Modeling the covariance matrix of financial asset returns

Gustav Alfelt
Modeling the covariance matrix of financial asset returns

Gustav Alfelt

Academic dissertation for the Degree of Doctor of Philosophy in Mathematical Statistics at Stockholm University to be publicly defended on Thursday 20 May 2021 at 13.00 online via Zoom, public link is available at the department website.

Abstract
The covariance matrix of asset returns, which describes the fluctuation of asset prices, plays a crucial role in understanding and predicting financial markets and economic systems. In recent years, the concept of realized covariance measures has become a popular way to accurately estimate return covariance matrices using high-frequency data. This thesis contains five research papers that study time series of realized covariance matrices, estimators for related random matrix distributions, and cases where the sample size is smaller than the number of assets considered.

Paper I provides several goodness-of-fit tests for discrete realized covariance matrix time series models that are driven by an underlying Wishart process. The test methodology is based on an extended version of Bartlett's decomposition, allowing to obtain independent and standard normally distributed random variables under the null hypothesis. The paper includes a simulation study that investigates the tests' performance under parameter uncertainty, as well as an empirical application of the popular conditional autoregressive Wishart model fitted to data on six stocks traded over eight and a half years.

Paper II derives the Stein-Haff identity for exponential random matrix distributions, a class which for example contains the Wishart distribution. It furthermore applies the derived identity to the matrix-variate gamma distribution, providing an estimator that dominates the maximum likelihood estimator in terms of Stein's loss function. Finally, the theoretical results are supported by a simulation study.

Paper III supplies a novel closed-form estimator for the parameters of the matrix-variate gamma distribution. The estimator appears to have several benefits over the typically applied maximum likelihood estimator, as revealed in a simulation study. Applying the proposed estimator as a start value for the numerical optimization procedure required to find the maximum likelihood estimate is also shown to reduce computation time drastically, when compared to applying arbitrary start values.

Paper IV introduces a new model for discrete time series of realized covariance matrices that obtain as singular. This case occurs when the matrix dimension is larger than the number of high frequency returns available for each trading day. As the model naturally appears when a large number of assets are considered, the paper also focuses on maintaining estimation feasibility in high dimensions. The model is fitted to 20 years of high frequency data on 50 stocks, and is evaluated by out-of-sample forecast accuracy, where it outperforms the typically considered GARCH model with high statistical significance.

Paper V is concerned with estimation of the tangency portfolio vector in the case where the number of assets is larger than the available sample size. The estimator contains the Moore-Penrose inverse of a Wishart distributed matrix, an object for which the mean and dispersion matrix are yet to be derived. Although no exact results exist, the paper extends the knowledge of statistical properties in portfolio theory by providing bounds and approximations for the moments of this estimator as well as exact results in special cases. Finally, the properties of the bounds and approximations are investigated through simulations.

Keywords: Realized covariance, Autoregressive time-series, Goodness-of-fit test, Matrix singularity, Portfolio theory, Wishart distribution, Matrix-variate gamma distribution, Parameter estimation, High-dimensional data, Moore-Penrose inverse.

Stockholm 2021
http://urn.kb.se/resolve?urn=urn:nbn:se:su:diva-191175


Department of Mathematics
Stockholm University, 106 91 Stockholm
MODELING THE COVARIANCE MATRIX OF FINANCIAL ASSET RETURNS

Gustav Alfelt
Modeling the covariance matrix of financial asset returns

Gustav Alfelt
To my family.
List of Papers

This thesis is based on the following papers, which are referred to in the text by their Roman numerals.

I: Goodness-of-fit tests for centralized Wishart processes.

II: Stein-Haff identity for the exponential family.

III: Closed-form estimator for the matrix-variate gamma distribution.

IV: Singular conditional autoregressive Wishart model for realized covariance matrices.

V: On the mean and variance of the estimated tangency portfolio weights for small samples.

Reprints were made with permission from the publishers.

Author's contributions: G. Alfelt has taken an active part in developing the content of all papers, including outlining the manuscripts, formulating and proving the theoretical results, writing and revising the manuscripts, as well as implementing the computer simulations and the empirical applications. Paper I was based on an idea of T. Bodnar and J. Tyrcha, where G. Alfelt formulated and implemented the simulation study and

---

1© 2019 The Author(s). Published with license by Taylor & Francis Group, LLC.
empirical part, and wrote the majority of the manuscript with assistance of T. Bodnar and J. Tyrcha. G. Alfelt is the sole author of Paper II and Paper III. The original idea of Paper IV was proposed by T. Bodnar and F. Javed. G. Alfelt proposed and formulated the model and its several extensions, as well as implementing the empirical part, and wrote the majority of the manuscript. Finally, Paper V is based on an idea of S. Mazur, where G. Alfelt formulated and proved the theoretical results, carried out the simulations and provided the majority of the writing.

**General comment:** An earlier version of Paper I, Paper II, and parts of the introduction were contained in the Licentiate thesis of Gustav Alfelt, Alfelt (2019a).
Acknowledgments

Pursuing a Ph.D. degree has been a fascinating journey, which I’ve had the pleasure to spend the last few years on. It has given me the opportunity to dig deep into a subject I’ve always held dear while exploring the frontier of modern research, but it has also introduced me to wonderful people and places. This journey would not have been possible without a number of individuals, to which I here would like to express my deep gratitude.

First of all, I would like to thank my supervisors, Joanna Tyrcha and Taras Bodnar. For entrusting me the Ph.D. student mantle, for supporting and guiding me in all aspects of my work, and for generously sharing your knowledge.

I want to thank Farrukh Javed and Stepan Mazur for the joint work and ideas, together with many rewarding discussions and brainstorming sessions, that lead to Paper IV and Paper V of this thesis.

Further, I want to thank my colleagues at the Department of Mathematics at Stockholm University, in particular all the other Ph.D. students, for many interesting discussions and for making the institution a great place to work at. A special thank you to Erik, Stanislas and Vilhelm, for being great office and travel mates throughout the years, and for all the laughs.

Thank you Marieke and Kasper, for introducing me to academic work life, providing invaluable guidance and for inspiring me to pursue a Ph.D degree.

Niklas and Sophia, thank you for being there both in good times and in bad, and for always having my back.

Finally, I want to thank the rocks of my life. For your endless love, for you unwavering support and for always believing in me. Thank you, mom and dad.
Contents

List of Papers i

Acknowledgments v

I Introduction 3

1 Covariation of asset returns 5

2 Covariance matrix 9
   2.1 Definition and basic properties 9
   2.2 Eigenvalues and eigenvectors 11
   2.3 Estimators of the covariance matrix 15
   2.4 Singularity and the Moore-Penrose inverse 17
   2.5 Wishart distribution 20

3 Integrated and realized covariance 23

4 Time series of realized covariance 26

5 Portfolio theory 32

6 Summary of papers 37

Sammanfattning 44

References 51

II Papers 53
Part I

Introduction
Chapter 1

Covariation of asset returns

Current prices of assets - be it food, raw material, housing or bank loans - can tell a revealing story about the current state of the world, and expectations of the times ahead. In prosperous times, the future might seem bright, encouraging investments in upstarting companies with hopes of high investment returns, potentially inflating the prices of such assets. On the other hand, in poorer times the outlook often seems more grim, perhaps leading investors to move their capital from risky endeavours to more stable assets such as gold and government bonds, again shifting market prices. In comprehensive financial crisis, such as the one in 2008, this may become direly evident, as prices of certain assets often drop rapidly.

Consequently, modeling price fluctuation remains a crucial part in both understanding the economic systems our society consists of, as well as assessing financial risks and identifying investment opportunities. One quantity central to asset price dynamics is the return of the asset between two time periods. It is commonly defined as the logarithm of the asset price at the later time point minus the logarithm of the price at the earlier time point, hence giving a measure of the relative price change over the time interval. In effect, it determines the proportional profit an agent receives investing in the asset. As future prices in general are always unknown, so are the returns between now and some future time point, or between two future time points. Hence, this quantity is essentially always modeled as a random variable. The properties for a set of assets’ return distribution is a central input parameter in most of financial applications, and much research is devoted to modeling them.

The most prominent features of an asset’s return distribution are arguably its first two moments, specifying the expected return and the variance of the asset. The former determines what profit an agent can expect from an investment, while the latter determines the dispersion of the random return, and is often used as a general measure of the riskiness
involved with investing in the asset. Sometimes the square root of the return variance is
used, generally noted asset volatility. When considering more than one asset, both the
individual variances, but also the covariances, determining how the asset returns fluctuate
in relation to each other, are highly important. These quantities are often structured into
a covariance matrix, an object which on its own provides essential information regarding
fluctuation of the returns for a set of asset. The covariance matrix is a key parameter
in for example option pricing theory, and fundamental for various financial regulatory
frameworks, such as regulatory capital requirement based on value-at-risk measures. Be-
ing a central quantity both in pricing financial instruments, as well as in understanding
the structural behaviour of our financial system and the risk it inherits, I have chosen to
dedicate my Ph.D. studies to research on the covariance matrix of asset returns, which
hence is the focus of this doctoral thesis.

Analyzing historical data on asset returns rather clearly suggests that conditional
return covariance matrices are unlikely to be constant over longer time periods, at least
regarding the one day return frequency. Concerning longer time intervals this might
seem intuitive - periods with financial turmoil which sharp price drops suggest large
return variances, while calm periods with steady economic growth often exhibit lower
return variance, for example. But similar changes in return variance seem to appear
also for shorter time periods, with rapid shifts over weeks or days. Since investment re-
balancing and trading strategy updates are often conducted on daily basis, one is regularly
interested in covariance models that adapt to the latest available data on, at least, daily
frequency. Hence, time-series models for one day return covariance matrices, that are
able to accurately capture the fluctuations and dynamics of these quantities, has become
a large research area. One prominent class of such models are the multivariate generalized
autoregressive heteroskedasticity (MGARCH) models, first introduced in Bollerslev et al.
(1988). This model type assumes that the vector of considered daily asset returns has a
latent covariance matrix that is re-specified for each trading day. This latent quantity is
updated incorporating the covariance matrix of previous days, as well as data on the return
vector of previous days. Hence, the model can potentially capture long-term trends of
the covariance matrix, as well as adapting to rapid spikes or drops in recent observations,
also incorporating short-term fluctuations. A related class are the multivariate stochastic volatility (MSV) models, where the latent process of covariance matrices instead are assumed to be random. Great summaries for these classical model types are provided by Bauwens et al. (2006) for the MGARCH models and by Asai et al. (2006) regarding the MSV models.

