



Comparing unconstrained parametrization methods for return covariance matrix prediction

Andrea Bucci¹ · Luigi Ippoliti¹ · Pasquale Valentini¹

Received: 6 July 2022 / Accepted: 21 September 2022
© The Author(s) 2022

Abstract

Forecasting covariance matrices is a difficult task in many research fields since the predicted matrices should be at least positive semidefinite. This problem can be overcome by including constraints in the predictive model or through a parametrization of the matrices to be predicted. In this paper, we focus on the latter approach in a financial application and analyse four parametrizations of the covariance matrices of asset returns. The aim of the manuscript is to understand if the parametrizations of the covariance matrices exhibit differences in terms of predictive accuracy. To this end, we critically analyse their predictive performance through both a Monte Carlo simulation and an empirical application with daily and weekly realized covariance matrices of stock assets. Our findings highlight that the Cholesky decomposition and the parametrization recently introduced by Archakov and Hansen are the overall best-performing methods in terms of forecasting accuracy.

Keywords Volatility forecasting · Covariance model · Realized volatility · Parametrization

1 Introduction

The estimation and forecasting of covariance matrices are common problems in many fields, such as finance, biostatistics, signal processing and geostatistics. This is a difficult task as covariance matrices lay on a curved manifold (Marron and Dryden 2021; Han and Park 2022). This means that an estimation setting of covariance matrices must account for the fact that the resulting matrices must be at least positive semidefinite (PSD).

In finance, estimating and predicting covariance matrices are of major interest for the crucial role of asset returns covariance matrices in portfolio optimization. Several methods aim at predicting returns conditional volatility in the multivariate framework while guaranteeing positive semidefiniteness of the covariance matrix. For example, Bollerslev

(1990) and Engle (2002) provide a multivariate version of autoregressive conditional heteroskedasticity models by imposing different processes for the variances and covariances in the Dynamic Conditional Correlation (DCC) model. Subsequently, a non-parametric measure of volatility, proposed by Andersen et al. (2001, 2003), Barndorff-Nielsen and Shephard (2002b, a), approximates the volatility of stock returns through high-frequency data and, in the multivariate framework, guarantees a positive definite and symmetric estimation of the real covariance matrix. The most attractive feature of realized volatility is that any time series model can be applied to estimate its relationship with explanatory variables and make predictions. Nevertheless, predicted matrices must still be positive semidefinite.

Generally, two approaches can be performed to ensure that predicted realized covariance matrices are symmetric and at least positive semidefinite: a constrained optimization in the predictive model or matrix parametrization. As further pointed out in Pinheiro and Bates (1996), the former approach usually leads to sub-optimal solutions and makes the statistical properties of the constrained predictions difficult to characterize. Therefore, most of the literature on the prediction of realized covariances has focused on matrix parametrizations and, consequently, on modelling the parameters obtained by these transformations. However, the majority of these studies applies a single method without

✉ Andrea Bucci
andrea.bucci@unich.it

Luigi Ippoliti
luigi.ippoliti@unich.it

Pasquale Valentini
pasquale.valentini@unich.it

¹ Department of Economics, Università degli Studi G. d'Annunzio Chieti-Pescara, Viale Pindaro 52, 65127 Pescara, Italy

discussing possible differences in the predictive accuracy due to the choice of the parametrization. This paper aims to fill the gap in the literature by providing a critical analysis of the unconstrained methods in a predictive setting both through simulation and empirical application, highlighting the strengths and weaknesses of the parametrizations.

There are several ways to parametrize the covariance matrix, each of which has distinctive properties. For example, Halbleib-Chiriac and Voev (2011) use the Cholesky decomposition to guarantee that predicted matrices through the VARFIMA model are positive definite. The same happens in Christiansen et al. (2012), C.Čech and Baruník, (2017) and in the application of neural networks to realized covariances in Bucci (2020). Conversely, Bauer and Vorkink (2011) and Asai et al. (2022) propose using the matrix logarithm function transformation of the realized covariances and using the exponential matrix function to guarantee positive definite predictions. As shown in Heiden (2015), both methods present some pitfalls that should be accounted for in the predictive setting. For instance, in the Cholesky decomposition, the ordering of elements in the original matrix yields different decompositions. Moreover, both models lack the interpretability of parameters estimated in the predicting model and may produce biased forecasts due to the nonlinear transformation of the covariance matrix.

For these reasons, the forecasting accuracy of a model for predicting realized covariance strongly depends on the parametrization of the realized covariance matrix. Since the predictive model is usually applied to the parameters of the covariance matrix transformation, any pitfall in these approaches may lead to biased forecasts and consequent poor predictive accuracy. The goal of this paper is then to investigate if there are differences in the predictive accuracy of the covariance matrix by comparing four different parametrizations. Specifically, we parametrize the realized covariance matrices, we model the related parameters and we apply the inverse transform to obtain positive semidefinite matrices (most of the methods compared here guarantee also the positive definiteness). As suggested by Pinheiro and Bates (1996), we apply Cholesky decomposition and matrix logarithm transformation to the covariance matrix. Both approaches can be easily implemented and guarantee symmetric positive definite predicted matrices. We also implement a positive semidefinite matrix approximation like the one introduced in Higham (1988) whose application for realized covariance matrices has been proposed in Barndorff-Nielsen and Shephard (2004) and implemented in Fan et al. (2012). This allows computing the nearest symmetric positive semidefinite matrix in the Frobenius norm to the estimated/predicted covariance matrix. The main advantage of this approach is that the time series model can be directly applied to the realized covariances since their predictions are *a posteriori* approximated to a positive semidefinite matrix.

Thus, the interpretability of coefficients in time series models remains untouched, as well as the unbiasedness of predictions for not-approximated positive semidefinite matrices. Finally, we use the parametrization recently suggested by Archakov and Hansen (2021). This approach is based on the matrix logarithmic transformation of the correlation matrix, offering a new tool for the regularization of large covariance matrices by imposing a structure on the off-diagonal elements. To the best of our knowledge, this is the first study that presents a comparison of parametrizations in the accuracy of the predicted covariance matrices. Moreover, the parametrization proposed by Archakov and Hansen (2021) has been here applied for the first time for predicting asset returns realized covariances.

To understand possible differences in the predictions obtained from these parametrizations, we analyse the forecasting accuracy by using both simulated and real data. Following Andersen et al. (2003) and Halbleib-Chiriac and Voev (2011), all the predictions are produced from a rather generic model. In particular, we fit a Vector Autoregressive (VAR) model either to the (unique) parameters of the parametrizations or, as in the case of the positive semidefinite matrix approximant, directly to the realized covariances. The accuracy of the predictions is then evaluated by applying robust multivariate loss functions (Pástor and Veronesi 2012) in a Model Confidence Set procedure (see Hansen et al. 2011). In this context, we also propose the use of a Procrustes Covariance Distance which, as discussed in Marron and Dryden (2021), is less prone to swelling than Frobenius distance and can be applied to non-full-rank covariance matrices like realized covariance matrices in high-dimensional allocation problems (Reiss and Winkelmann 2021).

The remainder of the paper is organized as follows. In Sect. 2, we introduce the realized covariance matrix and discuss the different parametrizations used to obtain positive semidefinite predictions. The parametrization predictive ability is evaluated by a simulation study in Sect. 3 where the data are simulated both from a stochastic volatility and a DCC model. An application with real data is then presented in Sect. 4 while Sect. 5 concludes the paper with a discussion on further developments.

2 Realized covariance matrix parametrization

The covariance matrix of asset returns plays a key role in several financial applications, such as portfolio optimization, risk management and option pricing. Throughout the last three decades, several approaches have been proposed to estimate such quantity (Engle and Kroner 1995; Engle 2002). Many of them (Andersen et al. 2003; Barndorff-Nielsen and Shephard 2004) are built upon the notion of a measure of

the latent volatility, *i.e.* the realized covariance, which 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 (intra-day) returns sampled at higher frequencies $\tau = 1, \dots, N_t$, where N_t is the number of high-frequency observations in the t -th reference period.

From the theoretical point of view, it is known that RC_t is an unconditional non-parametric estimator of the (ex-post) quadratic covariation matrix on a reference period t . Nonetheless, RC_t measures are also useful in gauging the conditional covariance matrix. In fact, as shown in Barndorff-Nielsen and Shephard (2002a) and Andersen et al. (2001), using high-frequency data, one can also obtain an estimate of the matrix of quadratic covariation that differs from the true conditional covariance matrix by a zero-mean error.

Since RC_t changes over time, describing and predicting its temporal dynamics is fundamental to financial decision-making. Nevertheless, the predicted covariance matrices should be guaranteed at least positive semidefinite, making the predictive setting still difficult. To overcome this problem, several parametrizations of the covariance matrices have been proposed (see Halbleib-Chiriac and Voev 2011; Bauer and Vorkink 2011).

