Seminar Recent Research in Econometrics

FREE UNIVERSITY BERLIN

Comparison of different Parametrization Methods for Covariance Matrix Prediction

How to ensure positive definite Covariance Matrix Predictions?

Author:	Hannah Meyer zu Tittingdorf
Student ID:	5461295
Adress:	Brüsseler Straße 28a, 13353 Berlin
E-Mail:	h.meyerzutittingdorf@fu-berlin.de
Lecturer:	Prof. Dr. Lars Winkelmann
Submission Date:	March 10, 2023

Contents

1 Introduction

Finding the optimal portfolio of assets is a major topic in finance. The key point of portfolio management is to minimize the variance through diversification and maximize the returns. The portfolio choice depends on the expected return and its variance. The decision which combination of return and variance of the portfolio to prefer depends on the preferences of the investor. For finding the optimal portfolio, not only the variances but also the correlations between the assets are relevant. If the correlation between two assets is high, the variance of the portfolio is higher than when the correlation is low (Markowitz, 1952, pp. 71). Therefore estimating and predicting covariance matrices plays a key role in portfolio optimization.

In the univariate case, many models such as the GARCH model are available to estimate variances. They use constraints to ensure the positivity of the estimated variance, which can be easily calculated in the univariate case. In the multivariate case, the same problem appears. Covariance matrices have to be at least positive semi-definite by definition. To solve this problem two approaches exist. The first one uses multivariate time series models with constraints and the second uses parametrization methods to ensure positivity (Bucci et al., 2022, p. 1). Constraint optimizations are difficult to calculate. Therefore Pinheiro and Bates (1996) recommend using unconstrained optimization with a parametrization. The parametrizations are applied to the realized covariance matrix which is a common measure of volatility in financial applications. The forecasting model is usually applied to the parameters of the covariance matrix transformation. Therefore, the forecasting accuracy of a model for predicting the covariance strongly depends on the parametrization method.
Bucci et al. (2022) compared the forecasting performance of four different parametrization methods. I repeated their procedure with stock data of the DAX and added a multivariate time series model with constraints to ensure positivity.

The first two parametrization methods were suggested by Pinheiro and Bates (1996), on the one hand, the Cholesky decomposition and on the other matrix logarithm transformation. The third method uses a positive semidefinite matrix approximant introduced by Higham (1988). The last method is the newest, developed by Archakov and Hansen (2021). This approach is based on the matrix logarithmic transformation of the correlation matrix. The predictions of the parametrized matrices are produced from a Vector Autoregressive (VAR) model. In addition to the parametrization methods, I used the CCC-GARCH model from Bollerslev (1990). Even if the constrained optimizations are difficult to calculate, it is interesting to analyze the performance of this different approach compared to the parametrization methods. Following Bucci et al. (2022) I use the Procrustes and Euclidean distance to compare the forecast performance of the different methods. Furthermore, I analyze the minimum variance portfolios based on the different covariance forecasts. Because portfolio optimization is the

field where covariance predictions are needed, it is of interest to compare the performance of the different approaches directly in their field of application.

The structure of the paper is as follows. First, chapter 2 explains the four different parametrization methods. Chapter 3 describes the VAR model and the multivariate CCC-GARCH model and makes clear the prediction procedure. Then, in chapter 4, the underlying stock data are presented. Finally, in chapter 5 the results are shown and discussed.

2 Realized Covariance Matrix Parametrization Methods

Especially in financial applications it is common to use the realized covariance as a measure of volatility. The realized covariance matrix can be computed as

$$RC_t = \sum_{\tau=1}^{N_t} r_\tau r_\tau' \tag{1}$$

where r_τ is the n-dimensional vector of returns. The returns are sampled at higher frequencies $\tau = 1, ..., N_t$, where N_t is the number of observations in the t-th period (Bucci et al., 2022, pp. 2). I used daily returns to calculate the monthly realized covariance matrix. Therefore $N_t = 21$ which is the average number of trading days per month and n is the number of included stocks.

By definition, a covariance matrix has to be symmetric and at least positive semi-definite. A n x n square matrix A is positive definite if $c'Ac > 0$ for all nonzero n x 1 vectors c. Similarly, A is positive semidefinite if $c'Ac \geq 0$ for all nonzero n x 1 vectors c. Another definition uses the spectral decomposition of a matrix into eigenvalues and eigenvectors. A matrix is positive definite if all its eigenvalues are positive, and positive semi-definite if all eigenvalues are non-negative (Horn & Johnson, 2012, pp. 425).

When predicting covariance matrices a crucial point is to guarantee to estimate at least positive semi-definite matrices. In the following subsections, I will introduce four different parametrization methods which deal with this problem. Instead of predicting the parametrized matrices the operator $vech(\cdot)$ is applied to the matrices. It returns the lower triangular of a symmetric matrix and its diagonal as a vector. Then a time series model is used to estimate this vector and afterward, we calculate the predicted matrix from this vector by reversing the parametrization. Moreover, every parametrization leads to a set of parameters at time t ψ_t that identifies RC_t (Bucci et al., 2022, pp. 2).

2.1 Cholesky Decomposition

The Cholesky decomposition parametrizes RC_t in the product of an upper triangular matrix L_t and its transpose. An upper triangular matrix has only values in the upper triangular including the diagonal and all other entries are zero. This decomposition is only defined for positive definite matrices.

$$RC_t = L_t' L_t \tag{2}$$

Instead of predicting RC_t, $\psi_t = vech(L_t)$ is predicted through a time series model. After forecasting, inverting the decomposition can only lead to a positive definite matrix by definition (Bucci et al., 2022, p. 4).

$$\widehat{RC}_t = \widehat{L}_t' \widehat{L}_t \tag{3}$$

If the underlying data includes positive semidefinite realized covariance matrices, a modification is needed. It is possible to add a small positive diagonal matrix to the positive semidefinite matrix to make it positive definite (cf. Schnabel and Eskow, 1990). Thereafter the Cholesky decomposition can be used as before.

