., & Hwang, E. (2022). Exponentially weighted multivariate har model with applications in the stock market. *Entropy*, *24*(7), 937. doi: 10.3390/e24070937

Jacod, J., Li, Y., Mykland, P. A., Podolskij, M., & Vetter, M. (2009). Microstructure noise in the continuous case: The pre-averaging approach. *Stochastic Processes and their Applications*, *119*(7), 2249-2276. doi: 10.1016/j.spa.2008.11.004

Jacod, J., & Todorov, V. (2009). Testing for common arrivals of jumps for discretely observed multidimensional processes. *The Annals of Statistics*, *37*(4), 1792–1838. doi: 10.48550/arXiv.0908.1847

Lunde, A., Shephard, N., & Sheppard, K. (2016). Econometric analysis of vast covariance matrices using composite realized kernels and their application to portfolio choice. *Journal of Business & Economic Statistics*, *34*(4), 504-518. doi: 10.1080/07350015.2015.1064432

Mancini, C., & Gobbi, F. (2012). Identifying the Brownian covariation from the co-jumps given discrete observations. *Econometric Theory*, *28*(2), 249–273. doi: 10.1017/S0266466611000326

Mansini, R., Ogryczak, W., & Speranza, M. G. (2015). Portfolio optimization with transaction costs. In *Linear and Mixed Integer Programming for Portfolio Optimization* (pp. 47–62). Springer International Publishing. doi: 10.1007/978-3-319-18482-1_3

Šípka, S. (2022). *Combining multivariate volatility forecasts in portfolio optimization* (Unpublished master's thesis). Charles University.

Shephard, N., & Barndorff-Nielsen, O. E. (2004). *Measuring the impact of jumps in multivariate price processes using bipower covariation.*

Zhang, L. (2011). Estimating covariation: Epps effect, microstructure noise. *Journal of Econometrics*, *160*(1), 33-47. doi: 10.1016/j.jeconom.2010.03.012

Zhang, L., Mykland, P., & Ait-Sahalia, Y. (2005). A tale of two time scales: Determining integrated volatility with noisy high-frequency data. *Journal of the American Statistical Association*, *100*(472), 1394-1411. doi: 10.2307/27590680

List of Tables

A. Attachments

A.1 Selected Companies

Ticker	Security	Sector	Founded
AAL	American Airlines Group	Industrials	1934
AAPL	Apple Inc.	Information Technology	1977
AES	AES Corporation	Utilities	1981
AMAT	Applied Materials	Information Technology	1967
AMD	AMD	Information Technology	1969
AMZN	Amazon	Consumer Discretionary	1994
AON	Aon	Financials	1982
APTV	Aptiv	Consumer Discretionary	1994
BAC	Bank of America	Financials	1998
BAX	Baxter International	Health Care	1931
BLK	BlackRock	Financials	1988
CLX	Clorox	Consumer Staples	1913
CSCO	Cisco	Information Technology	1984
CZR	Caesars Entertainment	Consumer Discretionary	1973
DPZ	Domino's	Consumer Discretionary	1960
DVA	DaVita Inc.	Health Care	1979
EXR	Extra Space Storage	Real Estate	1977
FRC	First Republic Bank	Financials	1985
GOOGL	Alphabet Inc. (Class A)	Communication Services	1998
GS	Goldman Sachs	Financials	1869
HES	Hess Corporation	Energy	1919
ICE	Intercontinental Exchange	Financials	2000
INTC	Intel	Information Technology	1968
IPG	Interpublic Group of Companies (The)	Communication Services	1930
KIM	Kimco Realty	Real Estate	1958
KO	Coca-Cola Company (The)	Consumer Staples	1886
LUMN	Lumen Technologies	Communication Services	1930
META	Meta Platforms	Communication Services	2004
MHK	Mohawk Industries	Consumer Discretionary	1878
MKC	McCormick & Company	Consumer Staples	1889
MLM	Martin Marietta Materials	Materials	1993
MSFT	Microsoft	Information Technology	1975
NFLX	Netflix	Communication Services	1997
NVDA	Nvidia	Information Technology	1993
PARA	Paramount Global	Communication Services	2019
PNC	PNC Financial Services	Financials	1845
RE	Everest Re	Financials	1973
SBNY	Signature Bank	Financials	2001
SPGI	S&P Global	Financials	1917
STZ	Constellation Brands	Consumer Staples	1945
TSLA	Tesla, Inc.	Consumer Discretionary	2003
VLO	Valero Energy	Energy	1980
VTRS	Viatris	Health Care	1961
VZ	Verizon	Communication Services	2000
WAT	Waters Corporation	Health Care	1958
WHR	Whirlpool Corporation	Consumer Discretionary	1911
WMT	Walmart	Consumer Staples	1962
WTW	Willis Towers Watson	Financials	2016
WY	Weyerhaeuser	Real Estate	1900
ZBH	Zimmer Biomet	Health Care	1927