In order to fully understand the methods here implemented, we need to specify some useful notation. The operator $\text{vech}(\cdot)$ denotes the vectorization operator of the $\tilde{n} = n(n + 1)/2$ lower elements of an $n \times n$ symmetric matrix. As in Archakov and Hansen (2021), we use the operator $\text{vecl}(\cdot)$ to denote the vectorization operator of the lower off-diagonal elements of the argument. The operator $\text{diag}(\cdot)$ works in a twofold way. On the one hand, when applied to an n -dimensional vector it denotes an $n \times n$ diagonal matrix whose diagonal entries are the elements of the vector. On the other hand, when applied to an $n \times n$ square matrix, it extracts the diagonal elements of it and stacks them in an $n \times 1$ vector. The operator $\text{expm}(\cdot)$ returns the matrix exponential function of a symmetric matrix, as further described in Sect. 2.2, while $\text{logm}(\cdot)$ gives the matrix logarithm transformation function of a square and symmetric matrix.

Since the $n \times n$ covariance matrix, RC_t , is symmetric, only \tilde{n} parameters are necessary for its representation. By denoting with ψ_t the set of parameters at time t that identifies RC_t , a general parametrization of RC_t can be written as

$$RC_t = L_t(\psi_t)' L_t(\psi_t) \tag{2}$$

where $L_t(\psi_t)$ is an $n \times n$ full rank matrix derived from the \tilde{n} -dimensional vector of unconstrained parameters ψ_t . The

choice of $L_t(\psi_t)$ in Eq. (2) leads to different parametrizations of RC_t . In general, the parametrization should guarantee that:

1. any nonsingular covariance matrix, Σ_t , maps to a unique vector $\psi_t = v(\Sigma_t) \in \mathbb{R}^{\tilde{n}}$;
2. any vector, $\psi_t \in \mathbb{R}^{\tilde{n}}$, maps to a unique covariance matrix $\Sigma_t = v^{-1}(\psi_t)$; therefore, the parametrization can be inverted and this inversion is unique;
3. the parametrization, $\psi_t = v(\Sigma_t)$, is ‘invariant’ to the ordering of the columns that define Σ_t ;
4. the elements of the parameters ψ_t can be easily interpreted.

Here, we consider three classes of unconstrained parametrizations that respect most of the former properties, one relying on the Cholesky decomposition of RC_t , the other based on its spectral decomposition, and the third based on the matrix logarithm transformation of the correlation matrix.

2.1 Cholesky decomposition

A computationally simple way to parametrize RC_t is supposing that the parameters of ψ_t in Eq. (2) are the elements of an upper triangular matrix, such that

$$RC_t = L_t' L_t \tag{3}$$

is the Cholesky decomposition of RC_t . The elements of L_t , $l_{ij,t}$, can be calculated recursively by

$$l_{ij,t} = \begin{cases} \frac{1}{l_{jj,t}} \left(c_{ij,t} - \sum_{k=1}^{j-1} l_{ik,t} l_{jk,t} \right), & i > j \\ \sqrt{c_{ij,t} - \sum_{k=1}^{j-1} l_{jk,t}^2}, & i = j \end{cases} \tag{4}$$

where $c_{ij,t}$ is the ij -th element of RC_t . Once obtained the h -steps-ahead predictions of L_t , *i.e.* \hat{L}_{t+h} , through a given time series model, the predicted RC_{t+h} is obtained by inverting the Cholesky decomposition in Eq. (3), $\widehat{RC}_{t+h} = \hat{L}_{t+h}' \hat{L}_{t+h}$.

The major pitfall of this parametrization, already assessed in Pinheiro and Bates (1996) and Heiden (2015), is that it is not invariant to the ordering of variables, thus violating the aforementioned property 3. In fact, since the calculation of Cholesky factors is recursive, the ordering of the elements in the original matrix, RC_t , influences the definition of the factors. This means that up to 2^n different set of parameters, ψ_t , may represent the same RC_t . In terms of predictions, different results may arise from the different ordering, especially when the dimension of RC_t is high.

The Cholesky decomposition does not even respect property 4, since, except for the first Cholesky factor, it is difficult to interpret the relationship between the parameters and elements of RC_t . For example, this lack of a direct relationship

between ψ_t and the elements of RC_t makes it difficult to interpret the coefficients in a time series model.

Furthermore, the prediction errors on the Cholesky factors quadratically propagate to the forecasts of the original elements of RC_t , making the predictions possibly biased, as already shown in Halbleib-Chiriatic and Voev (2011). In fact, the forecasts of the parameters ψ_t from a Vector Autoregressive model are unbiased (see Dufour 1985). However, in the construction of \widehat{RC}_{t+h} , the prediction error for each element of the covariance matrix, $e_{ij,t+h} = RC_{ij,t+h} - \widehat{RC}_{ij,t+h}$, has no longer mean zero and depends on the variance Ω_h of the forecast error of the Cholesky factors, $u_{\psi,t+h} = \psi_{t+h} - \hat{\psi}_{t+h}$. Specifically, each element of the predicted covariance matrix $\widehat{RC}_{ij,t+h}$ is biased by $E[e_{ij,t+h}] \equiv \sigma_{ij,h}^* \neq 0$, where $\sigma_{ij,h}^*$ is obtained from the elements of Ω_h . For this reason, Halbleib-Chiriatic and Voev (2011) propose a bias correction that strongly depends on the second moment of the forecasting error, which is model-dependent. Nevertheless, the authors conclude in their manuscript that bias correction is difficult to be empirically justified. Therefore, we do not consider this correction in our paper.

2.2 Matrix logarithm function

Another possible way to parametrize and predict RC_t is based on its matrix logarithm function and the matrix exponential function of the predicted parameters ψ_{t+h} . The matrix exponential function gives a power series expansion on a square matrix A_t such that

$$RC_t = \text{expm}(A_t) = \sum_{m=0}^{\infty} \frac{1}{m!} A_t^m \tag{5}$$

where $\text{expm} = e^{(\cdot)}$ is the matrix exponential function as defined in Chiu et al. (1996).

Conversely, given the square and symmetric matrix RC_t , its logarithm A_t is a square and symmetric matrix that satisfies the condition $\text{expm}(A_t) = RC_t$. It should be remembered that a real matrix has a real logarithm if and only if it is invertible and each Jordan block belonging to a negative eigenvalue occurs an even number of times.

To obtain the logarithm transformation of RC_t , a spectral decomposition is necessary. Because RC_t is positive definite by construction, it has n positive definite eigenvalues λ_t . Letting U_t denote the orthogonal matrix of orthonormal eigenvectors of RC_t and $\Lambda_t = \text{diag}(\lambda_t)$ be a diagonal matrix with eigenvalues on its diagonal, we can write

$$RC_t = U_t \Lambda_t U_t' \tag{6}$$

This is a reformulation of Eq. (2) by simply imposing

$$L_t = \Lambda_t^{1/2} U_t' \tag{7}$$

where $\Lambda_t^{1/2}$ denotes the diagonal matrix with $[\Lambda_t^{1/2}]_{ii} = \sqrt{[\Lambda_t]_{ii}}$. Given the spectral decomposition of RC_t , its matrix logarithm function can be defined as

$$A_t = \text{logm}(RC_t) = U_t \text{log}(\Lambda_t) U_t' \tag{8}$$

where $\text{log}(\Lambda_t) = \text{diag}(\text{log } \lambda_t)$ is the $n \times n$ diagonal matrix with diagonal entries the logarithm of the eigenvalues. The matrix A_t can take any value in the space of $n \times n$ symmetric matrices. Letting $\psi_t = \text{vech}(A_t)$ be the upper triangular elements of $\text{log}(RC_t)$, we obtain the matrix logarithm parametrization of RC_t . As for the Cholesky decomposition, ψ_t is predicted h -steps ahead. Then, the inverse of the *vech* operator is used to form the symmetric predicted logarithm matrix, \hat{A}_{t+h} , and the prediction of RC_t is obtained through its matrix exponential transformation such that $\widehat{RC}_{t+h} = \text{expm}(\hat{A}_{t+h})$.

Differently from Cholesky decomposition, this parametrization is unique as it defines a one-to-one mapping between ψ_t and RC_t . Indeed, it satisfies the first three properties of a covariance matrix reported before. Nevertheless, the vector of parameters ψ_t does not directly refer to the elements of the original matrix RC_t , hence the fourth property stated before cannot be satisfied.