2.2 Matrix Logarithm Function

The idea is to parametrize RC_t using matrix exponential and logarithm transformation to take advantage of the fact that the matrix exponential of a real matrix is always positive definite (Hall, 2015, pp. 31).

$$RC_t = expm(A_t) = \sum_{m=0}^{\infty} \frac{1}{m!} A_t^m \tag{4}$$

The function $expm(\cdot)$ is the matrix exponential function of a symmetric matrix. A_t is a square and symmetric matrix. It can be calculated by means of the matrix logarithm function. Because RC_t is an orthonormal positive definite matrix the matrix logarithm function can be calculated with the eigenvectors U_t and eigenvalues Λ_t of the realized covariance matrix.

$$A_t = logm(RC_t) = U_t log(\Lambda_t) U_t' \tag{5}$$

After predicting $\psi_t = vech(logm(RC_t)) = vech(A_t)$ with a time series model, the matrix exponential function can be used to calculate the predicted RC_t which has to be positive definite by definition of the matrix exponential function.

2.3 Positive semi-definite Matrix Approximant

The third approach calculates a positive semi-definite approximant of the covariance matrix. The main difference from the other parametrizations is, that the method is applied after forecasting the realized covariance matrices. Because of this we model and predict $\psi_t =$

$vech(RC_t)$ without caring about semi-definite positiveness. Once the forecast of the realized covariances from a time series model is obtained, the nearest positive semi-definite matrix approximant can be calculated (Bucci et al., 2022, pp. 89). Therefore we parametrize $\widehat{RC}_t$ using the spectral decomposition into its eigenvalues $\Lambda_t = \text{diag}(\lambda_t)$ and eigenvectors U_t. As already mentioned, positive semi-definite matrices have only non-negative eigenvalues. The nearest positive semi-definite approximant S_t can be simply obtained by replacing the negative eigenvalues with zeros (Higham, 1988, p. 106).

$$S_t = U_t \text{diag}(\lambda_t^*) U_t \text{ with } \lambda_{i,t}^* = \begin{cases} \lambda_{i,t}, & \lambda_{i,t} \geq 0 \\ 0, & \lambda_{i,t} < 0 \end{cases} \tag{6}$$

2.4 Matrix correlation parametrization

The last parametrization was recently introduced by Archakov and Hansen (2021). The idea is to calculate positive definite forecasts of the realized covariance matrix from the matrix logarithm transformation of the correlation matrix.
The first step is to decompose the covariance matrix into a diagonal matrix of standard deviations D_t and a matrix of correlations C_t

$$RC_t = D_t C_t D_t \tag{7}$$

In the next step we parametrize both matrices with the matrix logarithm function.

$$\begin{aligned} d_t &= log(diag(D_t)) \\ G_t &= logm(C_t) \end{aligned} \tag{8}$$

The parameters that define RC_t are $\psi_t = (d_t, \gamma_t)$ where $\gamma(G_t)$ is a vector containing the off-diagonal elements of G_t. After predicting ψ_{t+h} reconstructing the correlation matrix is not straightforward. The problem is that we have no forecasts for the diagonal elements of G_t. A vector of diagonal elements is needed such that after applying the matrix exponential transformation all values of the diagonal are equal to 1. This is because the aim is to calculate a correlation matrix.
According to Archakov and Hansen (2021) the following theorem holds:

Theorem. *Theorem For any real symmetric matrix, $A \in \mathbb{R}^{nxn}$, there exists a unique vector, $x^* \in \mathbb{R}^n$, such that $expm(A[x^*])$ is a correlation matrix.*

To find x^* the authors propose to apply the following iterative process starting with an arbitrary vector $x_0 \in \mathbb{R}$

$$x_{t+1} = x_t - log\ diag(expm(A[x_t])) \qquad t = 0, 1, 2, ... \tag{9}$$

They found that x_t converges relatively fast to x^*. I used the code they provided in their supplemental material to implement this iterative process in Python (Archakov & Hansen, 2021, p. 12).

Finally, we are able to reconstruct the predicted correlation matrix and diagonal matrix of standard deviations and to calculate the predicted realized covariance matrix (Archakov and Hansen (2021), pp. 1701, and Bucci et al. (2022), pp. 5).

3 Forecasting Model

3.1 VAR Model

Vector Autoregressive models (VAR) of order p are an extension of the univariate Autoregressive model to dynamic multivariate time series. It can be represented by

$$y_t = v + A_1 y_{t-1} + ... + A_p y_{t-p} + u_t \tag{10}$$

with u_t $WN(0,\Omega)$. y_t is n x 1 vector of time series, v is n x 1 vector of intercepts and A_i is n x n coefficient matrix.

Following Bucci et al. (2022) I used this model to forecast the different ψ_t vectors of the parametrizations. After predicting a vector, the covariance matrix will be reconstructed. Moreover, I estimated covariance matrices with the VAR model without any restrictions or parametrization to ensure positive definitness. The results serve as a comparison to see how often a model would fail to fulfill the requirement of positivity. The model order is automatically selected in the training set by Akaike information criterion.

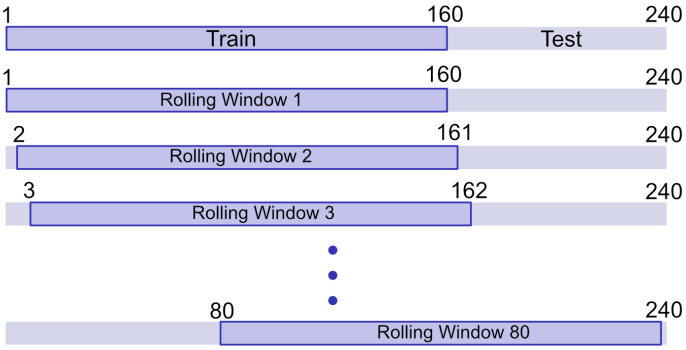
3.2 CCC-GARCH

Additional to the parametrizations used by Bucci et al. (2022) I use a multivariate GARCH model with assumptions to make sure the positivity. The constant conditional correlation model was introduced by Bollerslev (1990). Similar to Archakov and Hansen (2021), it decomposes the covariance matrix into the correlation matrix R_t and a diagonal matrix of the standard deviations D_t.