The assumption of a conditional daily asset return covariance matrix that varies from trading day to trading day, as assumed in e.g. the MGARCH-type models, does however pose a statistical challenge. When the daily asset returns are not considered identically and independently distributed, the statistician essentially has to estimate the covariance matrix of a particular trading day based on a single observation of the one day return from that trading day. Such a procedure naturally generates very imprecise estimates. However, during the last decades, increased availability of asset prices recorded on very high frequency has presented new possibilities in this area. Instead of considering the return computed from the closing prices between two consecutive trading days, novel methods rely on the numerous price variations that occur throughout the trading day. Various matrices aiming to estimate the one day asset return covariance matrix with such approaches are typically denoted realized measures, or realized covariance measures. As these facilitate collecting much larger samples sizes, they allow obtaining much less noisy estimates.

The techniques of realized measures has spurred a new area of research, analyzing how to refine and model these quantities. This area is where the research conducted in this thesis emanates from. The first research paper of this thesis supplies several goodness-of-fit tests adapted to models of discrete realized covariance matrix time series, supplying methods to evaluate how well such models can describe particular sets of collected data. In paper two and three, results and estimation methods for distributions suitable to model realized covariance matrices are derived. The fourth research paper introduces a model for discrete time series of realized covariance matrices computed when the number of assets out-weight the amount of high-quality intra-day return data available. The fifth and final research paper of this thesis also considers the situation of sample size smaller than the number of assets considered. While the first four papers is concerned with modeling the
asset return covariance matrix, this paper applies the covariance matrix in the portfolio theory setting, a framework which aims to derive optimal ways to allocate capital between a set of considered assets. In the paper, several properties for estimators of such allocation quantities are derived, extending recently published results in the research area.

The rest of the introduction part is organized as follows. Chapter 2 provides a primer on the covariance matrix and its properties, including its definition, how it relates to eigenvalue and eigenvectors, estimators for the covariance matrix, as well as singularity and the Wishart distribution, which often appears in junction with covariance matrix estimators. In Chapter 3, realized covariance is introduced together with its theoretical counterpart, integrated covariance. Discrete time series of realized covariance matrices is discussed in Chapter 4, together with a review of existing models to describe the dynamics of such series. In Chapter 5, portfolio theory is introduced, together with a few common allocation strategies, and how these are applied in the papers of this thesis. Finally, Chapter 6 provides a summary of the five research papers that this thesis consists of. Thereafter follows part two of this thesis, which contains each of the five papers in their full length.
Chapter 2

Covariance matrix

This chapter discusses the covariance matrix, the most common quantity used to describe the dispersion of a random vector, and presents some of its typical features. The aim is to provide a primer on the key concepts that are discussed in the rest of the thesis. In Section 2.1, the definition together with basic properties are presented. Eigenvalues and their role with regard to covariance matrices are discussed in Section 2.2. Section 2.3 presents estimators of the covariance matrix, while singularity of covariance matrices are reviewed in Section 2.4. Finally, the Wishart distribution, with its properties and various applications, are presented in Section 2.5. All matrices in this chapter are assumed to be real-valued.

Excellent walkthroughs on the covariance matrix, the statistical properties of its estimators, the Wishart distribution and related laws, together with general matrix algebra can be found in e.g. Muirhead (1982), Harville (1997), Gupta and Nagar (2000), Anderson (2003) and Kollo and von Rosen (2006).

2.1 Definition and basic properties

The covariance matrix of a $p \times 1$ random vector $\mathbf{x}$ is a symmetric $p \times p$ matrix defined as

$$\mathbf{V}[\mathbf{x}] = \mathbb{E}[(\mathbf{x} - \mathbb{E}[\mathbf{x}])(\mathbf{x} - \mathbb{E}[\mathbf{x}])']$$

extending the notion of variance and covariance to the general vector case. Here $\mathbb{E}[\cdot]$ denotes the expectation operator and $\mathbf{A}'$ denotes the transpose of the matrix $\mathbf{A}$. Let $\Sigma$ denote the covariance matrix of $\mathbf{x}$, and denote the element on row $i$ and column $j$ of $\Sigma$ as $\sigma_{ij}$, $i, j = 1, \ldots, p$. As such, if $x_i$ is the the $i$:th element of $\mathbf{x}$, we have that $\sigma_{ii}$ denotes the variance of $x_i$, while $\sigma_{ij}$ denotes the covariance between $x_i$ and $x_j$, $i \neq j$. In the case
of $p = 3$, $\Sigma$ will thus have the following symmetric structure:

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_{22} & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_{33} \end{pmatrix},$$

where the diagonal elements of $\Sigma$, $\sigma_{11}$, $\sigma_{22}$ and $\sigma_{33}$, represents the variances of $x_1$, $x_2$ and $x_3$, respectively, while the non-diagonal elements, $\sigma_{12}$, $\sigma_{13}$ and $\sigma_{23}$ represents the covariances between these random variables.

Moreover, as the variance of a univariate random variable is non-negative, the covariance matrix $\Sigma$ is correspondingly *positive semi-definite* (p.s.d.), which we denote $\Sigma \geq 0$. A square symmetric $p \times p$ matrix $A$ is said to be positive semi-definite if and only if, for all non-zero vectors $\alpha \in \mathbb{R}^p$, it holds that $\alpha' A \alpha \geq 0$. If the inequality is strict, then the matrix $A$ is instead said to be *positive definite* (p.d.), which we denote $A > 0$. The difference between a p.s.d. and p.d. covariance matrix will be discussed further in Section 2.4.

So, in the context of covariance matrices, what does the positive semi-definite property entail? First, the property ensures that each of the diagonal elements of $\mathbb{V}[x]$ are non-negative, in correspondence to the non-negativity of the variance of a univariate random variable. Regarding the effects on the non-diagonal elements, let us look at an example. Let the covariance matrix of the $3 \times 1$ random vector $x$ be

$$\Sigma = \begin{pmatrix} 1 & 0.9 & \sigma_{13} \\ 0.9 & 1 & 0.9 \\ \sigma_{13} & 0.9 & 1 \end{pmatrix}. \quad (2.1)$$

The structure of $\Sigma$ tells us that the variance of each element in $x$ is 1, while the covariance values of 0.9 suggest that the dependency between $x_1$ and $x_2$, as well as between $x_2$ and $x_3$, is positive, and quite large. Now, let us consider what values $\sigma_{13}$, the covariance between $x_1$ and $x_3$, can take. From basic probability theory, we know that $|\text{Cov}[x_1, x_3]| \leq \sqrt{\mathbb{V}[x_1] \mathbb{V}[x_3]}$, such that $\sigma_{13} \in (-1, 1)$, in our example. Would for example $\sigma_{13} = -0.9$ be possible then? Let $\alpha = [1, -1, 1]$. With $\sigma_{13} = -0.9$, we have that $\alpha' \Sigma \alpha = -2.4$, and
hence $\Sigma$ is not p.s.d., and therefore not a valid covariance matrix. This seems intuitive: If $x_1$ and $x_2$ are highly positively dependent, and $x_2$ and $x_3$ are highly positively dependent, $x_1$ and $x_3$ can not to be highly negatively dependent. Straightforward calculations and application of the determinant property (ii) presented below shows that for $\Sigma$ to be p.s.d., we must have that $\sigma_{13} \in [0, 62, 1]$, such that also $x_1$ and $x_3$ have a high degree of positive dependence. Hence, a heuristic interpretation of the p.s.d. property of covariance matrices is that studying the pairwise covariances independently is not enough, all the dependencies of the random vector’s elements must be considered jointly, and the dependencies must make sense structurally.

From the p.s.d. property, a number of other properties follow, where the most basic of them are listed below. Here, we denote the $p$ ordered eigenvalues of $\Sigma$ as $\lambda_1, \lambda_2, \ldots, \lambda_p$, while $|\cdot|$ denotes the determinant operator. We also assume that the matrices are of dimensions such that the following additions and multiplications are possible. Given that $\Sigma \geq 0$, the following holds:

(i) $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p \geq 0$. If $\Sigma > 0$, the last inequality is strict.

(ii) $|\Sigma| \geq 0$. If $\Sigma > 0$, the inequality is strict.

(iii) If $\Sigma > 0$, then $\Sigma^{-1} > 0$.

(iv) If $c > 0$, then $c\Sigma \geq 0$. If $\Sigma > 0$, the inequality is strict.

(v) If $A \geq 0$, then $\Sigma + A \geq 0$. If $\Sigma > 0$, the inequality is strict.

(vi) For any matrix $A$, we have that $A^\prime \Sigma A \geq 0$. If $\Sigma > 0$ and $A$ is of full column rank, we have that $A^\prime \Sigma A > 0$.

The above properties, especially in the case of $\Sigma > 0$, will be extensively applied throughout the papers included in this thesis.

### 2.2 Eigenvalues and eigenvectors

The eigenvalues and eigenvectors of a covariance matrix provide important information of the dependency structure of the associated random vector. A $p \times p$ covariance matrix
that is p.d. can be represented by the eigendecomposition

\[ \Sigma = \Gamma \Lambda \Gamma', \]  

(2.2)

where the \( p \) normalized eigenvectors of \( \Sigma \) are stacked as columns in \( \Gamma \), while the elements of the diagonal matrix \( \Lambda \) consists of the \( p \) positive eigenvalues of \( \Sigma \). We have that \( \Gamma \) is an orthogonal matrix, an object that is characterized by the property \( \Gamma \Gamma' = \Gamma' \Gamma = I_p \). It should be noted that the decomposition (2.2) is not unique. While the set of eigenvalues of \( \Sigma \) is unique, their associated eigenvectors are not, and consequently \( \Gamma \) in (2.2) can be represented by a number of orthogonal matrices. Furthermore, let \( \Lambda^{1/2} \) be a diagonal matrix where the element on row \( i \) and column \( j \) is the positive square root of the element on row \( i \) and column \( j \) in \( \Lambda \). But, what does \( \Gamma \) and \( \Lambda \) tell us about the dispersion patterns of the random vector? Let us illustrate with an example.