As for the Cholesky decomposition, the predictions of RC_t will be biased since the predictive model is implemented in the log-volatility space and by Jensen's inequality $E(\widehat{RC}_{t+h}) \neq \text{expm}[E(\psi_{t+h})]$. Once again, the solution proposed by Bauer and Vorkink (2011) is to correct somehow the predictions through a data-driven approach. Still, their correction is based on the unrealistic assumption that RC_t is observed and, more worryingly, it may lead to predicted matrices no longer being positive semidefinite.

2.3 Positive semidefinite matrix approximant

In addition to the previous parametrizations of RC_t , we implement the positive semidefinite approximant, proposed by Higham (1988) and Rousseeuw and Molenberghs (1993), and used on realized covariance matrices in Fan et al. (2012). This approach allows computing the nearest symmetric positive semidefinite matrix to an arbitrary matrix in the Frobenius norm. This means that the realized covariances can be directly modelled and that the predicted matrices are *a posteriori* approximated to the nearest positive semidefinite matrix. Therefore, the interpretability of the coefficients of the predictive model remains unchanged, as well as the forecasts are unbiased, at least when the approximation is not necessary.

When a matrix Z_t is symmetric but not positive semidefinite, some eigenvalues in the spectral decomposition as in Eq. 6 are negative. Higham (1988) demonstrates that a PSD

approximant of Z_t can be obtained by simply replacing the negative eigenvalues with zeroes. In such a case, the computation of the unique approximation of Z_t is straightforward. Given the spectral decomposition of $Z_t = U_t \Lambda_t U_t'$, with $\Lambda_t = \text{diag}(\lambda_t)$, where $\lambda_t = \{\lambda_{1,t}, \dots, \lambda_{n,t}\}$ is the vector of eigenvalues, its PSD approximant, S_t , can be computed as follows:

$$S_t = U_t \text{diag}(\lambda_t^*) U_t' \tag{9}$$

where $\lambda_t^* = \{\lambda_{1,t}^*, \dots, \lambda_{n,t}^*\}$ with $\lambda_{i,t}^*$, for $i = 1, \dots, n$, is defined as follows

$$\lambda_{i,t}^* = \begin{cases} \lambda_{i,t}, & \lambda_{i,t} \geq 0 \\ 0, & \lambda_{i,t} < 0. \end{cases} \tag{10}$$

Alternatively, the approximant of Z_t can be computed using the iterative algorithm introduced by Higham (1986) based on matrix inversions. Here, we do not consider this approach and rely only on the approximant presented in Eqs. (9) and (10).

From this specification, it follows that we can directly model and predict the \tilde{n} -dimensional vectorization of the realized covariance matrices without any *a priori* parametrization of the matrices themselves and without caring about semidefinite positiveness. Once obtained the forecast of the realized covariances from a specific model, we approximate the positive semidefinite version of the predicted matrix via Eq. (9). This also allows the practitioner to directly interpret the effect of possible explanatory variables on the realized covariances and avoid any lack of the method in terms of the properties defined in Sect. 3. Differently from the aforementioned approaches, the predictions of the realized covariances remain unbiased and no bias correction methods should be applied to them.

2.4 Matrix correlation parametrization

Finally, we implement the parametrization recently introduced by Archakov and Hansen (2021) that allows us to obtain positive definite forecasts of RC_t from the matrix logarithm transformation of the correlation matrix. Their approach relies on the decomposition of the covariance matrix into standard deviations and correlations, such that

$$RC_t = D_t C_t D_t \tag{11}$$

where C_t is the correlation matrix and D_t is the diagonal matrix with the standard deviation of the asset returns on its diagonal, see also Bollerslev (1990). By reformulating, Eq. (11) can be written in terms of Eq. (2) with $L_t = C_t^{1/2} D_t'$. Archakov and Hansen (2021) propose to parametrize RC_t by the n log-volatilities, $d_t = \log(\text{diag}(D_t))$, and by the

$n(n - 1)/2$ off-diagonal elements of the matrix logarithm transformation of C_t , such that $G_t = \text{logm}(C_t)$. Let $\gamma(G_t)$ be the off-diagonal elements of G_t , the parameters that define RC_t are then $\psi_t = (d_t, \gamma_t)$. As in the case of the Constant/Dynamic Conditional Correlation models by Bollerslev (1990) and Engle (2002), this approach allows modelling separately the variances and correlations. This may be crucial in the modelling of high-dimensional covariance matrices where a simple model may be imposed for the correlations and more complex models may be supposed for the variances.

Once obtained the predictions ψ_{t+h} , reconstructing C_{t+h} and then RC_{t+h} from this approach is not straightforward. To formalize the inverse process to obtain positive definite predicted matrices from d_{t+h} and γ_{t+h} , Archakov and Hansen (2021) introduce a new operator such that, given an $n \times n$ matrix, B , and a vector $\omega \in \mathbb{R}^n$, then $B[\omega]$ is the matrix for which ω has replaced its diagonal. This operator is useful for the definition of Theorem 1 in their manuscript. According to this theorem, for any real symmetric matrix, $B \in \mathbb{R}^{n \times n}$, there is a unique vector, $\omega^* \in \mathbb{R}^n$, such that the exponential of B , $\text{expm}(B[\omega^*])$, is a correlation matrix. Obtaining ω^* requires that the diagonal elements of $\text{expm}(B[\omega^*]) = e^{B[\omega^*]}$ are all equal to one. At this end, the authors propose to apply the following iterative procedure. Consider the sequence,

$$\omega^{(k+1)} = \omega^{(k)} - \log \text{diag} \left(e^{B[\omega^{(k)}]} \right), \quad k = 0, 1, 2, \dots \tag{12}$$

with an arbitrary initial vector $\omega^{(0)} \in \mathbb{R}$, it follows that $\omega^{(k)} \rightarrow \omega^*$, when ω^* is the solution of the aforementioned theorem. The authors find that this algorithm converges very fast, also with high-dimensional matrices. This procedure permits to map the values of $\hat{\gamma}_{t+h}$ to \hat{C}_{t+h} by simply calculating $\hat{C}_{t+h} = \text{expm}(\hat{G}_{t+h}[\omega^*]) = e^{\hat{G}_{t+h}[\omega^*]}$, where \hat{G}_{t+h} is the matrix with off-diagonal elements equal to $\hat{\gamma}_{t+h}$ and diagonal elements equal to ω^* , and then obtain \widehat{RC}_{t+h} in combination with the prediction of the diagonal matrix of volatilities, $\hat{D}_{t+h} = \text{diag}(e^{\hat{d}_{t+h}})$, such that $\widehat{RC}_{t+h} = \hat{D}_{t+h} \hat{C}_{t+h} \hat{D}_{t+h}$.

As further discussed in Archakov and Hansen (2021), this parametrization satisfies all the properties introduced before. Additionally, this approach also permits separately modelling variances and correlations, which is usually a desirable feature in the modelling of asset returns covariances (Heiden 2015). However, the practitioner should account for the fact that the procedure in Eq. (12) must be repeated for each prediction, and this may entail a higher computational time in high-dimensional problems.

We list the parametrizations of RC_t considered in this paper in Table 1.

Table 1 Summary of parametrizations

Method	Parameters modelled, ψ_t	RC_t parametrization
RC	$\psi_t = \text{vech}(RC_t)$	None
Cholesky	$\psi_t = \text{vech}(L'_t)$	Equation (3)
logm	$\psi_t = \text{vech}(\text{logm}(RC_t))$	Equation (8)
Approximant	$\psi_t = \text{vech}(RC_t)$	Equations (9) and (10)
Corr	$\psi_t = (d_t, \gamma_t), d_t = \log \text{diag}(D_t), \gamma_t = \text{vecl}(\text{logm}(C_t))$	Equation (11)

3 Monte Carlo simulations

In this section, we examine the performance of each approach with simulated data. Specifically, we first simulate the time series of high-frequency prices from the stochastic volatility (SV) model proposed by Zhang et al. (2005) and then obtain the daily RC_t matrices from the high-frequency returns as in Eq. (1). For comparison purposes, we also simulate directly the return covariance matrices from a DCC model.

3.1 Simulation from a stochastic volatility model

The data-generating process for each element of the $n \times 1$ vector of prices, $P_t = (p_{1,t}, \dots, p_{n,t})'$, is the stochastic volatility model by Heston (1993), such that

$$dp_{i,t} = (\mu - v_{i,t}/2)dt + \sigma_{i,t}dB_t \tag{13}$$

$$dv_{i,t} = \kappa(\alpha - v_{i,t})dt + \gamma v_{i,t}^{1/2}dW_t, \quad i = 1, \dots, n \tag{14}$$

where $v_{i,t} = \sigma_{i,t}^2$ is the stochastic volatility of the i -th asset. As in Zhang et al. (2005), we assume that the parameters are constant through time and that they are common among the simulated assets. In particular, following the simulation set-up of Zhang et al. (2005), we set $\mu = 0.05$, $\kappa = 5$, $\alpha = 0.04$, $\gamma = 0.5$ and $\rho = -0.5$, where ρ is the correlation coefficient between the two Brownian motions, B and W . To investigate the predictive accuracy of the procedures for a different number of assets, we simulate three samples of prices with n equal to 5, 10 and 15. Assuming that the trading day consists of 6.5 h, we are computing $T = 3000$ daily realized covariance matrices as in Eq. (1), which is almost the same amount of data available in the empirical application, from 39,000 prices sampled at a 30-min frequency.