$$\Sigma_t = D_t R_t D_t \text{ with } D_t = \text{diag}(\sqrt{\sigma_{i,t}}) \tag{11}$$

The conditional variances $\sigma_{i,t}$ can be estimated with a univariate GARCH model. The covariance matrix is positive definite if and only if each of the conditional variances are well defined and the correlation matrix is positive definite. The computational complexity is large to estimate a correlation matrix with constraints to ensure a positive definite forecast for every

Figure 1: Out-of-sample forecasts with rolling window (own representation)



time period. Bollerslev (1990) reduces this complexity by assuming constant conditional correlations over time. Thus, the complex calculation has to be accomplished only once. The validity of this assumption is an empirical question.

Engle (2002) expands this model to a dynamic conditional correlation (DCC) model. Their model assumes time-varying conditional correlation and ensures positivity with another decomposition of the correlation matrix similar to the first one of the covariance matrix. The decomposed correlation matrix can be estimated with a GARCH which makes it less computationally costly. I also implemented a DCC-GARCH but it led to nearly the same results as the CCC-GARCH. Therefore I focus on only one multivariate time series model.

3.3 Forecasting with Rolling Window

Following Bucci et al. (2022) I used a rolling window to forecast. Firstly, I divided the sample into two subsamples. The training sample contains 2/3 of the observations and the test sample the remaining 1/3. Based on my data I ended up with 240 monthly realized covariance matrices. Thus, it follows that the training sample contains 160 observations and the training sample 80. The underlying data will be further explained in chapter 4. Out-of-sample forecasts are calculated from a rolling window with a fixed size of 160. The parameters are re-estimated at each step. As shown in figure 1, the first 160 observations are used to predict the 161. Then the second to the 161 observations are used to predict 162, and so on. The advantage is that the latest data is taken for the prediction. Therefore, the model is more able to make good predictions even in volatile times.
I analyze three forecast horizons: one month ahead ($h = 1$), three months ahead ($h = 3$), and one year ahead ($h = 12$).

4 Data

To compare the different parametrization methods and the multivariate time series model I use stock data from the DAX index. I choose the 15 largest stocks due to their weight in the index on the condition that they have been part of the index for the last 20 years. A list of all stocks can be found in table 4 in appendix A. The underlying stock data for the parametrization methods is data sourced from Yahoo finance between the 1st of January 2003 and the 1st of January 2023. This leads to 5100 daily prices which are used to calculate the daily returns in percentage. With these daily returns, 240 monthly realized covariance matrices are calculated using equation 1. The CCC-GARCH model is applied to monthly returns in percentage. Therefore, I use monthly price data also sourced from Yahoo finance between the 1st of December 2002 and the 1st of January 2023. The additional month is needed because returns can only be calculated from the second observation onwards. The covariance stationarity for each time series of daily and monthly returns per stock is confirmed by Augmented Dickey-Fuller (ADF), KPSS, and Phillips Perron tests for unit roots following Bucci et al. (2022). The detailed results of these tests are shown in table 4 and 5 in appendix A.

Table 1: Maximum and minimum value and corresponding stock each from return and realized variance per month and average return and average realized variance over the whole time period

	Stock	Value
Maximum return per month	Infineon	131.61%
Minimum return per month	Deutsche Boerse	−52.39%
Maximum average return	Infineon	1.65%
Minimum average return	Telekom	0.35%
Maximum variance per month	Infineon	2510.33
Minimum variance per month	Telekom	4.02
Maximum average variance	Infineon	168.31
Minimum average variance	Telekom	50.67

Table 1 shows some descriptive statistics. The statistics considered are the return and realized variance per month and the average return and average realized variance over the whole period of 20 years. For each statistic, the value of the minimum and maximum is stated together with the corresponding stock. Infineon has the largest return per month and on average. Deutsche Boerse has the smallest return per month which might be the month of the start of the financial crisis. The stock with the smallest average return is Telekom. This matches the results for the variance. Infineon has the largest variance per month and on average. Telekom, on the other hand, has the smallest variance per month and on average. Another interesting point is that the returns over 20 years are very small. This makes portfolios and trading strategies more important in order to be able to achieve larger profits.

Figure 2: Returns in percentage and realized variance for Infineon and Telekom

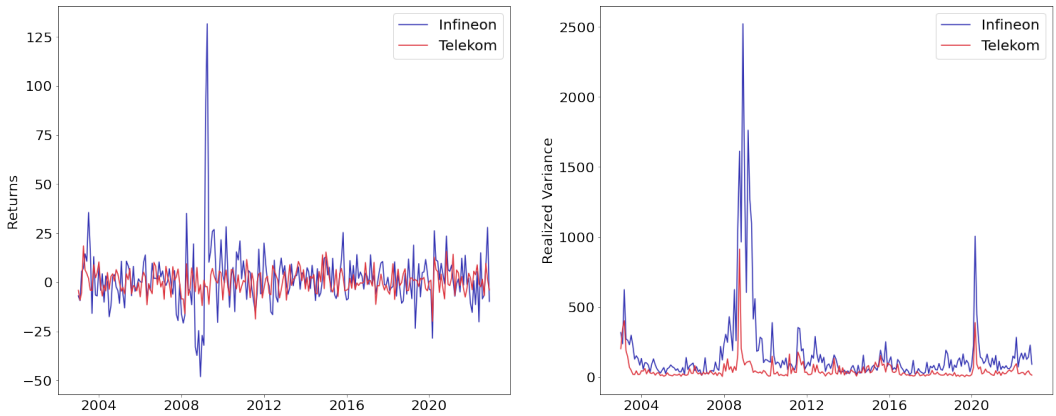


Figure 2 shows on the left side the returns and on the right side the realized variance of Infineon and Telekom. It is clearly visible in both plots that Infineon has a larger variance, especially between 2008 and 2012. The higher variance is connected with higher positive and negative returns. Moreover, clusters of higher and lower variances are recognizable. The rolling window approach should help to capture these clusters by including only the newest observations (cf. chapter 3.3).