Suppose that \( x \) is a \( 2 \times 1 \) multivariate normally distributed random vector with mean zero and covariance matrix equal to the identity matrix, which we denote \( x \sim N_2(0, I_2) \). The top left graph in Figure 2.1 displays 10 000 samples of \( x \), where the points are concentrated in a circle around the origin. Now, let

\[ \Lambda = \begin{pmatrix} 4 & 0 \\ 0 & 1/4 \end{pmatrix}, \]  

(2.3)

and define \( y = \Lambda^{1/2}x \), such that \( y \sim N_2(0, \Lambda) \), since \( \mathbb{V}[Ax] = \Lambda \mathbb{V}[x] \Lambda' \) for a random vector \( x \) and a constant matrix \( \Lambda \). Thus, whereas the elements \( x_1 \) and \( x_2 \) had variance 1, the scaling by \( \Lambda \) results in \( \mathbb{V}[y_1] = 4 \), \( \mathbb{V}[y_2] = 1/4 \) and \( \text{Cov}[y_1, y_2] = 0 \). Based on the previously drawn samples of \( x \), the top right graph of Figure 2.1 displays the corresponding transformations \( y \). Noticeably, the observations of \( y_1 \) are spread wider than those of \( x_1 \), while the spread is smaller for \( y_2 \), than it is for \( x_2 \). Similarly, no correlation between the draws of \( y_1 \) and \( y_2 \) seems discernible.

Next, consider the orthogonal matrix

\[ \Gamma = \begin{pmatrix} \cos(45^\circ) & -\sin(45^\circ) \\ \sin(45^\circ) & \cos(45^\circ) \end{pmatrix}. \]  

(2.4)
From basic results in linear algebra, pre-multiplying a vector with $\Gamma$ rotates the vector $45^\circ$ counter-clockwise, while retaining the vector’s length. Now, define $z = \Gamma y = \Gamma \Lambda^{1/2} x$, such that $z \sim N_2(0_2, \Sigma)$, where

$$
\Sigma = \Gamma \Lambda \Gamma' = \begin{pmatrix}
2.125 & 1.875 \\
1.875 & 2.125
\end{pmatrix}.
$$

The bottom left graph of Figure 2.1 displays the transformations $z$ based on the previously drawn samples of $x$. As expected, it consists of the cloud of observations $y$ in the top right graph, but with a $45^\circ$ counter-clockwise rotation. It is noticeable that the dispersion is the same as in the top right graph, except that it now occurs along the line $z_1 = z_2$. Further, while $y_1$ and $y_2$ had different variances but zero covariance, we now have $V[z_1] = V[z_2] = 2.125$ and $\text{Cov}[z_1, z_2] = 1.875$.

Finally we set $m = (2, -3)'$ and $w = m + z = m + \Gamma \Lambda^{1/2} x$, such that $w \sim N_2(m, \Sigma)$ consists of a shift of $z$ by $m$. The bottom right graph of Figure 2.1 displays the observations of $w$ based on the sample of $x$. As expected the observations resembles those of $z$, but with the center shifted by $+2$ along the horizontal axis and $-3$ along the vertical axis. Applying general values to $m$, $\Gamma$ and $\Lambda$, any linear transformation of the original random vector $x$ can be obtained.

Now, the matrices (2.3) and (2.4) are the components of an eigendecomposition of $\Sigma$, as displayed in (2.5). Hence, the diagonal elements of $\Lambda$ contains the eigenvalues of $\Sigma$, and $\Gamma$ contains eigenvectors associated with these eigenvalues. Reversing the approach in the above example gives some insights to the role that eigenvalues and eigenvectors play in the context of covariance matrices. The fact that $\Lambda$ can be interpreted as a scaling matrix and $\Gamma$ interpreted as a rotation matrix, allows to disentangle how a random vector with covariance matrix $\Sigma$ behaves. Most prominently, the eigenvalues in $\Lambda$ give us information regarding the de facto dimensionality of the random vector. In the above example, both the elements of $z$ have variance 2.125, as noted from its covariance matrix $\Sigma$ in (2.5). However, inspecting the eigenvalues and eigenvectors of $\Sigma$ reveals that the majority of the dispersion occurs along one dimension, namely the line $z_1 = z_2$. In the two dimensional case it can be fairly trivial to make the above observation without consulting the eigendecomposition,
Figure 2.1: Top left: Scatter plot for the samples of $x$. Top right: Scatter plot for the samples of $y$. Bottom left: Scatter plot for the samples of $z$. Bottom right: Scatter plot for the samples of $w$. Sample size is $n = 10000$. 


but in higher dimensions it is usually more challenging. For example, letting \( \sigma_{13} = 0.9 \) in (2.1) results in the eigenvalues \{2.8, 0.1, 0.1\}, such that essentially all of the variation of the three-dimensional vector \( \mathbf{x} \) could be represented by a single random variable. The case where one or several eigenvalues are equal to zero results in singular covariance matrices, a case that is further discussed in Section 2.4. Finally, while \( \Lambda \) represents the dispersion around orthogonal axes, \( \Gamma \) represents the rotation of these axes to the random vector’s coordinate system.

Eigendecomposition is a key concept in for example principle component analysis, presented in e.g. Jolliffe (2011), where it is commonly used for dimension reduction in observed data. In this thesis, eigenvalues play an important role in the application of Paper II, where a shrinkage-type estimator based on eigenvalues is derived. They are also prominent in Paper V, where covariance matrix bounds based on eigenvalues are derived. The simulations in Paper II, Paper III and Paper V are also based on pre-defined eigenvalues.

### 2.3 Estimators of the covariance matrix

A very common scenario is that the population covariance matrix of a random vector is unknown, and that this quantity needs to be estimated from observed data. The most standard such estimator is the sample covariance matrix, defined as follows. Suppose \( \Sigma \) is the population covariance matrix of the random vector \( \mathbf{x} \), and let \( \mathbf{x}_1, \ldots, \mathbf{x}_n \) be a sample of \( n \) independent and identically distributed random vectors, while denoting \( \bar{\mathbf{x}} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i \) the sample mean. The sample covariance matrix (SCM) is then computed as

\[
\hat{\Sigma} = \frac{1}{n-1} \sum_{i=1}^{n} (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})',
\]

which can be shown to be an unbiased and consistent estimator, under some regularity conditions. As long as \( n > p \), such that the sample size is larger than the vector dimension, \( \hat{\Sigma} \) will be p.d. almost surely. The complementary case of \( n \leq p \) is discussed in Section 2.4.

The SCM defined in (2.6) can be viewed as an empirical estimator, applicable irre-
gardless of the distribution of the random vector. However, if the distributional family of \( \mathbf{x} \) is known and the covariance matrix can be expressed as a function of the parameters in that distribution, the maximum likelihood estimator (MLE) of the covariance matrix is often preferable to the SCM, since the MLE will provide a lower asymptotic estimator variance than the SCM. In fact, the MLE is asymptotically efficient, meaning that when \( n \to \infty \), the MLE variance reaches the Cramer-Rao bound, the lowest possible variance of an estimator (see e.g. Rao, C.R. and Das Gupta, S. (1989)). In case of a multivariate normal distribution, the MLE of the covariance matrix differs from (2.6) only by the factor \((n - 1)/n\).

Although the MLE is asymptotically efficient, it is possible to obtain estimators that outperform the MLE or the SCM in terms of some estimation loss measure, such as mean squared error (MSE). One such type of estimators are the so-called shrinkage estimators. The idea behind this approach is essentially to shrink the MLE or the SCM towards a deterministic matrix, thus reducing the estimator variance. The shrinkage commonly also introduces a bias, but is specified such that the new estimator still dominates the original one in terms of the predetermined loss measure, such as MSE. An estimator of such type is presented in Ledoit and Wolf (2004) (extended in Bodnar et al. (2014)), and consists of a weighted sum of the SCM and the identity matrix, which is shown to outperform the SCM in terms of MSE, especially when the matrix dimension is large relatively to the sample size. The weighting of this estimator can be seen as a bias-variance trade-off between two extremes: estimating the covariance matrix with the identity matrix leads to larger bias but no dispersion; estimating the covariance matrix with the SCM leads to no bias but larger dispersion. This type of estimator can also be related to the estimation in the Bayesian setting, where the estimator is a combination of the sample information, captured in the likelihood function, and of the previous parameter knowledge, represented by the prior distribution.

Improved estimators of \( \Sigma \) in the multivariate normality case have received particular attention. One such case is when estimators are evaluated using Stein’s loss function, presented in James and Stein (1961) and defined as

\[
L(\Sigma, \tilde{\Sigma}) = \text{tr}(\tilde{\Sigma}\Sigma^{-1}) - \ln|\tilde{\Sigma}\Sigma^{-1}| - p,
\]

(2.7)
for a $p \times p$ covariance matrix $\Sigma$ with associated estimator $\tilde{\Sigma}$. In the multivariate normal case, Stein’s loss closely relates to the Kullback-Leibler divergence, a measure of difference in probability distributions widely applied in e.g. information theory and machine learning (see e.g. Kullback (1959)). For example, Dey and Srinivasan (1985) and references therein discuss several estimators that outperform the MLE for $\Sigma$ in terms of (2.7), where the general idea is to shrink the eigenvalues of the SCM towards some value. The derivation of the estimators are based on the expected Stein’s loss, and on obtaining convenient identities for this quantity. These equalities are commonly denoted Stein-Haff identities of various kinds, due to the original derivations in Stein (1977) and Haff (1979). Many extensions of Stein-Haff type identities and estimators under Stein’s loss have been proposed, for example covering the more general case of elliptically contoured distributions, in e.g. Kubokawa and Srivastava (1999) and Bodnar and Gupta (2009).

In this thesis, Paper II derives the Stein-Haff identity for a class of exponential matrix distributions. The application part of the paper also presents estimators for covariance matrices under Stein’s loss, based on the matrix-variate gamma distribution discussed more closely in Section 2.5. Paper III also proposes a covariance matrix estimator based on this distribution, and shows that it can be beneficial compared to the MLE in several ways.

### 2.4 Singularity and the Moore-Penrose inverse

In Section 2.1, the concepts of positive definite and positive semi-definite covariance matrices were discussed, and that a covariance matrix possesses either of these properties. In brief, the set of p.s.d. matrices contains the set of p.d. matrices, as well as the set of singular matrices. A few important properties of a singular $p \times p$ covariance matrix $\Sigma$ are:

1. Some of the $p$ eigenvalues are equal to zero.
2. $|\Sigma| = 0$.
3. $\text{rank}(\Sigma) < p$. Furthermore, $\text{rank}(\Sigma)$ is equal to the number of non-zero eigenvalues.
4. $\alpha' \Sigma \alpha = 0$ for some non-zero vector $\alpha \in \mathbb{R}^p$. 