3.1.1 Out-of-sample evaluation of simulated data from a stochastic volatility model

We produce out-of-sample predictions by fitting a general VAR model to the elements of ψ_t obtained with the different parametrizations. This means that we apply the VAR to make predictions on the elements of ψ_t and we reconstruct the guaranteed positive semidefinite predicted matrix from

the predicted ψ_{t+h} , where h is the forecast horizon. The order of the VAR is automatically selected in the training set by information criteria with a maximum lag order of 5. We then analyse the differences in the predictive accuracy among the parametrizations and compare them with the predictions from a VAR applied directly to the realized covariances. Considering that for a growing number of assets the dimension of the coefficient matrix in a VAR can be very large, we have also analysed the predictions from a penalized VAR as the one introduced in Davis et al. (2016). In addition, we have considered using the prediction from a random walk as an additional competing model, nevertheless, the performance of this is much worse than the one from the considered models and we do not report it in our results.

We divide the whole sample into two subsamples: the in-sample one contains 2/3 of the total observations, *i.e.* $T_1 = 2000$, while the out-of-sample contains the remaining 1/3 of the total observations, hence $T_2 = 1000$. The simulated volatilities for $n = 15$ stocks are depicted in Fig. 1. Since possible biases in the predictive setting further propagate in a multistep-ahead procedure due to the recursive mechanism, we consider both one- and multistep-ahead predictions. Supposing that the realized covariance matrices are sampled at daily frequency, we analyse three forecast horizons: short-term ($h = 1$, one day ahead), mid-term ($h = 5$, one week ahead), and long-term ($h = 22$, one month ahead). Out-of-sample forecasts are obtained from a rolling window with a size of 2000 where the parameters are re-estimated at each step. The size of the window is fixed, while T_2 changes with h as follows: $T_2 = 1000$ for $h = 1$, $T_2 = 996$ for $h = 5$ and $T_2 = 979$ for $h = 22$. The order of the asset returns in the construction of the covariance matrix of asset returns, relevant for the Cholesky decomposition, is random.

To assess the predictive accuracy of the parametrizations, we evaluate the Frobenius and the Procrustes (Dryden and Mardia 2016) distances between the true and predicted covariance matrices at time t . In general, for a pair of covariance matrices S_1 and S_2 , the latter is defined as the Euclidean distance (or Frobenius) which minimises

$$D^2(X_1, X_2) = \|X_2 - X_1\Gamma\|_F \tag{15}$$

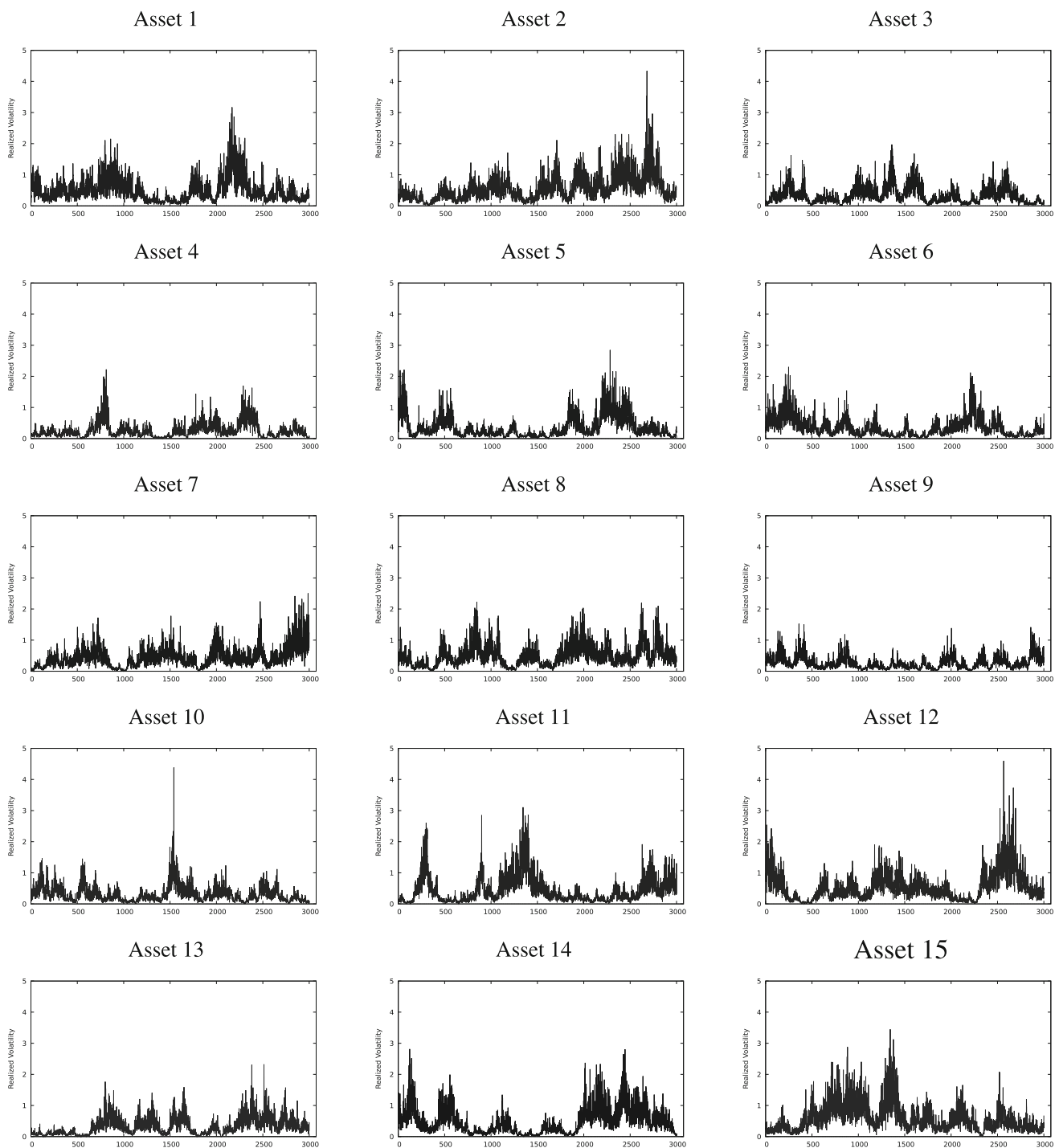


Fig. 1 Volatility plots for $n = 15$ assets simulated from a SV model

where $\|\cdot\|_F$ is the Frobenius norm, Γ is an $n \times n$ rotation matrix and $X_i, i = 1, 2$ are unique non-negative matrix roots such that $S_i = X_i X_i'$.

As discussed in Laurent et al. (2013), the Frobenius distance is robust to a volatility proxy, as in the case of RC_T . On the other hand, the Procrustes distance is less prone to swelling than the Frobenius distance (Marron and Dryden

2021) and can be applied to non-full-rank covariance matrices like realized covariance matrices in high-dimensional allocation problems (Reiss and Winkelmann 2021).

In this paper, we use the average of these loss functions over the out-of-sample observations to analyse the predictions in the rolling window setting and, as descriptive statistics, we also report the number of times the VAR applied

to $\text{vech}(RC_t)$ fails to predict a PSD matrix, denoted by T_2^* . Moreover, the loss functions have been used to compute multiple comparisons via the model confidence set (MCS) methodology introduced by Hansen et al. (2011). For a given confidence level, this procedure identifies the set of models with the best out-of-sample forecasts. Specifically, we test sequentially the equal predictive ability of the considered models. The performance is measured pairwise via the loss functions difference, $\eta_{ij,t} = L(RC_t, H_{i,t}) - L(RC_t, H_{j,t})$, for all $i, j = 1, 2, \dots, M$ with $i \neq j$, where M is the number of compared predictive models (in our case $M = 4$), $L(\cdot)$ is either a Frobenius loss function or a Procrustes metric between the observed realized covariance matrix and its prediction from the i/j -th model, $H_{\cdot,t}$. Assuming that $\eta_{ij,t}$ is stationary, the null hypothesis takes the form

$$H_0: E(\eta_{ij,t}) = 0. \quad (16)$$

When the null hypothesis is rejected, we discard the model with the largest loss function. Hence, the set of superior models (SSM) is composed of all non-discarded models. The MCS p -values are calculated as the number of times the model is included in the SSM through 10000 stationary block bootstraps.