Table A.1: Table containing individual tickers and core financial information of 50 considered companies.

In table A.1, we present the companies from S&P500 used in the empirical study. Unfortunately, as of the date of writing this thesis, one company from the list is no longer a part of the index. Regulators shut down the Signature Bank (SBNY) around March 10, 2023. The reason for the regulator's intervention was the massive amount of withdrawals from the bank accounts due to another bank's failure (the Silicon Valley Bank).

A.2 Portfolio Optimization Results

	POET	*RTS*	*TS*	*RTC*	*MRC*	*CholC*	*BP*	*RK*	*RC*
$s = 1$									
RW	0.18	11.73	25.11	39.68	35.34	15.55	11.81	103.56	57.00
MHAR-A	1.08	5.64	10.78	74.74	44.76	19.49	49.34	21.76	156.13
MHAR-B	0.54	26.51	40.67	66.95	66.59	53.33	46.65	33.55	84.42
MHAR-C	0.97	28.11	23.82	85.10	43.71	49.70	55.90	18.26	151.64
$s = 1.5$									
RW	0.18	11.73	25.11	39.68	35.34	15.55	11.81	103.56	57.00
MHAR-A	1.08	5.64	10.78	74.74	31.21	22.03	49.34	21.76	156.13
MHAR-B	0.54	26.51	40.67	66.95	66.59	53.33	46.65	41.01	84.42
MHAR-C	0.97	27.69	23.82	85.10	43.71	49.70	55.90	18.26	151.64
$s = 2$									
RW	0.18	11.73	25.11	39.68	35.34	15.55	11.81	103.56	57.00
MHAR-A	1.08	5.64	10.78	74.74	31.21	22.03	49.34	21.76	156.13
MHAR-B	0.54	26.51	40.67	66.95	66.59	53.33	46.65	41.01	84.42
MHAR-C	0.97	27.69	23.82	85.10	43.71	49.70	55.90	18.26	151.64
$s = 2.5$									
RW	0.18	11.73	25.11	39.68	35.34	15.55	11.81	103.56	57.00
MHAR-A	1.08	5.64	10.78	74.74	31.21	22.03	49.34	21.76	156.13
MHAR-B	0.54	26.51	40.67	66.95	66.59	53.33	46.65	41.01	84.42
MHAR-C	0.97	27.69	23.82	85.10	43.71	49.70	55.90	18.26	151.64
$s = 3$									
RW	0.18	11.73	25.11	39.68	35.34	15.55	11.81	103.56	57.00
MHAR-A	1.08	5.64	10.78	74.74	31.21	22.03	49.34	21.76	156.13
MHAR-B	0.54	26.51	40.67	66.95	66.59	53.33	46.65	41.01	84.42
MHAR-C	0.97	27.69	23.82	85.10	43.71	49.70	55.90	18.26	151.64
$s = \infty$									
RW	0.17	11.73	25.11	38.14	35.34	18.61	11.81	91.39	56.97
MHAR-A	1.08	4.79	10.78	74.71	31.21	22.03	49.34	21.76	156.13
MHAR-B	0.57	26.54	40.67	66.17	66.59	49.08	46.65	41.01	84.42
MHAR-C	1.00	27.69	23.82	85.15	43.71	49.70	55.90	18.26	151.33