17
Furthermore, a symmetric square matrix possessing any of the above properties is singular. With the aid of property (i) above, singularity in a covariance matrix can be interpreted as follows. Let \( x \) be a \( p \times 1 \) random vector with singular \( p \times p \) covariance matrix \( \Sigma \), that has \( k < p \) non-zero eigenvalues. Following the discussion in Section 2.2, \( x \) exhibits dispersion not in \( p \) dimensions, but rather in \( k \) dimensions. Conversely, the \( p \times 1 \) random vector \( x \) can be represented by a linear transformation of a \( k \times 1 \) random vector. For example, letting \( \sigma_{13} = 1 \) in (2.1) yields the eigenvalues \( \{2.867479, 0.1325206, 0\} \) for \( \Sigma \). As such, \( x \) is in effect 2-dimensional. In this case, it can be seen directly from the covariance matrix, since \( \mathbb{V}[x_1] = \mathbb{V}[x_3] = \text{Cov}[x_3, x_3] = 1 \), such that we indeed have \( x_1 = x_3 \) with probability one, and could equivalently define \( x = (x_1, x_2, x_1) \).

The above discussion regards singular population covariance matrices; another important case concerns estimators of covariance matrices. Suppose that the random vector \( x \) has non-singular population covariance matrix \( \Sigma \), which we want to estimate with the sample covariance matrix \( \hat{\Sigma} \), defined in (2.6). As long as the sample size \( n \) is larger than the vector dimension \( p \), then \( \text{rank}(\hat{\Sigma}) = p \) almost surely, and thus the estimator obtains as non-singular. However, if \( n \leq p \), we have that \( \text{rank}(\hat{\Sigma}) = n - 1 < p \), resulting in a singular \( \hat{\Sigma} \). This scenario naturally arises when dealing with high-dimensional data or when samples sizes are limited. Furthermore, a singular \( \hat{\Sigma} \) can be viewed in light of a key concept regarding statistical quantities, namely dimension reduction. A very general notion of a statistic is that it aims to describe a larger amount of data with a much smaller set of data - such as a single value or relatively small matrix. However, a singular \( \hat{\Sigma} \) contradicts this idea. To illustrate this, consider \( n = 2 \) samples of a random vector of dimension \( p = 4 \), a data set which consists of \( np = 8 \) elements. Then the sample covariance matrix \( \hat{\Sigma} \) defined in (2.6) is of dimension \( p \times p \) and, as it is symmetric, has \( p(p + 1)/2 = 10 \) elements. As such, the statistic \( \hat{\Sigma} \) summarizes the 8 elements in the data with 10 elements, inflating the dimension rather than reducing it. But also in the non-ideal case \( n \leq p \), an estimator of \( \Sigma \) might be necessary, given the application at hand.

Moreover, several applications require an estimator of the inverted covariance matrix, \( \Sigma^{-1} \). These include for example discriminant analysis, presented in e.g. Garson (2012),
and portfolio theory, more closely discussed in Chapter 5. In the case of a non-singular sample covariance matrix, an estimator of $\Sigma^{-1}$ can straightforwardly be obtained as $(\hat{\Sigma})^{-1}$. However, if for some reason $\Sigma$ is singular, the standard inverse can not be taken. One approach to deal with this is instead by applying a generalized inverse. The most well-known such inverse is the Moore-Penrose inverse, which for a covariance matrix can be constructed as follows. Suppose $\hat{\Sigma}$ is a $p \times p$ covariance matrix with $\text{rank}(\hat{\Sigma}) = k \leq p$ (hence either singular or non-singular). Now, apply the factorization

$$
\hat{\Sigma} = LDL',
$$

where $D$ is a $k \times k$ diagonal matrix that contains the $k$ non-zero eigenvalues of $\hat{\Sigma}$ while the $p \times k$ matrix $L$ contains the $k$ eigenvectors associated with the non-zero eigenvalues of $\hat{\Sigma}$ as columns. Here $L$ is a semi-orthogonal matrix, an object that is generally characterized by either $L'L = I_k$, or $LL' = I_p$. Note that (2.8) is an alternative characterization of the eigendecomposition (2.5). The Moore-Penrose inverse of $\hat{\Sigma}$ can now be computed as

$$
(\hat{\Sigma})^+ = LD^{-1}L'.
$$

If $\hat{\Sigma}$ is singular it does in general not hold that $(\hat{\Sigma})^+\hat{\Sigma} = I_p$, but we do have that $\hat{\Sigma}(\hat{\Sigma})^+\hat{\Sigma} = \hat{\Sigma}$. Furthermore, $(\hat{\Sigma})^+$ provides the best solution, in the least square sense, to the system of equations $\hat{\Sigma}\mathbf{v} = \mathbf{u}$, where $\hat{\Sigma}$ and $\mathbf{u}$ are given. Thus, as presented in Planitz (1979), for any vector $\mathbf{v} \in \mathbb{R}^p$, it holds that $\|\hat{\Sigma}\mathbf{v} - \mathbf{u}\|_2 \geq \|\hat{\Sigma}(\hat{\Sigma})^+\mathbf{u} - \mathbf{u}\|_2$, where $\|\cdot\|_2$ denotes the Euclidean norm of a vector. If on the other hand $\hat{\Sigma}$ is non-singular, we have by construction that $(\hat{\Sigma})^+ = (\hat{\Sigma})^{-1}$, such that indeed $(\hat{\Sigma})^+$ can be viewed as a generalized matrix inversion. For further reading on the Moore-Penrose inverse, see e.g. Boullion and Odell (1971).

Singular covariance matrix estimators are key concepts in Paper IV and Paper V, where the singularity stems from matrix dimensions exceeding sample sizes. The Moore-Penrose inverse is further applied in Paper V, in the context of estimating the weight vector of the tangency portfolio, an important problem in finance that is discussed further in Chapter 5.
2.5 Wishart distribution

Let \( x_1, \ldots, x_n \) be \( n \) independent and identically distributed samples of the \( p \times 1 \) random vector \( x \sim \mathcal{N}(\mu, \Sigma) \), let \( \bar{x} = \frac{1}{n} \sum_i^n x_i \) be the sample mean and let \( \hat{\Sigma} \) be the sample covariance matrix, defined as in (2.6). Then \( \bar{x} \) and \( \hat{\Sigma} \) are independent and \( \bar{x} \sim \mathcal{N}(\mu, \Sigma/n) \).

Regarding the sample covariance matrix, we have that

\[
(n - 1) \hat{\Sigma} \sim \mathcal{W}_p(n - 1, \Sigma),
\]

where \( \mathcal{W}_p(\nu, S) \) denotes a Wishart distribution of dimension \( p \times p \), with degrees of freedom \( \nu > p - 1 \) and scale matrix \( S > 0 \) as parameters. The Wishart distribution's role in the above context makes it a central concept in multivariate statistics. It was first introduced in Wishart (1928), and is a key probability distribution throughout this thesis, why this section will discuss it in more detail.

Letting \( W \sim \mathcal{W}_p(\nu, S) \), the density function for \( W \), defined on the set of p.d. symmetric \( p \times p \) matrices, is

\[
f(W) = \frac{|W|^{(\nu-p-1)/2}}{2^p p^{p/2} \Gamma_p(\nu/2)|S|^{\nu/2}} e^{-\text{tr}(S^{-1}W)/2},
\]

where \( \Gamma_p(\cdot) \) denotes the multivariate gamma function (see e.g. Gupta and Nagar (2000)). Furthermore, the first moments of \( W \) obtains as

\[
E[W] = \nu S
\]

\[
\mathbb{V}[\text{vec}(W)] = \nu (I_p^2 + K_{p,p}) (S \otimes S),
\]

where \( \otimes \) denotes the Kronecker product and \( K_{p,p} \) is the commutation matrix\(^1\), and \( \text{vec}(\cdot) \) is the operator that stacks the columns of a \( p \times q \) matrix into a \( pq \times 1 \) vector. A property that is applied throughout this thesis is that of affine transformations for the Wishart distribution. It states that if \( W \sim \mathcal{W}_p(\nu, S) \) and \( A \) is a \( q \times p \) matrix of rank \( q \), then \( AW A' \sim \mathcal{W}_q(\nu, AS A') \). An important consequence of this property concerns the marginal

\(^1\)Defined s.t. \( K_{p,q} \text{vec}(A) = \text{vec}(A') \) for any \( p \times q \) matrix \( A \) (see e.g. Harville (1997))
distribution of $W$. Consider the partitions

\[
W = \begin{pmatrix}
W_{11} & W_{12} \\
W'_{12} & W_{22}
\end{pmatrix}, \quad S = \begin{pmatrix}
S_{11} & S_{12} \\
S'_{12} & S_{22}
\end{pmatrix},
\]

where $W_{11}$ and $S_{11}$ are $q \times q$, while $W_{22}$ and $S_{22}$ are $(p - q) \times (p - q)$. Then $W_{11} \sim \mathcal{W}_q(\nu, S_{11})$ and $W_{22} \sim \mathcal{W}_{p-q}(\nu, S_{22})$. These basic results and extensions thereof are significant to the majority of the papers in this thesis.

Moreover, when the Wishart distribution is derived in the context of the sample covariance matrix for a sample of multivariate normal vectors, the degrees of freedom $\nu$ naturally obtains as an integer value related to the sample size. This can be seen as the classical way the Wishart distribution is presented. However, as discussed on p. 87 on Muirhead (1982), the density function (2.9) allows to extend the definition of the distribution to include real-valued degrees of freedom. The Wishart distribution with real-valued degrees of freedom coincides with another distribution, the matrix-variate gamma distribution. If $W \sim \mathcal{W}_p(\nu, S)$, then we also have $W \sim \mathcal{MG}_p(\nu/2, 2S)$, where $\mathcal{MG}_p(\alpha, S)$ denotes the matrix-variate gamma distribution with shape parameter $\alpha > (p - 1)/2$, $\alpha \in \mathbb{R}$, and scale matrix parameter $S > 0$. The classical Wishart distribution with integer degrees of freedom can be viewed as an generalization of the chi-squared distribution to symmetric p.d. matrices, while the matrix-variate gamma distribution can be seen as a generalization of the gamma distribution to symmetric p.d. matrices. Depending on context or branch of literature, either a matrix-variate gamma distribution, or a Wishart distribution with real-values degrees of freedom, might be used to describe the law of a random matrix. Both notations appear in the papers of this thesis.