5 Results

5.1 Procrustes and Euclidean Distance

In the following chapter, the results of the predictions with the different parametrization methods and the CCC-GARCH model are presented. To compare the different forecasts I use two different measures, the Procrustes and the Euclidean distance. The Euclidean distance is calculated with

$$d(T,F) = ||T-F|| = \sqrt{\sum_{i=1}^{n}(x_{i,T}-x_{i,F})^2} \tag{12}$$

where T is the true n x n covariance matrix and F is the predicted matrix (Bucci et al., 2022, pp. 6). The equation for the Procrustes distance is

$$d(T,F) = \inf_{\Gamma \in SO(m), \beta \in \mathbb{R}^+} ||Z_T - \beta Z_F \Gamma|| \tag{13}$$

where Z_T and Z_F are shape matrices of the true and predicted covariance matrix. Γ is the Procrustes rotation matrix from Z_F on Z_T. The Procrustes distance minimizes the Euclidean distance of the difference with the shape and rotation matrices (Dryden & Mardia, 2016, pp. 69).

Table 2 shows the results for the two measures. On the left are the results for the forecast horizon $h = 1$, in the middle are the quarterly forecasts ($h = 3$), and on the right side are the yearly forecasts ($h = 12$). For each forecast horizon, three different amounts of stocks are used, the first 5, 10, and 15 stocks. I compare the model without parametrization (RC) with the four parametrizations (Cholesky decomposition, matrix logarithm function (Logm), positive semi-definite matrix approximant (Approx) and matrix correlation parametrization (Corr)) and the CCC-GARCH model (CCC). T is the number of forecasts. Because the rolling window has a size of 160, 80 data points remain. For larger forecast horizons the number of predictions decreases because the first prediction is more steps ahead and therefore no prediction exists for the first observations in the test set. T^* states the number of times the model without any parametrization failed to predict at least a positive semi-definite covariance matrix. The bold numbers highlight the smallest distance for each combination of the forecast horizon and the number of stocks.

Table 2: Procrustes and Euclidean distance for forecast horizon $h = 1$ (left), $h = 3$ (middle) and $h = 12$ (right). T is the number of forecasts and T^* is the number of times the model without parametrization failed to predict at least a positive semi-definite matrix (following Bucci et al. (2022))

	h = 1		h = 3		h = 12	
Model	Procrustes	Euclidean	Procrustes	Euclidean	Procrustes	Euclidean
$n = 5$	$T = 80$	$T^* = 24$	$T = 78$	$T^* = 24$	$T = 69$	$T^* = 22$
RC	0.37119	261.67	0.37408	266.80	0.37623	278.17
Cholesky	0.60494	255.25	0.61223	273.22	0.63101	299.25
Logm	0.30860	**162.78**	0.31528	**178.26**	0.34433	224.94
Approx	0.37606	240.70	0.41246	245.83	0.41737	252.60
Corr	**0.30032**	165.76	**0.30763**	182.12	**0.33718**	229.25
CCC	0.35548	185.34	0.34927	190.39	0.35888	**210.12**
$n = 10$	$T = 80$	$T^* = 64$	$T = 78$	$T^* = 62$	$T = 69$	$T^* = 55$
RC	0.55618	688.73	0.56764	715.74	0.56720	777.72
Cholesky	0.78375	662.58	0.76994	703.42	0.76928	734.63
Logm	0.53659	341.64	0.53783	396.00	0.55913	459.84
Approx	0.73080	745.85	0.73528	756.09	0.70750	784.45
Corr	0.52997	**338.45**	0.52713	**368.33**	0.58108	457.06
CCC	**0.51044**	357.15	**0.51572**	374.77	**0.53298**	**419.74**
$n = 15$	$T = 80$	$T^* = 80$	$T = 78$	$T^* = 78$	$T = 69$	$T^* = 69$
RC	0.68039	1842.48	0.66303	1822.71	0.70105	1818.74
Cholesky	0.77461	1862.75	0.74799	1910.78	0.77198	1643.12
Logm	0.78224	2567.00	0.78592	2678.42	0.79582	2344.04
Approx	0.84849	1525.62	0.84862	1499.60	0.84486	1432.50
Corr	0.73883	573.15	0.73862	606.58	0.74897	704.30
CCC	**0.51272**	**522.63**	**0.52345**	**538.17**	**0.56641**	**590.08**

When including 5 stocks, the correlation parametrization of Archakov and Hansen (2021) performs best considering the Procrustes distance. Looking at the Euclidean distance the matrix logarithm parametrization or the CCC-GARCH is better. All in all, the three models are pretty close together for both distances. When using 10 stocks, the CCC-GARCH achieves the best results according to the Procrustes distance. The CCC-GARCH and the correlation parametrization are the best models considering the Euclidean distance. Again these two models and the matrix logarithm parametrization are close to each other. When including all 15 stocks, the CCC-GARCH outperforms the parametrizations. In particular, when looking at the Procrustes distance the result is pretty clear. Considering the Euclidean distance the correlation parametrization is closest to the CCC-GARCH. Moreover, the table shows that the larger the number of stocks, the more often the model without parametrization is not able to predict at least a positive semi-definite covariance matrix. Starting with slightly more than a quarter of the forecasts, if 5 stocks are used, it fails every time, when all 15 stocks are included.