Table A.2: The attained minima in the portfolio optimization problem described in the chapter IV for six different choices of gross exposure parameter s $(1, 1.5, 2, 2.5, 3, \infty)$. The choice of parameter s equal to ∞ correspond to classical Markovitz portfolio. All the presented results are displayed on the 10^7 scale.

	EqW	*POET*	*RTS*	*TS*	*RTC*	*MRC*	*CholC*	*BP*	*RK*	*RC*
					$s = 1$					
RW	139.84	24.10	22.07	21.94	20.65	20.36	19.67	25.74	22.73	24.45
MHAR-A	139.84	17.56	28.50	14.33	18.52	17.27	14.34	21.62	13.32	20.62
MHAR-B	139.84	19.30	18.88	21.99	18.51	18.70	19.42	20.13	17.57	18.00
MHAR-C	139.84	21.12	21.80	16.53	20.20	19.09	19.57	22.57	13.73	18.80
					$s = 1.5$					
RW	139.84	24.10	22.07	21.94	20.65	20.36	19.67	25.74	22.73	24.45
MHAR-A	139.84	17.56	28.50	14.33	18.52	12.94	14.21	21.62	13.32	20.62
MHAR-B	139.84	19.30	18.88	21.99	18.51	18.70	19.42	20.13	18.34	18.00
MHAR-C	139.84	21.12	20.88	16.53	20.20	19.09	19.57	22.57	13.73	18.80
					$s = 2$					
RW	139.84	24.10	22.07	21.94	20.65	20.36	19.67	25.74	22.73	24.45
MHAR-A	139.84	17.56	28.50	14.33	18.52	12.94	14.21	21.62	13.32	20.62
MHAR-B	139.84	19.30	18.88	21.99	18.51	18.70	19.42	20.13	18.34	18.00
MHAR-C	139.84	21.12	20.88	16.53	20.20	19.09	19.57	22.57	13.73	18.80
					$s = 2.5$					
RW	139.84	24.10	22.07	21.94	20.65	20.36	19.67	25.74	22.73	24.45
MHAR-A	139.84	17.56	28.50	14.33	18.52	12.94	14.21	21.62	13.32	20.62
MHAR-B	139.84	19.30	18.88	21.99	18.51	18.70	19.42	20.13	18.34	18.00
MHAR-C	139.84	21.12	20.88	16.53	20.20	19.09	19.57	22.57	13.73	18.80
					$s = 3$					
RW	139.84	24.10	22.07	21.94	20.65	20.36	19.67	25.74	22.73	24.45
MHAR-A	139.84	17.56	28.50	14.33	18.52	12.94	14.21	21.62	13.32	20.62
MHAR-B	139.84	19.30	18.88	21.99	18.51	18.70	19.42	20.13	18.34	18.00
MHAR-C	139.84	21.12	20.88	16.53	20.20	19.09	19.57	22.57	13.73	18.80
					$s = \infty$					
RW	139.84	22.19	22.07	21.94	20.83	20.36	19.58	25.74	21.19	24.44
MHAR-A	139.84	17.56	27.54	14.33	18.51	12.94	14.21	21.62	13.32	20.62
MHAR-B	139.84	20.44	18.89	21.99	18.13	18.70	17.09	20.13	18.34	18.00
MHAR-C	139.84	20.44	20.88	16.53	20.22	19.09	19.57	22.57	13.73	18.77

Table A.3: Table contains the evaluation of model-estimator combinations based on the calculation of $\widehat{\boldsymbol{w}}_{t+1}^T RC_{t+1} \widehat{\boldsymbol{w}}_{t+1}$, where RC_{t+1} is the realized covariance for the period $t+1$ and $\widehat{\boldsymbol{w}}_{t+1}$ are the optimal weights obtained as a result of the optimization problem described in the chapter IV (the portfolio weights corresponding to the attained minima presented in the table A.2). We show the performance based on this evaluation method for all six selected choices of gross exposure parameter s $(1, 1.5, 2, 2.5, 3, \infty)$. The *EqW* column corresponds to the portfolio with equal weights. All the presented results are displayed on the 10^6 scale.