A closely related distribution that is well studied in the literature is the inverse Wishart distribution. It is often denoted $W^{-1} \sim \mathcal{IW}_p(\nu, S)$, where it follows that $W \sim \mathcal{W}_p(\nu - p - 1, S^{-1})$, see e.g. Theorem 3.4.1 in Gupta and Nagar (2000). This distribution is frequently applied in Bayesian statistics, where it is the conjugate prior of the covariance matrix (see e.g. Koop and Korobilis (2010)). It is also common in portfolio analysis, where many applications requires an estimator of $\Sigma^{-1}$, which is further discussed in Chapter 5. Another related distribution is the singular Wishart distribution, defined in Srivastava
(2003), which is the distribution of the SMC (2.6) of a multivariate normal sample in the case of \( n \leq p \). It has been extensively analyzed in e.g. the portfolio theory setting, which again will be studied closer in Chapter 5. The Moore-Penrose inverse, discussed in Section 2.4, of a singular Wishart distributed matrix is another object of particular interest. Deriving the expectation and variance of this quantity is still an open problem, but e.g. Cook and Forzani (2011) and Imori and Rosen (2020) supplies bounds and approximation of these moments, as well as exact results in the special case of \( S = I_p \).

In this thesis, the Wishart distribution or matrix-variate gamma distribution figures as important pieces in each of the five papers. Paper I derives goodness-of-fit test for the Wishart distribution in a time-series setting. In the application part of Paper II, as well as in Paper III, estimators for the matrix-variate gamma distribution are presented. Paper IV concerns discrete time-series of singular Wishart distributed matrices, while Paper V provides bounds and approximations of the moments for products of the Moore-Penrose inverse of a singular Wishart distributed matrix and a multivariate normal random vector, in a portfolio application.
Chapter 3

Integrated and realized covariance

As mentioned in Chapter 1, return processes for financial assets tend to be highly heteroskedastic, and their behaviour can often exhibit large differences even across a single trading day. A very general approach to describe their variability over a time period is with integrated covariance, a continuous time-varying definition of the return covariance matrix that enters as a key quantity in many financial applications. This chapter aims to introduce integrated covariance along with the empirical analogy, realized covariance, a data-driven measure which has a central role in this thesis.

Let the arbitrage-free log-prices of $p$ assets be described by the following continuous time model:

$$
x(t) = x_0 + \int_0^t \mu(u)du + \int_0^t \Theta(u)d\mathbf{w}(u),
$$

(3.1)

where $x_0$ is a $p \times 1$ vector of the log-prices at $t = 0$, $\mu(t)$ is a $p \times 1$ vector describing the price drift, while $\Theta(t)$ is a $p \times p$ matrix of spot volatilities and $\mathbf{w}(t)$ is a vector of independent standard Brownian motions, where $\mu(t)$ and $\Theta(t)$ are independent of $\mathbf{w}(t)$. Moreover, let the log-return vector of the price process between time $s$ and $t$ be denoted $\mathbf{r}(s,t) = x(t) - x(s)$. Then, by e.g. Theorem 2 in Andersen et al. (2003), we get

$$
\mathbf{r}(s,t) \mid \mathcal{F}\{\mu(u), \Theta(u)\}_{s \leq u \leq t} \sim N \left( \int_s^t \mu(u)du, \int_s^t \Theta(u)\Theta'(u)du \right),
$$

(3.2)

where $\mathcal{F}\{\mu(u), \Theta(u)\}_{s \leq u \leq t}$ is the $\sigma$-algebra generated by $\{\mu(u), \Theta(u)\}_{s \leq u \leq t}$. The integrated covariance between time $s$ and $t$, is then defined as

$$
\mathbf{I}(s,t) := \int_s^t \Theta(u)\Theta'(u)du.
$$

(3.3)

As notable from equation (3.2), the integrated covariance $\mathbf{I}(s,t)$ solely determines the conditional covariance of the asset returns of the price process model (3.1), and it is a
central component in for example option pricing (see e.g. Muhle-Karbe et al. (2010)).

However, the integrated covariance defined in equation (3.3) depends on the full sample path of $\Theta(t)$, which in practice is not directly observable. In order to consistently estimate $I(s, t)$ without prior knowledge of $\Theta(u)$, $s \leq u \leq t$, Andersen et al. (2001a) presents a framework utilizing high-frequency asset price data, denoted realized covariance. The approach is based on the properties of quadratic covariation of the log-return process, which is defined as

$$[r(s, t)] := \lim_{M \to \infty} \sum_{j=1}^{M} r(t_{j-1}, t_{j})r(t_{j-1}, t_{j})^{'},$$

for any sequence of partitions $s = t_{0} < \ldots < t_{M} = t$, with $\sup_{j}(t_{j+1} - t_{j}) \to 0$ as $M \to \infty$, where the limit is in probability. Moreover, the approach utilizes standard results on quadratic covariation for stochastic processes to establish that $[r(s, t)] = I(s, t)$. Now, consider a sample of $M$ log-return vectors recorded at a times $s = t_{0} < \ldots < t_{M} = t$. As a finite sample analogy to equation (3.4), Andersen et al. (2001a) defines the realized covariance between time $s$ and $t$ as

$$R(s, t) := \sum_{j=1}^{M} r(t_{j-1}, t_{j})r(t_{j-1}, t_{j})^{'},$$

such that $R(s, t)$ is a $p \times p$ matrix, where $R(s, t) > 0$ as long as $M \geq p$. The equation (3.4) together with the equality of integrated covariance and quadratic covariation implies that, as $M \to \infty$,

$$R(s, t) \xrightarrow{p} I(s, t),$$

concluding that $R(s, t)$ is a consistent estimator of $I(s, t)$. Hence, $R(s, t)$ can be thought of as an ex-post measurement of the covariability of the asset log-returns between time point $s$ and $t$. It is noticeable that $R(s, t)$ is computed without specifying the underlying processes $\mu(t)$ or $\Theta(t)$, and can thus be considered a completely data-driven measure. Letting the time points $s$ and $t$ represent the opening and closing time of a trading day, $R(s, t)$ can be viewed as an estimator of the covariance matrix for the asset return vector on said trading day, based on $M$ intra-day return vectors. In this regard, it is different to the SCM (2.6) in Section 2.3. Computing the SCM of the covariance matrix of a one day asset return requires a sample of independent and identically distributed
(i.i.d.) daily return vectors. Unless one assumes that the daily return covariance matrix is constant across several days or weeks, such i.i.d. samples are generally unobtainable. Thus, when assuming heteroskedastic asset returns, the realized covariance $\mathbf{R}(s, t)$ is a very useful quantity in relation to traditional estimators. Finally, Barndorff-Nielsen and Shephard (2004) evaluates the measurement error between $\mathbf{R}(s, t)$ and $\mathbf{I}(s, t)$, and derives the asymptotic distribution of $\sqrt{M}(\mathbf{R}(s, t) - \mathbf{I}(s, t))$, for stochastic volatility models of the type (3.1), as mixed Gaussian.

The consistency property of $\mathbf{R}(s, t)$ advocates that larger sample size, or equivalently higher sample frequencies, provide better estimates of $\mathbf{I}(s, t)$. Empirically, this would equate sampling price quotes on the highest frequency possible, perhaps every minute, second or even more frequently. However, when sampling observed asset prices, various systematic disturbances related to the practical aspects of the financial market might deter sampling on very high frequencies. Such disturbances are often jointly denoted market microstructure noise, and are studied in e.g. Aït-Sahalia, Yacine and Yu, Jialin (2009). These include for example discreteness of price recording, bid-ask bounces, and so-called asynchronous price sampling, stemming from the fact that each of the $p$ assets might not be traded simultaneously at every sampled time point. Asynchronous trading induces e.g. the Epps effect, stating that covariation statistics computed from return data sampled on high frequencies tend to be biased towards zero, e.g. found for stock return data in Epps (1979) and for foreign exchange rates in Guillaume et al. (1997). On the other hand, sampling on low frequencies possibly ignores large amounts of data. Thus, several methods that mitigate the market microstructure noise while still utilizing the richness of intra-day price data have been purposed, such as the subsampling strategy in Chiriac and Voev (2011) or the multivariate realized kernel estimator in Barndorff-Nielsen et al. (2011). Sampling issues for realized covariance is considered in Paper IV of this thesis. Large portfolio sizes, market microstructure noise or illiquid assets might result in situations where the realized covariance (3.5) is computed with $M < p$, resulting in a singular matrix $\mathbf{R}(s, t)$, an object that is studied in this paper.
Chapter 4

Time series of realized covariance

While the ex-ante estimation methods presented in Chapter 3 are useful on their own, the interest in financial applications often lies in predicting future outcomes given currently available information. Hence, in an ideal situation one would possibly like to consider the distribution of $I(s,t) \mid \mathcal{F} \{ \Theta(u) \}_{0 \leq u \leq s}$, the integrated covariance of coming time period given the volatility process up to the current time. However, since the full sample path of $\Theta(t)$ is generally not observable, alternatives include predicting future integrated covariance based on the information of the integrated covariance from previous time periods, or based on previously observed realized covariances. Given that the integrated covariance is latent while the realized covariance is observable, an approach that has gained popularity is predicting future realized covariances conditional on historically observed realized covariances, and apply this as a proxy for the future integrated covariance. Such forecast modeling alternatives are investigated in e.g. Andersen et al. (2004), for a general class of univariate stochastic volatility models. The conclusion is that while there is some loss of predictive power when using a realized measure as proxy, compared to the ideal case, it still performs well for moderately large sample sizes. The empirically feasible approach of directly modeling discrete time series of realized covariances based on high-frequency price data, as advocated by e.g. Andersen et al. (2003), has given rise to a vast literature of time-series models. This chapter will discuss several common models of this kind, in particular models that are based on the assumption of an underlying Wishart distribution, presented in Section 2.5.