These results are partially consistent with those of Bucci et al. (2022). They used simulated data from a stochastic volatility model and a DCC model and real-world data from the largest US-stocks. They also made predictions for 5, 10, and 15 stocks in their simulation studies. Moreover, they used intraday data to calculate daily realized covariance matrices and make predictions on a daily (h = 1), weekly (h = 5), and monthly basis (h = 22). Additionally, they calculated daily and weekly realized covariance matrices with intraday data of the 10 largest US-stocks and forecast for the same forecast horizons as in the simulation study. They only compared the four different parametrization methods.
They found in the simulation studies that in almost all cases the correlation matrix parametrization performs best according to Euclidean and Procrustes distance. In some cases looking at Euclidean distance, the Cholesky decomposition is the best. For the real-world data, the results were similar. According to my results, the correlation parametrization has also often the best performance when the CCC-GARCH is disregarded. But in contrast to their paper, the Cholesky decomposition does not perform well in my analysis. Furthermore, they found that modeling the realized covariance matrix directly or using the matrix approximant after modeling produces the poorest predictions. This might be due to the predictive errors made in non-positive semidefinite observations. Based on my analysis I agree with them because the performance of the matrix approximant is always one of the worst. A difference is that the model without parametrization has not failed that often in their study to predict an at least positive semi-definite matrix. This might be the case because they selected the number of lags in the VAR model with the analysis of different information criteria and I only let Python choose the lag based on the Akaike information criterion. Moreover, they had much more observations to train the model. All in all, my results coincide to a large extent with theirs.

5.2 Global Minimum Variance Portfolio

In addition to the distance measures, I also consider the global minimum variance portfolios. Portfolios are used because the combination of different stocks reduces the variance compared to only one stock. This diversification can result in portfolios with lower variance than the lowest variance of a single stock in the portfolio. The global minimum variance portfolio is the combination of stocks that leads to the lowest variance of all possible combinations of these stocks. The weights ω of the stocks for this portfolio can be obtained by solving the minimization problem

$$\min_{\omega} \ \omega'\Sigma\omega \ \ s.t. \ \ \omega'\mathbf{1} = 1, \ \ 0 < \omega_i < 0.5 \tag{14}$$

where ω is a n x 1 vector of weights, Σ is the n x n covariance matrix and n is the number of stocks in the portfolio. For the choice of the constraints, I followed Callot et al. (2017) approach. $\omega'\mathbf{1} = 1$ means that the portfolio weights add up to 1. The weights have to be smaller than 0.5 such that no stock represents more than 50% of the portfolio to ensure diversification. Moreover, they have to be positive because a negative weight means to short-sell a stock which is not a common practice for most investors (Chan et al., 1999, p. 17).
I calculated the weights using the one-step-ahead forecasts of the covariance matrices for 5, 10, and 15 stocks. Furthermore, I updated the portfolio only once per quarter. As a benchmark, I added the optimal portfolio if the true monthly covariance is known. Because the model without parametrization often fails to predict a correct covariance matrix I dropped this model. To compare the different portfolios I use five different statistics. First I analyze the average and accumulated return in percentage and the average variance. For a better understanding of the variance, I compare it with the average variance of the stock with the lowest average variance. This stock is Telekom (cf. chapter 4). It represents the case that the investor only invests in stocks from Telekom. The difference is given in the percentage change compared to the average variance of the Telekom stock. The fifth statistic is the Sharpe ratio which is the ratio between average return and average standard deviation. The portfolio is better, the larger the Sharpe ratio, as it provides higher ratios of return over risk (Callot et al., 2017, pp. 155). The bold numbers mark the highest returns, Sharpe ratio, and variance reduction and the lowest variance of all portfolios based on the different models, not considering the portfolio based on the true covariance matrix.

Table 3 shows the results for the different optimal portfolios. The main point is that with a portfolio the average variance can be reduced by between 20% and 70%. High reductions over 45% are only possible if the true covariance matrix is known. Considering the returns, for 5 stocks no parametrization method or the CCC-GARCH are able to beat the true model. But all parametrization methods deliver a higher average return than the Telekom stock (0.35% cf. Table 1). Therefore, buying portfolios based on the different covariance matrices reduces

Table 3: Different statistics of the optimal portfolios updated once per quarter based on the one-step-ahead forecasts for the different parametrization methods for 5, 10, and 15 stocks

Model	Average Return	Accumulated Return	Variance	%-change Variance Telekom	Sharpe Ratio
n = 5					
True	0.6180	1.4111	23.0695	−54.4746	0.1287
Cholesky	**0.5378**	**1.3590**	31.2588	−38.3138	**0.0962**
Logm	0.3804	1.1892	**30.1829**	**−40.4371**	0.0692
Approx	0.3734	1.1582	36.1830	−28.5964	0.0621
Corr	0.4181	1.1967	30.1900	−40.4231	0.0761
CCC	0.3059	1.1230	32.1067	−36.6407	0.0540
n = 10					
True	−0.2255	0.7052	20.3567	−59.8282	−0.0500
Cholesky	0.5885	1.2547	39.5996	−21.8541	0.0935
Logm	0.5841	1.3301	**32.1036**	**−36.6468**	0.1031
Approx	**1.1310**	**1.8305**	40.0876	−20.8911	**0.1786**
Corr	0.5382	1.3424	33.3249	−34.2365	0.0932
CCC	0.3609	1.1225	32.5487	−35.7684	0.0633
n = 15					
True	0.4274	1.2179	15.9708	−68.4832	0.1069
Cholesky	**1.0402**	**1.9217**	35.0293	−30.8731	**0.1758**
Logm	0.6201	1.3644	32.0632	−36.7264	0.1095
Approx	0.8819	1.6023	39.3511	−22.3445	0.1406
Corr	0.5399	1.3088	35.7066	−29.5367	0.0904
CCC	0.7033	1.4874	**27.5986**	**−45.5369**	0.1339

the variance and leads to higher returns than buying only one stock. For more stocks, all parametrization methods and the CCC-GARCH provide higher returns than the true optimal portfolio and the Telekom stock.

The parametrization method using the matrix approximant has always the smallest variance reduction compared to the Telekom stock but one of the highest returns and Sharpe ratio. Looking only at the variance reduction this result is consistent with the results in chapter 5.1 and the results of Bucci et al. (2022) where the matrix approximant is one of the worst models.