3.1.2 Results for data simulated from a stochastic volatility model

In this section, we discuss the prediction results from fitting the VAR model to the elements of ψ_t . An analysis of the information criteria (Akaike, Bayesian and Hannan-Quinn information criterion) for each parametrization over T_1 suggests that, independently of the number of assets, a first-order VAR is a preferable specification to represent the dynamics of the series of the training set. The average loss functions and the MCS p -values for $h = 1, 5, 22$ are reported in Tables 2, 3 and 4.

When the forecast horizon is equal to 1 (Table 2), the only predictions that always belong to the SSM are the ones from the method proposed by Archakov and Hansen (2021). Specifically, this method has the lowest average Frobenius distance for $n = 5$ and 10, and the lowest Procrustes metric for all the simulated samples. The only method comparable with `Corr` is the Cholesky decomposition which also exhibits a lower Frobenius distance for $n = 15$. It is worth noticing that modelling directly the RC_t series or using the approximant by Higham (1988) leads to an average Frobenius distance which is almost double that of the Cholesky decomposition for $n = 15$.

In the cases of $h = 5$ and $h = 22$, the method of Archakov and Hansen (2021) exhibits the lowest loss functions for all the samples. Indeed, this is the only method included in the SSM for $p_{MCS} > 0.75$. Once again, the Cholesky decompo-

sition is the only parametrization with average loss functions similar to `Corr`.

Conditioning on h , the number of non-PSD predictions obtained by directly modelling the RC_t series (T_2^*) increases with the number of assets. On the other hand, conditioning on n , T_2^* decreases for larger values of h . There is, thus, the hint that by extending the forecast horizon the recursive predictions tend to the average realized covariance matrix which is, in turn, generally positive definite. Interestingly, a parametrization of the covariance matrix always leads to improved predictions with respect to a linear model applied directly on the covariances time series, even when $T_2^* = 0$. Finally, for $n = 15$, we have also estimated a sparse VAR (Davis et al. 2016) where the penalization parameter is tuned at each step of the rolling procedure. The results show that the performance of the different parametrizations under this model is very much the same as the general VAR model and they are thus not reported here.

3.2 Simulation from a DCC model

In order to analyse the forecasting accuracy of the parametrizations with a different simulation setting that produces known covariance matrices, we also simulate the return covariance matrices by a DCC model. With this simulation setting, we obtain time series of volatilities with lower dispersion around the mean and no seasonality, see Fig. 2. Therefore, we are investigating the robustness of the results obtained in the previous section also when the data-generating process changes. The steps of the simulation process from a DCC model are described in Appendix A.

This simulation setting yields return covariance matrices fully observed, allowing us to apply any parametrization and multivariate time series model to make predictions. Considering that the generating process is based on two GARCHs(1,1), out-of-sample predictions are obtained by fitting a VAR(1) model.

3.2.1 Results for data simulated from a DCC

The simulated sample is divided as in Sect. 3.1.1, thus we have a training sample of $T_1 = 2000$ observations and a testing sample of $T_2 = 1000$ observations. The simulated volatilities for $n = 15$ stocks are depicted in Fig. 2. The order of the assets in the construction of the return covariance matrix, relevant for the Cholesky decomposition, is random. It is worth noticing that the simulation design is similar to that suggested by Archakov and Hansen (2021). Accordingly, this may positively influence the accuracy of their parametrization. The evaluation of the out-of-sample forecasts is performed through the methods introduced in Sect. 3.1.1.

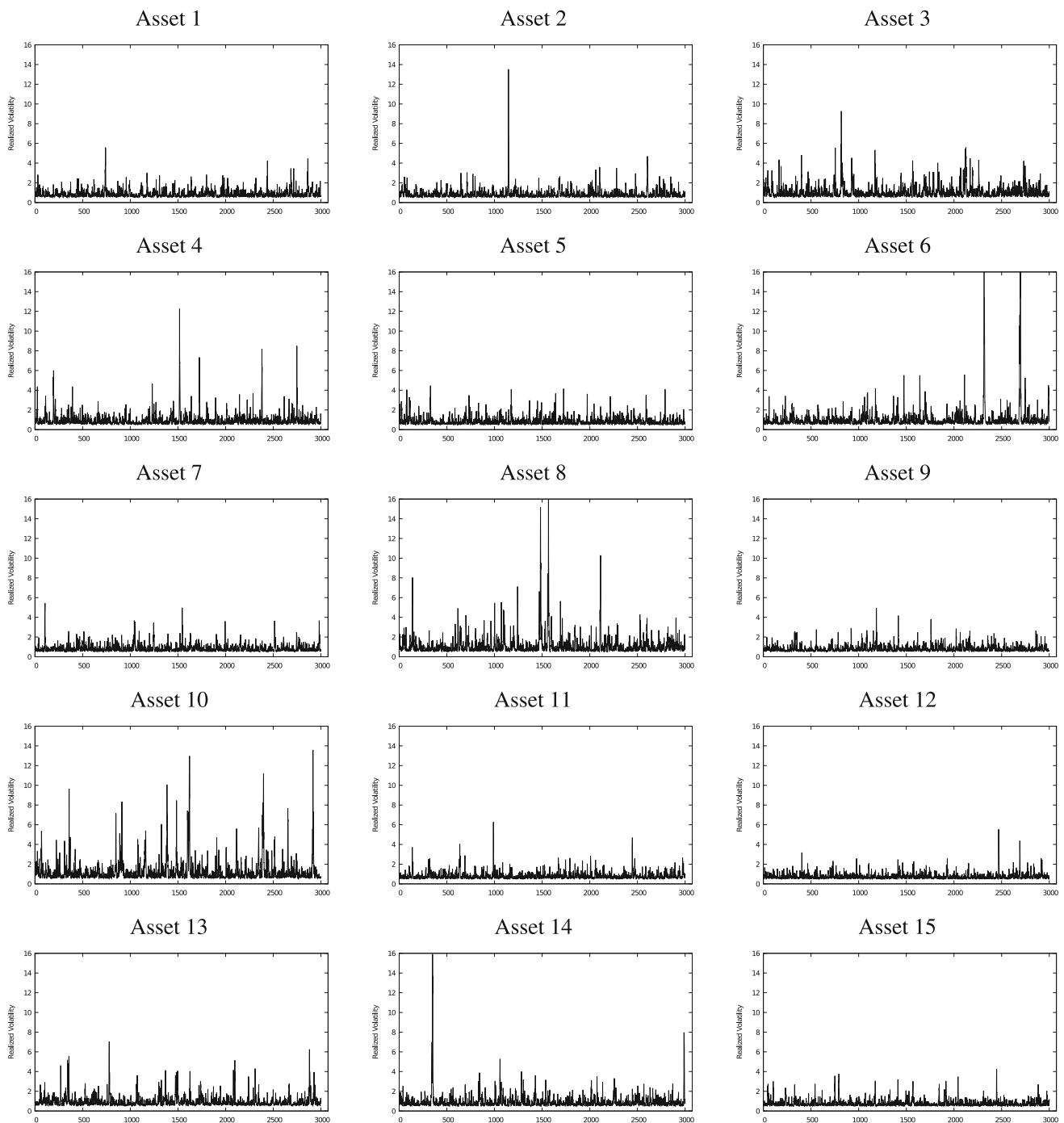


Fig. 2 Volatility plots for $n = 15$ assets simulated from a DCC model

Tables 8, 9, and 10 in Appendix B report the average of both loss functions for each parametrization along with the MCS p -values respectively for $h = 1, 5, 22$. Overall, the results confirm the superiority of the method proposed by Archakov and Hansen (2021) and the Cholesky decomposition, as already emerged from Tables 2, 3 and 4 in the application of simulated data from a stochastic volatility model.

The first evidence from the tables is that the VAR(1) fails to predict positive semidefinite matrices independently of the number of assets and the forecasting horizon, underlining how crucial a parametrization is in a forecasting approach. Not surprisingly, the number of times (T_2^*) a VAR(1) fails to produce PSD predicted matrices increases with the number of assets.