A.3 Properties of *expm* and *logm* Functions

We will follow Chiu et al. (1996) and show some interesting and important properties of both *logm* and *expm* functions. Let us have a $N \times N$ dimensional matrix $\boldsymbol{A}$. The matrix exponential function is defined as:

$$\boldsymbol{B} = expm(\boldsymbol{A}) = \sum_{s=0}^{\infty} \frac{\boldsymbol{A}^s}{s!},$$

where $\boldsymbol{A}^0$ reduces to the $N \times N$ identity matrix and $\boldsymbol{A}^s$ denotes ordinary matrix multiplication of $\boldsymbol{A}$ (s times). Therefore, the elements of $\boldsymbol{B}$ do not typically coincide with exponentiated elements of individual elements of $\boldsymbol{A}$.

Let us also assume that $\boldsymbol{B}$ is an $N \times N$ positive definite matrix with spectral decomposition $\boldsymbol{B} = \boldsymbol{T}\boldsymbol{D}\boldsymbol{T}^T$, where $\boldsymbol{D}$ is an $N \times N$ dimensional diagonal matrix with eigenvalues of $\boldsymbol{B}$ on the diagonal and the columns of the $N \times N$ orthogonal matrix $\boldsymbol{T}$ consist of the corresponding eigenvectors. We can now define the matrix logarithm of $\boldsymbol{B}$ as:

$$\boldsymbol{A} = logm(\boldsymbol{B}) = \boldsymbol{T} logm(\boldsymbol{D})\boldsymbol{T}^T,$$

where $logm(\boldsymbol{D})$ is again an $N \times N$ dimensional diagonal matrix with logarithms of eigenvalues on the diagonal. The matrix logarithm is an inverse function to the matrix exponential function.

Both transformations have some important properties that are worth mentioning. We present several of these properties and observations under the assumption of positive definiteness of $\boldsymbol{B}$.

1. Let $\boldsymbol{W}$ be an $N \times N$ orthogonal matrix. Then based on the spectral decomposition of $\boldsymbol{B}$ it holds:

   $$logm(\boldsymbol{W}\boldsymbol{B}\boldsymbol{W}^T) = \boldsymbol{W}\boldsymbol{A}\boldsymbol{W}^T = \boldsymbol{W} logm(\boldsymbol{B})\boldsymbol{W}^T$$

2. The determinant $|B|$ of the matrix $\boldsymbol{B}$ satisfies:

   $$log(|\boldsymbol{B}|) = tr(\boldsymbol{A}),$$

   where $tr(\boldsymbol{A})$ is a trace of the matrix $\boldsymbol{A}$ (which is equal to the sum of the diagonal elements of $\boldsymbol{A}$).

3. The inverse $\boldsymbol{B}^{-1}$ satisfies:

   $$\boldsymbol{B}^{-1} = expm(-\boldsymbol{A}).$$

4. On the other hand, it does not generally hold that

   $$\boldsymbol{B}_p = expm(\boldsymbol{A}_p),$$

   where $\boldsymbol{B}_p$ is some $p \times p$ ($p < N$) positive semi-definite submatrix of $\boldsymbol{B}$ consisting of p rows and corresponding p columns and $\boldsymbol{A}_p$ is the corresponding submatrix of $\boldsymbol{A}$.

5. It is also generally not true that the diagonal elements of $\boldsymbol{A}$ correspond to the logarithms of eigenvalues of $\boldsymbol{B}$.