A stylized fact regarding daily asset log-returns is that time-series of their conditional covariances tend to be clustered and highly persistent. This typical property is naturally inherited for in realized covariance. As an example, consider the univariate time series of realized variance, computed on one day intervals, for the Old National Bancorp stock (ONB) from mid 1997 to mid 2017, shown in Figure 4.1. The left graph shows the
Figure 4.1: *Left:* Daily realized variance for the Old National Bancorp stock from mid 1997 to mid 2017. *Right:* The sample autocorrelation function for the realized variance of the Old National Bancorp stock. The dotted lines represent 95% confidence intervals.

realized variance, which has clear tendencies of clustering - time periods of highly volatile movements are mixed with time periods of low and modest fluctuation. Across the series, there are also several extreme values or spikes, in comparison with the neighbouring observations. The graph also captures two turbulent time periods on the stock market - the so called Dot-com bubble around the millennium shift, and the global financial crisis of 2008. During these periods, the realized variance of the considered stock obtains as substantially larger than for other intervals, indicating sizable asset price movements. The right graph shows the sample autocorrelation function of the series, with lags in number of trading days. Although the autocorrelation decreases rapidly in the first couple of lags, the series shows tendencies of persistence for at least 300 days. This behaviour is not extreme for the considered stock, but rather a pattern among realized stock variances and covariances. With this discussion in mind, a multivariate model that aims to capture the properties of a realized covariance matrix time series should be able to account for the high serial dependence of the observations, as well as the occurrence of extreme values or spikes. Further, it must ensure that any predicted covariance matrices remain positive definite. Finally, from a practical point of view, the model should be parameterized in a computationally feasible manner. This point is important; many financial applications depend on the covariance matrix of a large number of assets. A model is of limited
usefulness if it is not possible to estimate the model parameters with good accuracy and reasonable computation time as the process dimension $p$ grows large.

An approach that has gained much attention is to model the evolution of observed realized covariance matrices with a centralized Wishart process. The stochastic properties of the Wishart distribution, presented in Section 2.5, ensures that realizations drawn from it are positive-definite, making it suitable for the problem at hand. In the following, let the realized covariance computed for trading day $t$ be denoted $R_t$, and denote the filtration based on historical observations up to and including trading day $t$ by $\mathcal{F}_t$. According to such a model, for a time series of $p \times p$ realized covariance matrices $\{R_t\}$ with filtration $\mathcal{F}_t$, let

$$R_t \mid \mathcal{F}_{t-1} \sim \mathcal{W}_p(\nu, S_t / \nu), \quad (4.1)$$

where $\mathcal{W}_p(\nu, S_t / \nu)$, denotes the Wishart distribution of dimension $p$, with $\nu > p - 1$, $\nu \in \mathbb{R}_+$ degrees of freedom and $p \times p$ scale matrix $S_t / \nu$, with $S_t > 0$. Since $\mathbb{E}[R_t \mid \mathcal{F}_{t-1}] = \nu S_t / \nu = S_t$, the scale matrix $S_t$ can be though of as the conditional mean of the realized covariance matrix, while its variability is determined by both $S_t$ and $\nu$. Furthermore, Section 2.5 introduces the classical Wishart distribution as a sum of outer products of i.i.d. multivariate normal vectors. However, the time series models with the structure (4.1) do in general not assume any particular distribution for the intra-day returns that $R_t$ is constructed from. Instead, the assumption of a conditional Wishart distribution is applied directly to the object $R_t$.

Given the basic setup described by equation (4.1), what remain is to specify the evolution of $S_t$. Apart from being able to capture the dynamics in observed data, the specification should ensure that $S_t$ remains positive-definite. In recent years, a several approaches on how to model $S_t$ have been suggested in the literature. For example, Jin and Maheu (2012) suggest a multiplicative component model, specifying the scale matrix with

$$S_t = \left[ \prod_{j=K}^{1} \Gamma^d_{t,j} / 2 \right] A \left[ \prod_{j=1}^{K} \Gamma^d_{t,j} / 2 \right]$$

$$\Gamma_{t,t} = \frac{1}{l} \sum_{i=0}^{l-1} R_{t-i},$$

28
with \( 1 = l_1 < \cdots < l_K \), where \( \mathbf{A} \) is a \( p \times p \) positive-definite symmetric matrix and \( d_j, j = 1, \ldots, K \) positive scalar parameters, ensuring that \( \mathbf{S}_t \) is positive-definite (see the properties of p.d. matrices in Section 2.1). The persistence structure of \( \mathbf{R}_t \) can be captured by the matrices \( \mathbf{\Gamma}_{t,l} \), consisting of sample averages of lagged realized covariances, while the values of \( d_j \) adjust the magnitude of their effect. A model with additive components is also proposed by the authors. Another model that has gained much attention is the conditional autoregressive Wishart (CAW) model presented in Golosnoy et al. (2012), where the scale matrix dynamics are described by

\[
\mathbf{S}_t = \mathbf{C}\mathbf{C}^\prime + \sum_{i=1}^{r} \mathbf{B}_i \mathbf{S}_{t-i} \mathbf{B}_i^\prime + \sum_{i=j}^{q} \mathbf{A}_j \mathbf{R}_{t-j} \mathbf{A}_j^\prime,
\tag{4.2}
\]

where \( \mathbf{A}_1, \ldots, \mathbf{A}_q, \mathbf{B}_1, \ldots, \mathbf{B}_r \) and \( \mathbf{C} \) are \( p \times p \) parameter matrices, where \( \mathbf{C} \) is lower triangular. In this model, the scale matrix can be described as a linear function of historical realized covariances and their conditional means, such that \( \mathbf{S}_t > 0 \) is guaranteed (again see Section 2.1). The structure (4.2) is often denoted as the Baba, Engle, Kraft and Kroner (BEKK) specification, presented in Engle and Kroner (1995) regarding the multivariate GARCH model. The authors also suggests extending (4.2) with specifications that explicitly accounts for long-run memory type of dynamics by including realized covariances computed on for example monthly horizons, inspired by the heterogeneous autoregressive (HAR) approach of Corsi (2009) and the mixed data sampling (MIDAS) approach adapted to GARCH models in e.g. Engle et al. (2013). The multivariate high-frequency (HEAVY) models presented in Noureldin et al. (2012) exhibit similarities to (4.2), but facilitates mixing observations on high and low frequencies. In Anatolyev and Kobotaev (2018) the CAW model (4.2) is further extended by allowing for asymmetry in the covariance dynamics depending on recent up- or downward changes in asset prices. It is denoted the conditional threshold autoregressive Wishart (CTAW) model, were \( \mathbf{A}_i \) and \( \mathbf{B}_i \) in (4.2) are modeled as

\[
\mathbf{A}_i = \mathbf{A}_i + \sum_{j=1}^{p} \mathbf{H}_{i,j} \mathbf{I}_{j,t-i}
\]

\[
\mathbf{B}_i = \mathbf{B}_i + \sum_{j=1}^{p} \mathbf{G}_{i,j} \mathbf{I}_{j,t-i},
\]
where $I_{j,t}$ is a direction indicator for the price of asset $j$ at time $t$, while $A_1, \ldots, A_q, B_1, \ldots, B_r, H_{i,j}, i = 1, \ldots, q, j = 1, \ldots, p$ and $G_{i,j}, i = 1, \ldots, r, j = 1, \ldots, p$ are parameter matrices. Closely related is the Wishart autoregressive (WAR) model of Gouriéroux et al. (2009), where instead the assumption of a non-central Wishart distribution is employed. In this model, the the dynamics are instead described by the non-centrality parameter. In Yu et al. (2017), the generalized conditional autoregressive Wishart (GCAW) model is presented. It is specified with both a scale matrix and a non-centrality parameter, and is thus a generalization of both the WAR and the CAW model described above.

The various specifications of the conditional mean $S_t$ in the above models facilitates to capture serial dependence often observed in realized covariance. However, the discrete time series $\{R_t\}$ also tend to exhibit extreme values, exemplified in Figure 4.1 regarding the univariate case of realized variance for the ONB stock. But the Wishart distribution, that the above models are based on, does not possess the property of fat tails, meaning that the probability of observing an extremely deviating value in a sample of this distribution is very low. Hence, to facilitate extreme value observations with reasonable likelihood, corresponding elements in the conditional mean $S_t$ of the Wishart models must obtain as particularly large at the trading days where the spikes are observed. An alternative approach is to instead apply a matrix distribution with fat tails, prescribing larger probability to extreme value observations. This is the approach of Opschoor et al. (2018), where a matrix-F distribution for the realized covariance matrices is applied. Other models for $\{R_t\}$ include e.g. Bauer and Vorkink (2011) and more recently Archakov et al. (2020), which works with matrix log-transformations of the realized series. In the latter, univariate time series are first obtained for the realized variance of each considered asset. From these series, a discrete time series of correlation matrices can be obtained and modeled separately, in the spirit of the DCC-GARCH model presented in Engle (2002). Applying a normal distribution assumption in the modeling of these log-transformed quantities appears to have some empirical support, which is similarly noted in e.g. Andersen et al. (2001b).

In this thesis, modeling of realized covariance is relevant in a majority of the papers. To a large extent, the Wishart models described in this chapter are evaluated by forecast
accuracy. In Paper I, a framework of goodness-of-fit tests is presented, that allows evaluating the assumption of no serial correlation and the distributional assumption of models based on an underlying centralized Wishart process. Paper II provides identities regarding a class of p.d. matrix distributions of exponential type, in which the Wishart distribution is included, providing possible candidate distributions when modeling realized covariance. In the application part, the paper also provides estimators for the scale matrix parameter of the matrix-variate gamma distribution. Such results can be applied for rudimentary models of realized covariance, for example where it is assumed the scale matrix is constant across time periods. A similar modeling approach can be facilitated using the results in Paper III, where a closed-form estimator for the matrix-variate gamma distribution is presented. Paper IV considers the important case of singular realized covariance matrices, which can occur when the size of the asset portfolio outgrows the amount of available high quality data, e.g. due to the reasons discussed in Chapter 3. The paper extends the rich family of Wishart models discussed above to the case of singular realized matrices with the singular conditional autoregressive Wishart (SCAW) model.
Chapter 5

Portfolio theory

Chapters 2 to 4 discuss the construction, properties, estimation and modeling of the covariance matrix, in general and in the case of asset returns. Portfolio theory, first introduced in Markowitz (1952), on the other hand, applies the covariance matrix by considering how to optimally allocate an investment between a number of assets in a portfolio. The analysis is based on the mean vector and covariance matrix for the asset return vector, together with the preferences of the investor. This chapter discusses this framework together with three of the most common portfolio allocations and how they appear in the papers of this thesis: the global minimum variance portfolio, the tangency portfolio and the equally weighted portfolio.