The model with the highest average and accumulated return always has the highest Sharpe ratio. For 5 and 15 stocks this is the Cholesky decomposition and for 10 stocks it is the parametrization method using the matrix approximant. The largest reduction in variance achieves the matrix logarithm parametrization using 5 and 10 stocks and the CCC-GARCH with 15 stocks. The correlation matrix parametrization, matrix logarithm parametrization, and the CCC-GARCH are close to each other considering the variance reduction, especially for 5 and 10 stocks. This matches the results in chapter 5.1 where these three models always

perform best. It is interesting to see that compared to the distance measures the Cholesky decomposition performs much better. The analysis of the portfolios does not reveal that only one model performs best.

Additionally, I calculated the statistics for portfolios which I updated every month, every half year, and every year. None of the frequencies produce a clear result. The results are shown in tables 6 to 8 in appendix B. All have in common that when including 15 stocks the CCC-GARCH model achieves the largest variance reduction. This result also matches the results in chapter 5.1 where the CCC-GARCH performs best for high dimensions. The Cholesky decomposition delivers the highest returns. In the other cases, no clear patterns are discernible. But often all models are relatively close together.

All in all, the analysis of the optimal portfolios does not give a clear answer which parametrization method or model is best to use. No model outperforms the other all the time. The CCC-GARCH works well in high dimensions. But summarized all of them produce good results. Therefore the important point is to use a parametrization method or constraint model at all. The choice of the method might then depend on other criteria such as the computational costs or the interpretability.

6 Conclusion

Not only in portfolio management but also in risk management or asset pricing predicting covariance matrices is a major task. The main problem is to ensure that the predicted covariance matrices are at least positive semi-definite.

In this paper, I have discussed the performance of four different parametrization methods and one multivariate time series model. I have looked at two distance measures and minimum variance portfolios. According to the distance measures, the matrix logarithm transformation, the correlation parametrization proposed by Archakov and Hansen (2021), and the CCC-GARCH performs best. Especially in high dimensions, the CCC-GARCH outperforms the parametrization methods. This might be the case because the estimated VARs for the parametrization methods involve a large number of dependent variables which is larger the more stocks are included. Researchers typically use prior shrinkage on the coefficients to deal with possible problems. This could be an interesting point to consider in future analysis (Bucci et al., 2022, p. 14).
According to the distance measure, the two parametrization methods matrix logarithm transformation and correlation parametrization and the CCC-GARCH perform best. Their distances are close together. Another important point is that the VAR model without any parametrizations failed pretty often to estimate an at least positive semi-definite matrix. Thus, the

importance of dealing with positivity becomes clear.
Looking at the global minimum portfolios based on the different forecasts no model outperforms the others all the time. All portfolios reach a large variance reduction compared to a single stock. The magnitude of the reduction matches quite well the results of the distance measures. The two parametrization methods mentioned above and the CCC-GARCH achieve often the largest variance reductions. The matrix approximant leads to the lowest reduction in the average variance. Looking at the portfolio returns the result is not clear. All approaches are able to deliver a larger average return than the lowest average return of all included stocks. It is interesting to see that the Cholesky decomposition has always one of the largest returns even if it does not perform pretty well according to the distance measures.

Summarized the CCC-GARCH, the correlation parametrization or the matrix logarithm transformation seem to be a good choice. But because the result is not clear and the differences are often small, other aspects can be considered when choosing a parametrization method or multivariate time series model. The important point is to use some kind of parametrization method or constrained model because otherwise non-positive definite covariance matrices are estimated. Especially in high dimensions, the computational costs are a crucial point. The matrix logarithm function or Cholesky decomposition are easy to implement and have low computational costs. In contrast, the correlation parametrization or the CCC-GARCH are more costly. For other multivariate time series models like the DCC-GARCH this problem can be even larger.
To this time not many researchers compared different parametrization methods and multivariate models. More analysis with different real-world or simulated data and different accuracy measures might be necessary to have a clearer result on which parametrization method or multivariate time series model to use best. In addition, either the model for estimating the parameterized matrices could be trained more carefully or another model could be used. Moreover, this is a relatively new field of research that has attracted increasing interest in recent years. This leads to new developments such as new matrix parametrization methods as seen in the approach of Archakov and Hansen (2021) or new multivariate models which could be considered in future works.

A Appendix: Results of the Stationarity Tests

Following Bucci et al. (2022), I used four different tests to test the time series of prices for stationarity. The Augmented Dickey-Fuller (ADF) test, the KPSS test, and the Phillips Perron test. The KPSS test has stationarity as the null hypothesis. For this reason, large p-values are preferred so the null hypothesis can not be rejected (Kwiatkowski et al., 1992, p. 3). Python only returns a boundary point of 0.1 if the test statistic is outside the table of critical values, that is if the p-value is outside the interval (0.01, 0.1). In this cases, the null hypothesis of stationarity can not be rejected. With the other tests, it is the other way around. Here, small p-values are preferred so that the null hypothesis of non-stationarity can be rejected.
In table 4 the test statistics and p-values of each test for the daily returns are provided. Looking at the ADF and Phillips Perron tests the null hypothesis of non-stationarity can be rejected for every stock at every common significance level. Moreover, the time series of the stocks are stationary according to the KPSS test. All in all, there is evidence to conclude that all time series of daily returns are stationary.