6. Let us assume that $\boldsymbol{B}$ is a linear combination of an $N \times N$ identity matrix $\boldsymbol{I}_N$ and an $N \times N$ idempotent matrix $\boldsymbol{G}$, $\boldsymbol{B} = a\boldsymbol{I}_N + b\boldsymbol{G}$, where a and b are positive constants. Then
$$\boldsymbol{A} = logm(\boldsymbol{B}) = \alpha \boldsymbol{I}_N + \beta \boldsymbol{G},$$
where $\alpha = log(a)$ and $\beta = log(a+b) - log(a)$. This relation also holds for submatrices $\boldsymbol{A}_p, \boldsymbol{B}_p, \boldsymbol{G}_p, \boldsymbol{I}_p$, $p \leq N$.

7. In general,
$$expm(\boldsymbol{A})expm(\boldsymbol{B}) \neq expm(\boldsymbol{A}+\boldsymbol{B}).$$
Additionally, it holds
$$tr(expm(\boldsymbol{A})expm(\boldsymbol{B})) \geq tr(expm(\boldsymbol{A}+\boldsymbol{B})),$$
Equality in both cases holds if and only if $\boldsymbol{A}$ and $\boldsymbol{B}$ are commutative matrices ($\boldsymbol{AB} = \boldsymbol{BA}$).

8. Finally, we will formulate a significant relation between matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ as a lemma. This property is of great importance in the context of the multivariate HAR models.

   Lemma 1. *For any real positive definite $N \times N$ dimensional matrix $\boldsymbol{B}$, there exists a real symmetric square matrix $\boldsymbol{A}$ of the same dimensions as $\boldsymbol{B}$ such that:*
$$\boldsymbol{B} = expm(\boldsymbol{A}).$$
   Reversely, for any real symmetric matrix $\boldsymbol{A}$, the matrix $\boldsymbol{B} = expm(\boldsymbol{A})$ is positive definite.

   Proof. We have already described the spectral decomposition of $\boldsymbol{B}$,
$$\boldsymbol{B} = \boldsymbol{TDT}^T,$$
   where $\boldsymbol{D}$ is a diagonal matrix of eigenvalues of $\boldsymbol{B}$ and $\boldsymbol{T}$ is an orthogonal matrix of corresponding normalized eigenvectors in columns. Let $\boldsymbol{A} = \boldsymbol{T}logm(\boldsymbol{D})\boldsymbol{T}^T$, then:
$$expm(\boldsymbol{A}) = expm(\boldsymbol{T}logm(\boldsymbol{D})\boldsymbol{T}^T) = \boldsymbol{TDT}^T = \boldsymbol{B}.$$
   On the contrary, let us assume that $\boldsymbol{A}$ is a real symmetric matrix, and then there exists an orthogonal matrix $\boldsymbol{T}$ such that $\boldsymbol{A} = \boldsymbol{TDT}^T$, where $\boldsymbol{D}$ is a diagonal matrix. Subsequently, it holds:
$$\begin{aligned} expm(\boldsymbol{A}) &= expm(\boldsymbol{TDT}^T) = \boldsymbol{T}expm(\boldsymbol{D})\boldsymbol{T}^T \\ &= \boldsymbol{T}diag(exp(\lambda_1), \ldots, exp(\lambda_N))\boldsymbol{T}^T. \end{aligned}$$
   For any vector $\boldsymbol{x} \neq \boldsymbol{0}$ we construct the quadratic form:
$$\boldsymbol{x}^T expm(\boldsymbol{A})\boldsymbol{x} = (\boldsymbol{x}^T\boldsymbol{T})diag(exp(\lambda_1), \ldots, exp(\lambda_N))(\boldsymbol{x}^T\boldsymbol{T})^T.$$
   The element $exp(\lambda_i)$ of the diagonal matrix in the quadratic form is always positive for any i and, therefore, the matrix $\boldsymbol{B} = expm(\boldsymbol{A})$ is positive definite. □