In the following, assume that an investor considers dividing a wealth, normalized to one, between \( p \) different risky financial assets with expected return \( \mu \) and covariance matrix \( \Sigma \). In some setups, there is also assumed to exist a risk-free asset, such as a government bond, that exhibits zero variance and typically a relatively low return, denoted \( r_f \). Given the preferences of the investor and knowledge regarding the mean vector and covariance matrix of the asset returns, the portfolio theory framework aims to produce a \( p \times 1 \) weight vector \( w \) which dictates how the wealth is optimally allocated between the risky assets. We have that \( w \in \mathbb{R}^p \), such that negative weights, and hence short sales of assets, are allowed. Furthermore, under the assumption of a risk-free asset, it is assumed that the proportion \( 1 - w'1_p \) of the wealth is invested into the risk-free asset, where \( 1_p \) is a \( p \times 1 \) vector of ones. If this amount is negative, it is assumed the investor borrows the amount at the risk-free rate. The expected return of the portfolio obtains as \( w'\mu + (1 - w'1_p)r_f \) under the assumption of a risk-free asset and \( w'\mu \) otherwise, while the variance of the portfolio obtains as \( w'\Sigma w \).

Moreover, the preferences of an investor are captured with a target function to optimize against, possibly given some constraints. Such functions are sometimes formulated as
utility functions, stating how much value, or utility, an agent obtains from some quantity of interest. The quantity is often assumed random, why it is typical to instead optimize against the expected utility function. An important parameter in the context of investor preferences is the risk-aversion parameter $\alpha > 0$. It aims to capture the investors attitude towards risk, where larger values of $\alpha$ implies that an investor is less willing to risk their wealth, and vice versa. In practice, $\alpha$ is usually obtained based on qualitative information from the investing agent, and will be assumed as given in this presentation.

One fundamental allocation strategy is the global minimum variance portfolio (GMV). It combines the considered risky assets to obtain the portfolio with the smallest possible variance, and is thus an optimal solution for an investor who wants to minimize portfolio variance, or risk, assuming there is no risk-free asset to invest in. It corresponds to minimizing $w'\Sigma w$ such that $w'1_p = 1$. The condition is due to the fact that all the wealth is assumed to be invested into the $p$ risky assets, and hence the weights must sum to 1. Denoting the global minimum variance portfolio’s weight vector as $w_{GMV}$, one can show that

$$w_{GMV} = \frac{1}{1_p'\Sigma^{-1}1_p} \Sigma^{-1}1_p.$$

Straightforward calculations allow to obtain the variance of the portfolio return as $(1_p'\Sigma^{-1}1_p)^{-1}$. It is furthermore notable that this portfolio is solely determined by the covariance matrix of the asset returns, and is thus independent of the expected returns.

Another important portfolio is the so-called tangency portfolio (TP). It is here denoted $w_{TP}$, and assuming the possibility to invest into a risk-free asset with return $r_f$, it obtains as

$$w_{TP} = \alpha^{-1} \Sigma^{-1}(\mathbf{\mu} - r_f 1_p). \quad (5.1)$$

Hence it depends on both the mean and covariance of the asset returns, as well as the risk aversion parameter of the investor. The vector $w_{TP}$ is the solution to the mean-variance
optimization problem

\[
\max_w w'\mu + (1 - w'1_p)r_f - \frac{\alpha}{2}w'\Sigma w.
\]

It represents a trade-off between the portfolio return \( w'\mu + (1 - w'1_p)r_f \), which investors desire to be large, and the portfolio variance, or risk, \( w'\Sigma w \), which investors commonly desire to be small. The TP allocation moreover appears as the solution to maximization problems based on the commonly used Sharp ratio, the ratio between portfolio mean and portfolio risk, and as the solution to maximization problems based on utility functions of quadratic and exponential forms. Further discussions on portfolio optimization problems can be found in e.g. Bodnar et al. (2013).

The third and final portfolio allocation presented here is the so-called equally weighted portfolio (EW). In this case, the wealth is allocated proportionally between the \( p \) assets, such that, denoting the equally weighted portfolio weight vector as \( w_{EW} \), we have

\[
w_{EW} = \frac{1}{p}1_p.
\]

This allocation is not the solution to some specific optimization problem, but is nonetheless one of the most important portfolios, since it appears to empirically outperform many more sophisticated portfolio allocations in terms of various risk and return measures, as discussed in e.g DeMiguel et al. (2009). A very attractive feature of this portfolio is further that it requires no knowledge regarding the mean or variance of the asset returns, therefore it is not affected by for example parameter estimation error, as discussed further below. In this regard, the EW can be viewed as a suitable allocation if the investor is averse to estimation error.

It is notable that both the GMV and TP depend on parameters of the asset return vector distribution, namely \( \mu \) and \( \Sigma \). In practice these quantities are unknown, and have to be estimated from historical return data. Consequently, it is of great importance to study the statistical properties for various estimators of the portfolio weight vectors \( w_{GMV} \) and \( w_{TP} \). Regarding the GMV weight, for example Frahm and Memmel (2010) considers shrinkage estimators of \( w_{GMV} \), while Glombeck (2014) and Bodnar et al. (2018)
studies statistical inference and estimation of the GMV in high dimensions. Concerning estimation of the TP weight vector, e.g. Britten-Jones (1999) presents an exact test of the estimated weights in the multivariate normal case; Bodnar and Okhrin (2011) derives the density for, and several exact tests on, linear transformations of estimated TP weights; Bauder et al. (2018) considers estimating $w_{TP}$ with a Bayesian approach. A large proportion of the studies of these quantities are based on the assumption that the asset returns follow an i.i.d. normal distribution. Empirically, this assumption appears to have little support regarding daily returns, but seems more suitable when considering returns over lower frequencies, such as weekly or monthly, as discussed in e.g. Aparicio and Estrada (2001). Naturally, estimating the model parameters on for example weekly return data limits the sample size, particularly since an important question is over what time intervals the mean $\mu$ and covariance matrix $\Sigma$ can be considered constant, as discussed in Chapters 3 and 4. This affects the precision of the parameter estimators of course, but regarding estimators for $\Sigma$, singularity must also be kept in mind. Assuming that returns are identically distributed over one year, a sample of weekly return data of size about $n = 50$ could be obtained to estimate the parameters. As discussed in e.g. Section 2.4, the SCM in (2.6) is non-singular if $p < n$, i.e. the portfolio size is smaller than 50 in this case. But it is not uncommon to regard portfolios with substantially larger amount of assets, such as $p = 100$ or even $p = 1000$, as considered in e.g. Hautsch et al. (2015) or Ding et al. (2020). The potential singularity of an estimator for $\Sigma$ is particularly noteworthy, since both $w_{GMV}$ and $w_{TP}$ rely on the inverse of the covariance matrix.

In this thesis, Paper V studies the estimation of the TP weight vector $w_{TP}$ for the singular case $p > n$, under the assumption of normally distributed returns. The usual procedure of estimating $\Sigma^{-1}$ with the standard inverse of the SCM is not possible, since the SCM is singular when $p > n$. Instead $\Sigma^{-1}$ is estimated applying the Moore-Penrose inverse discussed in Section 2.4. Unfortunately, as of yet there exists no derivation for the moments of the Moore-Penrose inverse in this case, and consequently neither for this estimator of $w_{TP}$. However, Paper V provides several bounds and approximations for the TP weight estimator based on the Moore-Penrose inverse, in the case $p > n$. The GMV and EW portfolio are considered in Paper IV, where they are used to evaluate the forecast
accuracy of the time series model presented in the paper.
Chapter 6

Summary of papers

Paper I: Goodness-of-fit tests for centralized Wishart processes

The fit of the centralized Wishart models discussed in Chapter 4 have, up to this point, commonly been evaluated through forecasting accuracy on out-of-sample data, and in some cases by testing for autocorrelation in standardized residuals. While such procedures facilitates some diagnostics on the usefulness of the model, they do not properly appraise the distributional assumption of an underlying centralized Wishart process. This paper presents several goodness-of-fit tests that evaluates this important assumption. The tests are based on an extension of Bartlett decomposition, which under the null hypothesis allows obtaining independent standard normal random variables, to which classical tests of normality and serial correlation is applied.

Moreover, while the null distributions of the described tests are derived with knowledge of the true model parameters, in practice, the parameters of the assumed model need to be estimated. The model parameters can be consistently estimated with the maximum likelihood method, but some amount of uncertainty is present when the model is fitted to finite samples. In order to evaluate how the presented tests perform under parameter uncertainty, a simulation study based on the CAW model, represented by equation (4.2) in Chapter 4, is presented in the paper. It shows that the tests for autocorrelation are able to detect violations of the correct lag order (denoted $r$ and $q$ in equation (4.2)), unless they are small. The study also suggests that the tests for normality are able to detect violations to the underlying Wishart assumption. Included as a benchmark procedure is the diagnostic approach applied in Golosnoy et al. (2012), which tests serial autocorrelation of univariate standardized residuals. The simulation study suggests that
this approach is able to recognize misspecification of lag orders to some extent, but is not able to detect violations of the distributional assumption.

Finally, the goodness-of-fit tests introduced in the paper are applied to CAW models of various lag orders fitted to a time series of realized covariance data based on six liquid stocks traded on the New York Stock Exchange from 2000 to mid-2008. The null hypothesis of a correct model assumption is rejected for all suggested tests, showing no support of a good model fit.

**Paper II: Stein-Haff identity for the exponential family**

While the matrix models discussed in Chapter 4 assume specific distributions, this paper is more general in nature and rather considers p.d. symmetric random matrices of the exponential family. For these distributions, with certain conditions on the density function, the Stein-Haff identity is derived, making up the main contribution of the paper. This identity, discussed in Section 2.3, can be applied in order to e.g. improve estimators given some loss function or to compute distribution moments. Originally derived in Stein (1977) and Haff (1979) with the aim of improving estimation of the covariance matrix of a multivariate normal population under Stein’s loss function, it has since been computed for several other important distributions.

In addition to the identity derived in the general case, an application to the matrix-variate gamma distribution, mentioned in Section 2.5, with known shape parameter is included. Moreover, the expected Stein’s loss for the maximum likelihood estimator of the scale matrix parameter of this distribution is computed. Furthermore, the derived identity is applied in order to obtain a condition under which an orthogonally invariant estimator of the scale matrix outperforms the MLE, under Stein’s loss. With this condition established, an estimator that dominates the MLE is presented.