Table 4: Test statistic and p-value of the three stationarity tests ADF, KPSS and Phillips Perron test applied on the time series of the daily returns for each stock

Stock	Augmented Dickey-Fuller		KPSS		Phillips Perron	
	Test Statistic	p-value	Test Statistic	p-value	Test Statistic	p-value
Allianz	-13.9903	4.03e-26	0.0218	0.1000	-69.9840	0.0000
SAP	-15.0525	9.25e-28	0.0864	0.1000	-72.2748	0.0000
Siemens	-18.0905	2.57e-30	0.0438	0.1000	-72.2076	0.0000
Telekom	-20.4206	0.0000	0.0600	0.1000	-75.5034	0.0000
Mercedes	-12.7920	7.04e-24	0.0358	0.1000	-68.4504	0.0000
Bayer	-19.0585	0.0000	0.3127	0.1000	-74.1997	0.0000
BASF	-14.1816	1.92e-26	0.2690	0.1000	-71.8758	0.0000
Muenchner Rueck	-25.9713	0.0000	0.0562	0.1000	-70.6503	0.0000
Infineon	-12.0093	3.21e-22	0.0407	0.1000	-68.5031	0.0000
Deutsche Post	-20.7917	0.0000	0.0639	0.1000	-70.6604	0.0000
Deutsche Boerse	-23.0534	0.0000	0.0599	0.1000	-72.4460	0.0000
RWE	-20.1522	0.0000	0.1814	0.1000	-70.5895	0.0000
VW	-19.1410	0.0000	0.1599	0.1000	-67.9290	0.0000
BMW	-13.4894	3.13e-25	0.0624	0.1000	-69.7622	0.0000
Merck	-73.9207	0.0000	0.0958	0.1000	-74.2042	0.0000

Tabel 5 shows the test statistics and p-values of each test for the monthly returns. Considering the ADF and Phillips Perron tests, the null hypothesis of non-stationarity can also be rejected at every common significance level. Looking at the KPSS test the time series are as well stationary. Only for the stock of Bayer the null hypothesis of stationarity can be rejected at a 10%-significance level. Summarized, there is also evidence to conclude stationarity.

Table 5: Test statistic and p-value of the three stationarity tests ADF, KPSS and Phillips Perron test applied on the time series of the monthly returns for each stock

Stock	Augmented Dickey-Fuller		KPSS		Phillips-Perron	
	Test Statistic	p-value	Test Statistic	p-value	Test Statistic	p-value
Allianz	-7.6696	1.61e-11	0.0477	0.1000	-17.2438	6.08e-30
SAP	-8.7176	3.45e-14	0.1443	0.1000	-17.5390	4.20e-30
Siemens	-15.2206	5.49e-28	0.0615	0.1000	-15.3545	3.68e-28
Telekom	-12.6105	1.66e-23	0.1275	0.1000	-16.8055	1.21e-29
Mercedes	-15.1494	6.83e-28	0.0414	0.1000	-15.1484	6.85e-28
Bayer	-15.8363	9.81e-29	0.3801	0.0857	-15.9142	8.06e-29
BASF	-14.3449	1.04e-26	0.3132	0.1000	-14.3270	1.12e-26
Muenchner Rueck	-7.8019	7.45e-12	0.0930	0.1000	-22.7523	0.0000
Infineon	-8.1127	1.21e-12	0.0388	0.1000	-11.2129	2.10e-20
Deutsche Post	-6.6121	6.34e-09	0.0669	0.1000	-15.0892	8.24e-28
Deutsche Boerse	-9.5341	2.84e-16	0.0832	0.1000	-14.6974	2.98e-27
RWE	-15.7435	1.25e-28	0.1920	0.1000	-15.7479	1.23e-28
VW	-11.7415	1.27e-21	0.2522	0.1000	-15.4835	2.54e-28
BMW	-8.8967	1.20e-14	0.0692	0.1000	-16.3750	2.79e-29
Merck	-16.5850	1.82e-29	0.1182	0.1000	-16.6140	1.72e-29

B Appendix: Results of the Portfolio Analysis

Table 6: Different statistics of the optimal portfolios updated every month based on the one-step-ahead forecasts for the different parametrization methods for 5, 10, and 15 stocks

Model	Average Return	Accumulated Return	Variance	%-change Variance Telekom	Sharpe Ratio
n = 5					
True	0.6180	1.4111	23.0695	−54.4746	0.1287
Cholesky	0.4394	1.2530	29.1467	−42.4820	0.0814
Logm	**0.5178**	**1.3283**	28.0646	−44.6172	**0.0977**
Approx	0.3087	1.0980	42.0586	−17.0015	0.0476
Corr	0.3171	1.1203	**29.0564**	**−42.6601**	0.0588
CCC	0.3425	1.1670	29.7484	−41.2944	0.0628
n = 10					
True	−0.2255	0.7052	20.3567	−59.8282	−0.0500
Cholesky	0.3342	1.0532	34.9169	−31.0950	0.0566
Logm	0.6986	1.4632	**30.2204**	**−40.3631**	0.1271
Approx	**1.1885**	**1.9615**	42.0134	−17.0907	**0.1834**
Corr	0.1971	1.0301	31.0083	−38.8082	0.0354
CCC	0.2943	1.0699	31.4037	−38.0280	0.0525
n = 15					
True	0.4274	1.2179	15.9708	−68.4832	0.1069
Cholesky	**0.9541**	**1.7977**	33.8994	−33.1030	**0.1639**
Logm	0.1535	0.9664	34.1595	−32.5896	0.0263
Approx	0.7241	1.4294	40.4106	−20.2538	0.1139
Corr	−0.0883	0.7927	36.7932	−27.3924	−0.0146
CCC	0.6727	1.4576	**26.9899**	**−46.7381**	0.1295

Table 7: Different statistics of the optimal portfolios updated every 6 months based on the one-step-ahead forecasts for the different parametrization methods for 5, 10, and 15 stocks

Model	Average Return	Accumulated Return	Variance	%-change Variance Telekom	Sharpe Ratio
n = 5					
True	0.6180	1.4111	23.0695	−54.4746	0.1287
Cholesky	**0.6526**	**1.4778**	31.7632	−37.3185	**0.1158**
Logm	0.4628	1.2352	**30.6760**	**−39.4640**	0.0836
Approx	0.4375	1.2113	36.2181	−28.5272	0.0727
Corr	0.4899	1.2262	31.3435	−38.1467	0.0875
CCC	0.3852	1.1746	31.1920	−38.4456	0.0690
n = 10					
True	−0.2255	0.7052	20.3567	−59.8282	−0.0500
Cholesky	0.4323	1.1050	35.0419	−30.8483	0.0730
Logm	0.5536	1.3211	**33.0966**	**−34.6872**	0.0962
Approx	0.7869	1.4082	37.0879	−26.8108	0.1292
Corr	**0.7888**	**1.5725**	37.3251	−26.3426	**0.1291**
CCC	0.3951	1.1371	33.2063	−34.4706	0.0686
n = 15					
True	0.4274	1.2179	15.9708	−68.4832	0.1069
Cholesky	**1.2499**	**2.3385**	36.2945	−28.3764	**0.2075**
Logm	0.5533	1.2670	33.2000	−34.4831	0.0960
Approx	0.6080	1.3205	34.6202	−31.6806	0.1033
Corr	0.5674	1.3774	36.7027	−27.5710	0.0937
CCC	0.6900	1.4531	**27.7626**	**−45.2133**	0.1310