Table 2 Average loss functions for a different number of assets and $h = 1$ in the simulation study from a SV model

Model	Frobenius	p_{MCS}	Procrustes	p_{MCS}
$n = 5$				
		$T_2 = 1000, T_2^* = 0$		
RC	9.004	0.299	0.314	0.000
Cholesky	8.939	0.852	0.309	0.000
logm	9.039	0.260	0.311	0.000
Approx	9.004	0.299	0.314	0.000
Corr	<u>8.523</u>	1.000	<u>0.303</u>	1.000
$n = 10$				
		$T_2 = 1000, T_2^* = 75$		
RC	19.164	0.001	0.372	0.000
Cholesky	23.150	0.000	0.391	0.000
logm	21.722	0.000	0.397	0.000
Approx	19.152	0.000	0.372	0.000
Corr	<u>18.524</u>	1.000	<u>0.363</u>	1.000
$n = 15$				
		$T_2 = 1000, T_2^* = 861$		
RC	114.701	0.000	0.284	0.000
Cholesky	<u>57.343</u>	1.000	0.225	0.326
logm	69.296	0.104	0.234	0.005
Approx	113.100	0.532	0.278	0.000
Corr	62.466	0.874	<u>0.221</u>	1.000

Underlined values denote the lowest loss functions for each setting. RC identifies the predictions from a VAR model directly on realized covariances, without any transformation; Cholesky and logm refer to the Cholesky decomposition and matrix logarithm transformation respectively; Approx concerns predictions from the PSD approximant presented in Sect. 2.3, while Corr refers to the parametrization proposed by Archakov and Hansen (2021). The p -value refers to the probability of being included in the SSM over 10000 block bootstrap replicates (in bold values with a p_{MCS} greater than 0.75). T_2^* is the number of times the VAR applied to $\text{vech}(RC_t)$ fails to predict a PSD matrix

For any combination of n and h , the differences between the parametrizations are relatively small, while there emerges a clear difference with the approach in which no-parametrization is performed and with the approximant of Higham (1988). These differences explode for $n = 10$ and $n = 15$. In general, the parametrization proposed by Archakov and Hansen (2021) outperforms the compared methodologies for the average of both loss functions. When looking at the average Frobenius distance, the use of Cholesky decomposition obtains the overall best performance no matter what the number of assets, only for $h = 1$. Still, its performance is closely followed by the parametrization proposed by Archakov and Hansen (2021), which is also the only parametrization always included in the SSM for $p_{MCS} > 0.75$ in terms of the Procrustes measure. The matrix logarithm transformation seems to be a valid tool to parametrize the covariance matrix, since

Table 3 Average loss functions for a different number of assets and $h = 5$ in the simulation study from a SV model

Model	Frobenius	p_{MCS}	Procrustes	p_{MCS}
$n = 5$				
		$T_2 = 996, T_2^* = 0$		
RC	13.449	0.000	0.386	0.000
Cholesky	12.405	0.003	0.361	0.001
logm	13.143	0.000	0.375	0.000
Approx	13.449	0.000	0.386	0.000
Corr	<u>11.769</u>	1.000	<u>0.352</u>	1.000
$n = 10$				
		$T_2 = 996, T_2^* = 1$		
RC	25.747	0.000	0.441	0.000
Cholesky	25.451	0.004	0.429	0.000
logm	28.544	0.002	0.461	0.000
Approx	25.747	0.000	0.441	0.000
Corr	<u>22.581</u>	1.000	<u>0.405</u>	1.000
$n = 15$				
		$T_2 = 996, T_2^* = 387$		
RC	51.039	0.489	0.972	0.000
Cholesky	56.089	0.439	0.318	0.000
logm	258.125	0.016	0.299	0.000
Approx	50.895	0.592	0.971	0.000
Corr	<u>50.029</u>	1.000	<u>0.294</u>	1.000

Underlined values denote the lowest loss functions for each setting. RC identifies the predictions from a VAR model directly on realized covariances, without any transformation; Cholesky and logm refer to the Cholesky decomposition and matrix logarithm transformation respectively; Approx concerns predictions from the PSD approximant presented in Sect. 2.3, while Corr refers to the parametrization proposed by Archakov and Hansen (2021). The p -value refers to the probability of being included in the SSM over 10000 block bootstrap replicates (in bold values with a p_{MCS} greater than 0.75). T_2^* is the number of times the VAR applied to $\text{vech}(RC_t)$ fails to predict a PSD matrix

it performs similarly to the approach proposed by Archakov and Hansen (2021). In fact, it also belongs to the SSM for $h = 22$ and $n = 5$. Furthermore, Cholesky decomposition is strongly outperformed in terms of average loss functions for $n = 15$ when the forecast horizon is different from $h = 1$.

4 Empirical evidence

4.1 Forecasting US stocks realized covariances

The dataset used to compare the forecasting accuracy of the parametrizations comprehends the time series of hourly open spot prices of the 10 largest (as of November 2021) US stocks by market capitalization belonging to the S&P500 for the period August 31, 2010–November 30, 2021, sourced from

Table 4 Average loss functions for a different number of assets and $h = 22$ in the simulation study from a SV model

Model	Frobenius	p_{MCS}	Procrustes	p_{MCS}
$n = 5$				
$T_2 = 979, T_2^* = 0$				
RC	13.345	0.000	0.387	0.000
Cholesky	12.227	0.000	0.360	0.001
logm	12.935	0.000	0.374	0.000
Approx	13.345	0.000	0.387	0.000
Corr	<u>11.703</u>	1.000	<u>0.353</u>	1.000
$n = 10$				
$T_2 = 979, T_2^* = 1$				
RC	25.990	0.000	0.443	0.000
Cholesky	25.424	0.000	0.430	0.000
logm	28.352	0.000	0.461	0.000
Approx	25.990	0.000	0.443	0.000
Corr	<u>22.838</u>	1.000	<u>0.409</u>	1.000
$n = 15$				
$T_2 = 979, T_2^* = 7$				
RC	75.874	0.000	0.647	0.000
Cholesky	56.288	0.000	0.559	0.000
logm	94.764	0.000	0.749	0.000
Approx	59.928	0.000	0.563	0.000
Corr	<u>56.285</u>	1.000	<u>0.559</u>	1.000

Underlined values denote the lowest loss functions for each setting. RC identifies the predictions from a VAR model directly on realized covariances, without any transformation; Cholesky and logm refer to the Cholesky decomposition and matrix logarithm transformation respectively; Approx concerns predictions from the PSD approximant presented in Sect. 2.3, while Corr refers to the parametrization proposed by Archakov and Hansen (2021). The p -value refers to the probability of being included in the SSM over 10000 block bootstrap replicates (in bold values with a p_{MCS} greater than 0.75). T_2^* is the number of times the VAR applied to $\text{vech}(RC_t)$ fails to predict a PSD matrix

Bloomberg (19,766 hourly observations). The choice of estimating daily RC_t covariance matrices from hourly data is in line with De Pooter et al. (2008) which show that optimal intra-day frequencies correspond to 30 min or an hour. Table 5 reports the complete list of stock names and tickers. Daily realized volatility matrices are computed as in Eq. (1), for a total amount of $T = 2833$ daily observations. The covariance stationarity of each series is confirmed by Augmented Dickey-Fuller (ADF), ADF-GLS, Phillips-Perron (PP), and KPSS tests for unit roots. The realized volatilities of the $n = 10$ stocks are depicted in Fig. 3.

In order to evaluate the forecasting accuracy of the parametrizations, we use almost 1/3 of the total observations ($T_2 = 943, 939, 922$ respectively for $h = 1, 5, 22$) from March 6, 2018 to November 30, 2021 for the out-of-sample analysis. To avoid any look-ahead bias or distortion in the

predictions due to a bad sampling scheme, we produce one- and multistep ahead forecasts for a rolling window of 1890 observations that foresees the estimation of the parameters at each step.

We select the number of lags to be used in the predictive VAR through information criteria applied to the training sample. All the criteria for all the parametrizations suggest a single lag, hence we make predictions of the dependent variables, ψ_t , from a VAR(1).

4.2 Forecasting results for daily realized covariances

In this section, we evaluate out-of-sample forecasts through the methods introduced in Sect. 3.1.1 and report the average loss functions in Table 6. As in the case of simulated data, Cholesky decomposition is the one with the lowest Frobenius loss for $h = 1$. Nevertheless, in the comparison of the predictions for $h = 1$, this does not lead to statistically significant differences, since all the models, except for the matrix logarithm transformation, belong to the SSM. For one-step-ahead predictions, the approach proposed by Archakov and Hansen (2021) is the one with the lowest average Procrustes metric. The supremacy of the Cholesky decomposition and the Corr approach is also evident from multistep-ahead predictions. In fact, for $h = 5$, these parametrizations outperform the competing ones and are both included in the SSM when looking at the Frobenius distance. Surprisingly, the PSD approximant has the lowest Procrustes measure for $h = 5$. As already observed in the simulated data, the differences in terms of predictive accuracy between these parametrizations and the simple VAR(1) or the *a posteriori* PSD approximation increase with the number of steps ahead. For $h = 22$, the lowest losses are observed for the method of Archakov and Hansen (2021). Intuitively, the performance of no-parametrization and the approximant by Higham (1988) is mainly influenced by the poor accuracy in the observations belonging to T_2^* . In Appendix A, we also report a heatmap of the Root Mean Square Error (RMSE) calculated between the observed and the predicted elements of the covariance matrix through the different parametrizations for $h = 1$. The heatmaps show that Cholesky and Corr outperform other methods for almost all the elements of the lower triangular part of the covariance matrix.