Moreover, to support the theoretical results, a small simulation study is presented. It shows that the mean value of the considered loss function is indeed smaller for the introduced estimator, than for the MLE. Furthermore, the difference seems larger when the true scale matrix parameter is equal to the identity matrix, rather than a scale matrix with non-zero off-diagonal elements.
In the context of realized covariance discussed in Chapters 3 and 4, the results in this paper can be applied in order to e.g. derive estimators for a basic model that assumes observations are generated from either a matrix-variate gamma distribution, or some other random matrix distribution of the exponential family. Such a model could be used on its own, or for example as a step in the estimation of a more complex model.

**Paper III: Closed-form estimator for the matrix-variate gamma distribution**

Commonly, the parameters of the matrix-gamma distribution discussed in Section 2.5 are estimated with the maximum likelihood method, as it provides the smallest asymptotic estimator variance. However, there exist no analytical solution for this estimator, why it needs to be computed with numerical optimization. Such methods tend to exhibit various drawbacks: they require computation time, which increases with the matrix dimension and the sample size; they need some start value as an input, which affects the computation time; there may be issues regarding numerical convergence. In addition to the requirement of a numerical procedure, the MLE tends to be imprecise when the scale matrix parameter is close to singular, or when the shape parameter is close to its lower bound.

This paper presents a closed-form estimator for the parameters of the matrix-variate gamma distribution. Hence, its computation does not require any numerical optimization. Moreover, the presented estimator does not appear to share the imprecision of the maximum likelihood estimator (MLE) for the parameter regions discussed above. The estimator is based on the moments of a transformation of the observed samples. The matrices resulting from the transformation have independent diagonal elements, a property that appears to reduce the variance of the estimator.

The properties of the new estimator are compared to the MLE in a simulation study, in terms of mean-squared error (MSE) and computation time. It shows that the presented estimator has a much lower MSE for when the scale matrix parameter is close to singular, or when the shape parameter is close to its lower bound, but that the MLE has a lower MSE in other parameter regions investigated. Furthermore, the computation time of the
closed-form estimator is lower than the computation time of the numerically obtained MLE, which is to be expected. Finally, the study considers the difference in computation time of the MLE for different start values as input to the numerical optimization procedure. It compares the computation time when arbitrary start values are used, compared to when the estimate obtained from the presented closed-form estimator is used as start value. The study shows that applying the new estimator to obtain start values reduces the computation time of the MLE substantially, particularly so when the matrix dimension is large.

Similar to the results in Paper II, the methods presented in this paper can for example be applied to provide an estimator for a model of daily realized covariance, where it is assumed realized covariance matrices follow a matrix-variate gamma distribution with constant parameters over certain time periods. Such models can be represented by the CAW model (4.2), that is presented in Chapter 4, with lag parameters \( r = q = 0 \).

**Paper IV: Singular conditional autoregressive Wishart model for realized covariance matrices**

The discrete time series models in Chapter 4 focus on the case where the daily realized covariance matrices are assumed to be non-singular. But, as discussed in Chapter 3, large portfolio dimension, asset illiquidity or market microstructure noise might result in a situation where the constructed daily realized covariance matrix obtains as singular, a matrix property introduced in Section 2.4. As large portfolio dimensions are common in practice, while high quality data on intra-day returns are limited, this case does indeed warrant attention.

This paper aims to capture the dynamics of discrete time series of singular realized covariance matrices with the singular conditional autoregressive Wishart (SCAW) model, extending the large family of econometric Wishart models discussed in Chapter 4. It adapts the scale matrix BEKK-structure described by (4.2) to the singular Wishart distribution, and presents several results on the stochastic properties of this model, which allows deriving parameter conditions under which the model is weakly stationary.
As discussed above, the singular case is closely related to large portfolio dimensions, why it is important to keep dimension scalability in mind, such that parameter estimation remains computationally feasible and accurate as the number of considered assets grows large. With this in regard, the SCAW model is further adapted to the high-dimension case by covariance targeting, and sectorwise parameterization. Covariance targeting replaces the parameter matrix product $C C'$ in (4.2) by an expression based on the parameter matrices $A_j$ and $B_i$, $j = 1, \ldots, q$, $i = 1, \ldots, r$, and the unconditional mean of the time series. In the application of covariance targeting, the time series is standardized in order to obtain straightforward parameter restrictions that ensure the positive-definite property of the scale matrix is maintained, similar to the approach in Noureldin et al. (2014) in the ARCH case. The sectorwise specification introduced in this paper, on the other hand, utilizes that assets that belong to the same market sector may exhibit similar price dynamics. With this specification, assets of the same sector are assumed to have identical parameters in $A_j$ and $B_i$. In turn, this means that the number of parameters in the model relies on the number of market sectors the considered assets belong to, not the number of actual assets. Both of these approaches significantly reduces the number of model parameters that needs to be estimated, increasing the computationally feasibility, particularly for very large portfolios. In addition, the heterogeneous autoregressive approach presented in Corsi (2009) is adapted to the SCAW model.

The introduced model is evaluated by out-of-sample forecast accuracy, and compared to the multivariate GARCH model. Several measures of accuracy is used: the Frobenius norm of forecast error; squared error of forecasted standard deviation of an equally weighted portfolio; the variance obtained when the weights of the global minimum variance portfolio are computed using the forecasted realized covariance. Moreover, a number of different forecast horizons are considered. For almost all of these measures, the SCAW model, with various specifications, outperforms the benchmark model with equivalent specifications, where the difference is statistically significant. The sectorwise specification appears to be particularly successful, which is promising since it scales excellent with dimension, making it a valid candidate for very large asset portfolios.

Finally, it is noteworthy that the goodness-of-fit tests presented in Paper I is derived
for discrete time series with non-singular Wishart entries, and is hence not applicable to
the model presented in this paper. Even with an adaptation to the singular case, it should
be kept in mind that some of the tests proposed in Paper I rely on asymptotic results, and
may have to be revised for the type of high-dimensional setting that the SCAW model
generally is applied in.

Paper V: On the mean and variance of the estimated
tangency portfolio weights for small samples

The tangency portfolio, as discussed in Chapter 5, is a central allocation strategy in
portfolio theory. Its derivation depends on the mean and covariance matrix of the asset
return vector, which in general are unknown and needs to be estimated from historical
data. Consequently, it is highly relevant to study the statistical properties of the tangency
portfolio weight vector computed using estimated parameters. In the literature, the most
common approach is to assume that the sample size is larger than the portfolio dimension,
which, for example, facilitates a sample covariance matrix that obtains as non-singular.
This is important, since tangency weight vector described by (5.1) is computed using the
inverse of the covariance matrix, and this quantity is naturally replaced by the inverse of
the sample covariance matrix.

However, as mentioned in Chapter 5, various factors can result in a situation where
the portfolio size outweights the sample size. The empirical observation that the covari-
ance matrix of the asset return vector tends to shift over time is one such aspect. In this
situation, the sample covariance matrix obtains as singular, which can not be inverted in
the standard sense. In contrast, as discussed in more detail in Section 2.4, the general-
ized inverse, often denoted the Moore-Penrose inverse, can be applied even if the sample
covariance matrix is singular. This solution allows to obtain an estimator of the tangency
portfolio weight vector, even when sample size is smaller than the portfolio dimension.
However, an issue arises in obtaining the statistical properties of such an estimator, since
there, as of yet, exists no derivation of the moments of the Moore-Penrose inverse for the
sample covariance matrix in the case of normally distributed vector observations. Conse-
sequently, there is also no exact derivation of the moments of the estimator of the tangency portfolio weight vector computed with the Moore-Penrose inverse.

Hence, this paper aims to provide, under the assumption of normally distributed returns, bounds and approximations of the moments for the Moore-Penrose inverse based tangency portfolio weight vector estimator, denoted $\tilde{w}$ in the paper. In addition, it supplies exact results when the population covariance matrix of the return vector is the identity matrix, as well as exact results of a tangency portfolio weight vector estimator computed using the reflexive generalized inverse, another inverse candidate for singular sample covariance matrices.

These results are then studied by simulation, where they are evaluated through several measures and for a number of portfolio sizes, sample sizes and population parameters sets. They suggest that the moment bounds on $\tilde{w}$ are closest to the observed sample moments when there is a low dependency implied by the population covariance matrix. The results also suggest that in some cases, the moments of the estimator based on the reflexive generalized inverse can be used as an approximation for the moments of $\tilde{w}$. 
Sammanfattning

Kovariansmatrisen för tillgängsavkastningar, som beskriver fluktuationer i tillgängsriter, spelar en avgörande roll för att förstå och förutsäga finansmarknader och ekonomiska system. Under de senaste åren har begreppet realiserade kovariansmått blivit ett populärt sätt att med precision skatta kovariansmatriser för avkastningar med hjälp av högfrekvent data. Denna avhandling innehåller fem forskningsartiklar som studerar tidsserier av realiserade kovariansmatriser, skattare för relaterade slumpmatris-fördelningar och fall där stickprovets storlek är mindre än antalet beaktade tillgångar.


Artikel IV introducerar en ny modell för diskreta tidsserier av realiserade kovariansmatriser som erhålls singulära. Detta fall uppstår när matrisdimensionen är större än antalet högfrekventa avkastningar som är tillgängliga för varje handelsdag. Eftersom
modellen främst förekommer då ett stort antal tillgångar beaktas fokuserar artikeln också på att tillgodose genomförbar skattning i höga dimensioner. Modellen anpassas till 20 års högfrekvensdata på 50 aktier, och utvärderas med hjälp av prognosprecision utanför stickprovet, där den överträffar den typiskt använda GARCH-modellen med hög statistisk signifikans.

Artikel V rör skattning av tangensportfölj-vektorn i fall där antalet tillgångar är större än den tillgängliga stickprovsstorleken. Skattaren innehåller Moore-Penrose-inversen av en Wishart-fördelad matris, ett objekt för vilket medelvärdes- och dispersionsmatrisen ännu inte är härledda. Även om inga exakta resultat existerar kompletterar artikeln kunskapen om statistiska egenskaper i portföljteori genom att tillhandahålla gränsvärden och approximationer för momenten för denna skattare, samt exakta resultat i speciella fall. Slutligen undersöks gränsvärdena och approximationerna genom simuleringar.
References


49


Part II

Papers