Table 8: Different statistics of the optimal portfolios updated once per year based on the one-step-ahead forecasts for the different parametrization methods for 5, 10, and 15 stocks

Model	Average Return	Accumulated Return	Variance	%-change Variance Telekom	Sharpe Ratio
n = 5					
True	0.6180	1.4111	23.0695	−54.4746	0.1287
Cholesky	0.4490	1.2661	**28.7076**	**−43.3485**	0.0838
Logm	0.4445	1.2413	31.5555	−37.7284	0.0791
Approx	**0.5713**	**1.2637**	32.6284	−35.6111	**0.1000**
Corr	0.4084	1.1703	31.5450	−37.7490	0.0727
CCC	0.3911	1.2023	30.6784	−39.4593	0.0706
n = 10					
True	−0.2255	0.7052	20.3567	−59.8282	−0.0500
Cholesky	0.5201	1.1533	36.2216	−28.5202	0.0864
Logm	0.5207	1.2818	32.6862	−35.4970	0.0911
Approx	**1.0518**	**1.7627**	40.6306	−19.8196	**0.1650**
Corr	0.8206	1.5880	38.3564	−24.3075	0.1325
CCC	0.3081	1.0721	**32.6580**	**−35.5527**	0.0539
n = 15					
True	0.4274	1.2179	15.9708	−68.4832	0.1069
Cholesky	**1.1824**	**2.1996**	36.7153	−27.5461	**0.1951**
Logm	0.6765	1.4177	31.2540	−38.3233	0.1210
Approx	0.8089	1.5493	34.7272	−31.4693	0.1373
Corr	0.6843	1.4860	44.3117	−12.5553	0.1028
CCC	0.8647	1.6783	**27.8736**	**−44.9943**	0.1638

References

Archakov, I., & Hansen, P. R. (2021). A New Parametrization of Correlation Matrices. *Econometrica*, *89*(4), 1699–1715. https://doi.org/10.3982/ECTA16910

Bollerslev, T. (1990). Modeling the Coherence in Short-Run Nominal Exchange Rates: A Multivariate Generalized Arch Model. *The Review of Economics and Statistics*, *72*(3), 498–505. https://doi.org/10.2307/2109358

Bucci, A., Ippoliti, L., & Valentini, P. (2022). Comparing unconstrained parametrization methods for return covariance matrix prediction. *Statistics and Computing*, *32*(5), 90. https://doi.org/10.1007/s11222-022-10157-4

Callot, L. A. F., Kock, A. B., & Medeiros, M. C. (2017). Modeling and Forecasting Large Realized Covariance Matrices and Portfolio Choice. *Journal of Applied Econometrics*, *32*(1), 140–158. https://doi.org/10.1002/jae.2512

Chan, L. K. C., Karceski, J., & Lakonishok, J. (1999). On Portfolio Optimization: Forecasting Covariances and Choosing the Risk Model. *The Review of Financial Studies*, *12*(5). https://doi.org/10.1093/rfs/12.5.937

Dryden, I. L., & Mardia, K. V. (2016). *Statistical Shape Analysis, with Applications in R* (1st ed.). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781119072492

Engle, R. (2002). Dynamic Conditional Correlation. *Journal of Business & Economic Statistics*(20). https://doi.org/10.1198/073500102288618487

Hall, B. C. (2015). *Lie Groups, Lie Algebras, and Representations: An Elementary Introduction* (Vol. 222). Springer International Publishing. https://doi.org/10.1007/978-3-319-13467-3

Higham, N. J. (1988). Computing a nearest symmetric positive semidefinite matrix. *Linear Algebra and its Applications*, *103*, 103–118. https://doi.org/10.1016/0024-3795(88)90223-6

Horn, R. A., & Johnson, C. R. (2012). *Matrix analysis* (2nd ed). Cambridge University Press. https://doi.org/10.1017/CBO9781139020411

Kwiatkowski, D., Phillips, P. C. B., & Schmidt, P. (1992). Testing the Null Hypothesis of Stationarity Against the Alternative of a Unit Root: How Sure Are We That Economic Time Series Have a Unit Root? *Journal of Econometrics*, *54*(1-3), 159–178. https://doi.org/10.1016/0304-4076(92)90104-Y

Markowitz, H. (1952). Portfolio Selection. *The Journal of Finance*, *7*(1), 77–91. https://doi.org/10.1111/j.1540-6261.1952.tb01525.x

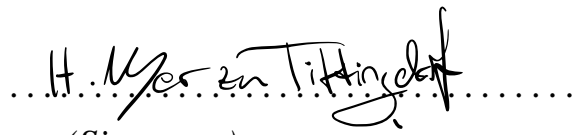
Pinheiro, J. C., & Bates, D. M. (1996). Unconstrained parametrizations for variance-covariance matrices. *Statistics and Computing*, *6*(3), 289–296. https://doi.org/10.1007/BF00140873

Schnabel, R. B., & Eskow, E. (1990). A New Modified Cholesky Factorization. *SIAM Journal on Scientific and Statistical Computing*, *11*(6), 1136–1158. https://doi.org/10.1137/0911064

Declaration of Orginality

I confirm that the submitted seminar paper is original work and was written by me without further assistance. Appropriate credit has been given where reference has been made to the work of others. The thesis was not examined before, nor has it been published.

Berlin, 09.03.2023



(Signature)