4.3 Forecasting results for weekly realized covariances

For robustness purposes, we also use the hourly data presented in Sect. 4.1 to compute weekly realized covariances. In particular, weekly realized volatility matrices are computed as in Eq. (1), for a total amount of $T = 595$ observations. To allow an out-of-sample evaluation, one-third of the total observations ($T_2 = 200, 196, 179$ respectively for

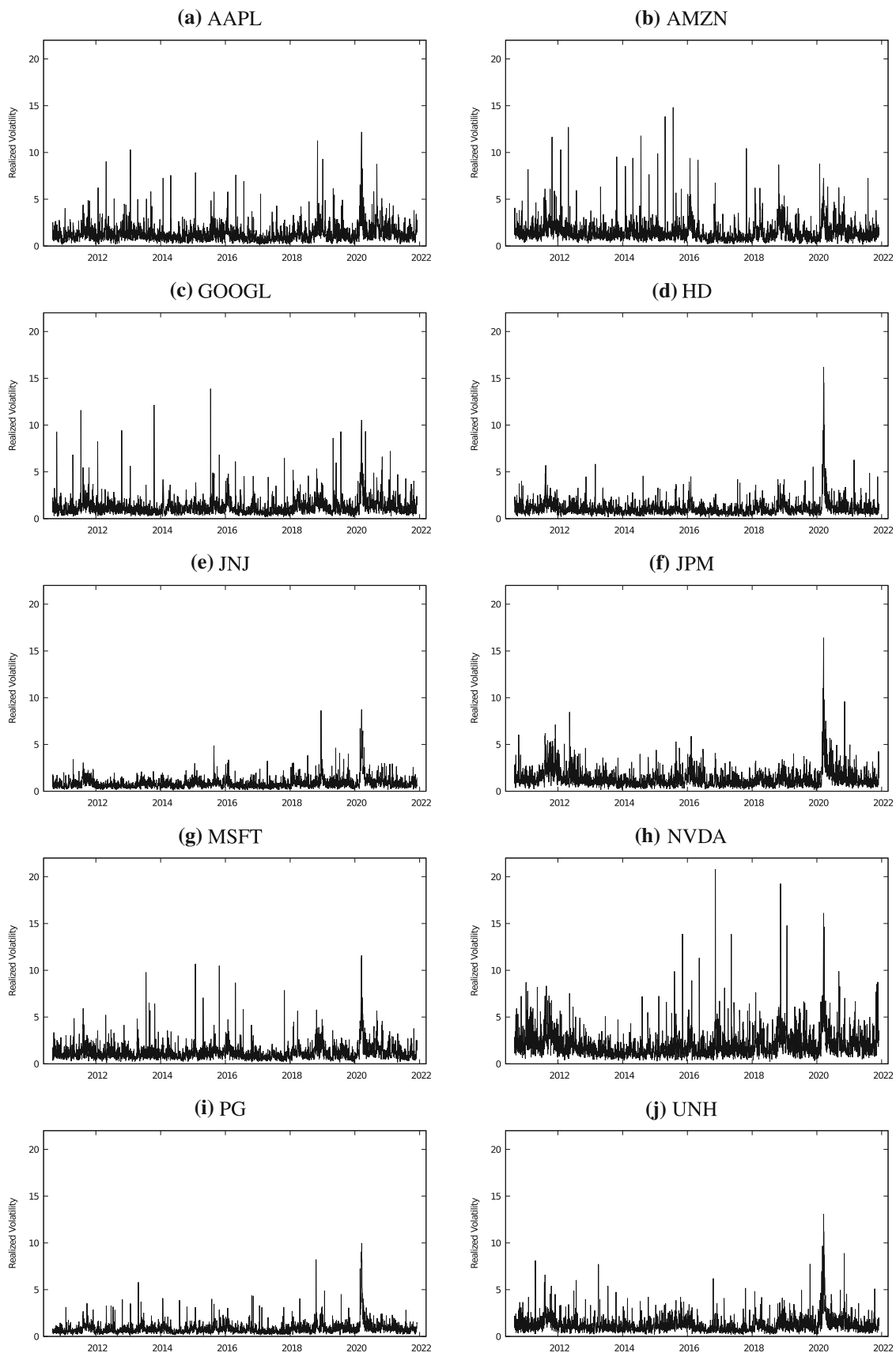


Fig. 3 Daily realized volatility plots

Table 5 List of the 10 US stocks used in this study

Ticker	Stock	Ticker	Stock
AAPL	Apple Inc.	JPM	JPMorgan Chase & Co.
AMZN	Amazon.com, Inc.	MSFT	Microsoft Corporation
GOOGL	Alphabet Inc. Class A.	NVDA	NVIDIA Corporation
HD	Home Depot Inc.	PG	Procter & Gamble Co.
JNJ	Johnson & Johnson	UNH	UnitedHealth Group Inc.

Table 6 Average loss functions for daily empirical data in different forecast horizons

Model	Frobenius	p_{MCS}	Procrustes	p_{MCS}
$h = 1$				
		$T_2 = 943, T_2^* = 591$		
RC	4505.439	1.000	6.054	0.000
Cholesky	<u>4204.497</u>	1.000	5.518	0.228
logm	5869.325	0.299	5.860	0.000
Approx	4287.349	1.000	5.975	0.000
Corr	4623.503	1.000	<u>5.494</u>	1.000
$h = 5$				
		$T_2 = 939, T_2^* = 4$		
RC	9475.311	0.580	5.552	0.747
Cholesky	<u>5589.961</u>	1.000	5.630	0.299
logm	5951.430	0.038	5.907	0.000
Approx	9036.394	0.713	<u>5.541</u>	1.000
Corr	5601.428	0.976	5.617	0.347
$h = 22$				
		$T_2 = 922, T_2^* = 4$		
RC	153461.345	0.606	8.239	0.862
Cholesky	5861.231	0.158	5.652	0.872
logm	6038.107	0.007	5.919	0.000
Approx	153460.363	0.731	8.239	0.862
Corr	<u>5744.141</u>	1.000	<u>5.643</u>	1.000

Underlined values denote the lowest loss functions for each setting. RC identifies the predictions from a VAR model directly on realized covariances, without any transformation; Cholesky and logm refer to the Cholesky decomposition and matrix logarithm transformation respectively; Approx concerns predictions from the PSD approximant presented in Sect. 2.3, while Corr refers to the parametrization proposed by Archakov and Hansen (2021). The p -value refers to the probability of being included in the SSM over 10000 block bootstrap replicates (in bold values with a p_{MCS} greater than 0.75). T_2^* is the number of times the VAR applied to $\text{vech}(RC_t)$ fails to predict a PSD matrix

$h = 1, 5, 22$) is used as a testing sample. As for the case of daily realized covariances, we produce the forecasts from a rolling window of 395 weekly observations and estimate the parameters of the VAR at each step. We have performed unit

Table 7 Average loss functions for weekly empirical data in different forecast horizons

Model	Frobenius	p_{MCS}	Procrustes	p_{MCS}
$h = 1$				
		$T_2 = 200, T_2^* = 131$		
RC	164065.06	1.000	20.990	0.126
Cholesky	<u>56092.65</u>	1.000	<u>17.117</u>	1.000
logm	95840.01	0.517	23.543	0.001
Approx	161207.99	1.000	20.815	0.213
Corr	85213.35	1.000	20.394	0.852
$h = 5$				
		$T_2 = 196, T_2^* = 35$		
RC	141896.36	0.129	20.694	0.000
Cholesky	76799.89	1.000	17.692	0.000
logm	77035.34	1.000	18.047	0.000
Approx	141950.56	0.138	20.694	0.000
Corr	<u>75113.08</u>	1.000	<u>17.386</u>	1.000
$h = 22$				
		$T_2 = 179, T_2^* = 64$		
RC	244928.00	0.225	21.472	0.000
Cholesky	83154.01	1.000	18.599	0.000
logm	84335.26	1.000	19.046	0.000
Approx	244544.71	0.225	21.472	0.000
Corr	<u>82327.19</u>	1.000	<u>18.325</u>	1.000

Underlined values denote the lowest loss functions for each setting. RC identifies the predictions from a VAR model directly on realized covariances, without any transformation; Cholesky and logm refer to the Cholesky decomposition and matrix logarithm transformation respectively; Approx concerns predictions from the PSD approximant presented in Sect. 2.3, while Corr refers to the parametrization proposed by Archakov and Hansen (2021). The p -value refers to the probability of being included in the SSM over 10000 block bootstrap replicates (in bold values with a p_{MCS} greater than 0.75). T_2^* is the number of times the VAR applied to $\text{vech}(RC_t)$ fails to predict a PSD matrix

roots test to confirm the covariance stationarity of the time series.

The average loss functions for weekly RC_t are reported in Table 7. The findings are mostly overlapping with those from the previous Sections. In fact, the parametrization introduced by Archakov and Hansen (2021) outperforms all the alternatives in terms of predictive accuracy. Nevertheless, the difference with the Cholesky decomposition and the matrix

logarithm transformation for the Frobenius distance is not significant for $h = 5$, since they are all included in the SSM for a p_{MCS} greater than 0.75.

5 Conclusions

Predicting return covariance matrices is of particular importance to the fields of risk management, portfolio management and asset pricing. However, modelling the dynamics of a covariance matrix is a challenging task as the chosen model must guarantee the positive semi-definiteness of the predicted covariance matrices.

In this paper, by using simulated and real data, we have discussed the performance of different parametrization methods which allow the predicted covariance matrices to be at least semidefinite positive. Our findings are that different parametrizations can lead to different results in terms of forecasting accuracy of asset return covariance matrices. In particular, we have found that Cholesky decomposition and the new parametrization proposed by Archakov and Hansen (2021) provided the lowest loss functions and that the difference between such methods is often statistically not significant. As a result, directly modelling the return covariances or using the *a posteriori* approximant of Higham (1988) produce the poorest predictions. This evidence is mainly due to the predictive errors made in non-positive semidefinite observations. Interestingly, we have found that a parametrization of the covariance matrix always leads to improved predictions with respect to a linear model applied directly on the covariances time series, also when this method does not fail to predict PSD matrices. This finding, which is worth further investigations for future works, clearly emerges in multistep-ahead forecasts, where non-PSD matrices are used recursively to produce successive predictions.

Among the two best-performing methods, possible advantages and disadvantages also emerge. Although computationally more complex, the one proposed by Archakov and Hansen (2021) respects all the properties that a parametrization should have. Moreover, it allows the implementation of different models for variances and correlations. In the ever-evolving context of time series modelling, this can be crucial for the practitioner to analyse possible nonlinear relations or to investigate the role of explanatory variables on variances and/or correlations. Conversely, Cholesky decomposition does not satisfy all the properties of a parametrization. In particular, the predictions from this method strongly rely on the ordering of the series in the construction of the covariance matrix. Nevertheless, this is the simplest method among the ones we consider in this analysis and, unsurprisingly, it is the most popular method so far implemented in the prediction of realized covariances.

The analysis presented in this paper has been conducted under simple hypotheses about the price process. However,

raw high-frequency data typically deviates from the ideal situations since, for example, jumps and market microstructure noise appear as stylized features of financial data. Studying the performance of the parametrization methods under more complicated settings will be a task for future works.

Finally, it should be noted that the estimated VARs, especially for $n = 15$, involve a large number of dependent variables. To deal with possible problems related to the high number of coefficients, researchers typically use prior shrinkage on the coefficients. In our simulation study, considering the case of $n = 15$ assets, we have compared unrestricted formulations with L_1 -penalized versions of the model. However, investigating the use of other approaches in a high-dimensional setting, such as the Stochastic Search Variable Selection model (George et al. 2008) and the Bayesian Compressed regression (Guhaniyogi and Dunson 2015), could be worth considering in future works.

Funding Open access funding provided by Università degli Studi G. D'Annunzio Chieti-Pescara within the CRUI-CARE Agreement.

Declarations

Competing interests None of the authors have conflicts of interest to declare.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Appendix A: Simulation process from a DCC model

A first step in the simulation of the covariance matrix of returns from a DCC foresees the simulation of univariate Generalized AutoRegressive Conditional Heteroskedasticity (GARCh) models for the diagonal entries of the covariance matrix. Let us define $r_{i,t}$ the return of the i -th asset at time t , from the ARCH specification by Engle (1982) it follows that

$$r_{i,t} = \sigma_{i,t} \epsilon_{i,t}$$

and

$$\frac{r_{i,t}}{\sigma_{i,t}} = \epsilon_{i,t} \stackrel{iid}{\sim} N(0, 1),$$

where $\epsilon_{i,t}$ are the standardized residuals for the i -th asset. Therefore, the dynamics of the asset returns volatility follows a GARCH(1,1) on the form:

$$\sigma_{i,t}^2 = \omega_i + \alpha_i r_{i,t-1}^2 + \beta_i \sigma_{i,t-1}^2 \tag{A1}$$

where α_i and β_i are respectively the coefficients associated with $r_{i,t-1}^2$ and the lag of $\sigma_{i,t}^2$, while ω_i is the long-run variance.

To initialize the simulation, we sample the first observation of the asset returns from a Normal distribution, $r_{i,1} \sim N(0, \sigma_{i,1}^2)$, where

$$\sigma_{i,1}^2 = \frac{\omega_i}{1 - \alpha_i - \beta_i}$$

is the unconditional variance. For $t = 2, \dots, T$, we update the dynamics of $\sigma_{i,t}^2$ as in Eq. (A1). This allows us to compute $D_t = \text{diag}(\sigma_{1,t}^2, \dots, \sigma_{n,t}^2)$.

After the simulation of the conditional variances for each of the n assets, we need to use the standardized residuals to simulate the DCC dynamics and reconstruct the correlation matrix R_t . Specifically, we need to update the $n \times n$ covariance matrix of standardized residuals, Q_t , and obtain

$$R_t = \bar{Q}_t^{-1/2} Q_t \bar{Q}_t^{-1/2}$$

where $\bar{Q}_t^{-1/2}$ contains the diagonal elements of Q_t . Q_t can be further modelled as a GARCH(1,1)-like structure, such that

$$Q_t = \bar{Q}(1 - a - b) + a(\epsilon_{t-1} \epsilon'_{t-1}) + bQ_{t-1} \tag{A2}$$

where the assumption of weak-stationarity, given by $a + b < 1$, is necessary to ensure that Q_t is positive definite, and \bar{Q} is the unconditional covariance matrix of the multivariate standardized residuals, $\epsilon_t = D_t^{-1/2} r_t \sim N(0, R_t)$, computed as

$$\bar{Q} = \frac{1}{T} \sum_{t=1}^T \epsilon_t \epsilon'_t. \tag{A3}$$

From this, we can reconstruct the covariance matrix via the using decomposition proposed by Bollerslev (1990) and shown in Eq. (11), such that the covariance matrix of returns is obtained as $D_t^{1/2} R_t D_t^{1/2}$.

As for the case of simulated data from a stochastic volatility model, we generate three samples with a different number of assets, $n = 5, 10, 15$. For each sample, we generate

$T = 3000$ observations. The data generating process for the i -th asset return volatility is the following:

$$\begin{aligned} \sigma_{i,t}^2 &= 0.2 + \alpha_i r_{i,t-1}^2 + \beta_i \sigma_{i,t-1}^2, \\ \alpha_i &\sim U(0.5, 0.7), \beta_i \sim U(0.2, 0.3) \end{aligned} \tag{A4}$$

where $U(\phi_1, \phi_2)$ denotes a uniform distribution with extremes ϕ_1 and ϕ_2 . Then, we set $a = 0.35$ and $b = 0.4$ for the simulation of Q_t as in Eq. (A2).

Appendix B: Tables of loss functions for simulated data from a DCC

See Tables 8, 9 and 10.

Table 8 Average loss functions for a different number of assets and $h = 1$ in the simulation study from a DCC model

Model	Frobenius	p_{MCS}	Procrustes	p_{MCS}
$n = 5$				
		$T_2 = 1000, T_2^* = 156$		
RC	94.861	0.006	1.494	0.000
Cholesky	<u>87.179</u>	1.000	1.339	0.079
logm	97.736	0.239	1.320	0.000
Approx	94.250	0.568	1.493	0.033
Corr	93.842	0.555	<u>1.304</u>	1.000
$n = 10$				
		$T_2 = 1000, T_2^* = 549$		
RC	397.024	0.215	2.390	0.000
Cholesky	<u>358.484</u>	1.000	<u>2.166</u>	1.000
logm	523.325	0.150	2.335	0.046
Approx	389.295	0.542	2.378	0.000
Corr	373.328	0.902	2.231	0.764
$n = 15$				
		$T_2 = 1000, T_2^* = 861$		
RC	1147.012	0.000	2.840	0.000
Cholesky	<u>573.432</u>	1.000	2.250	0.326
logm	692.956	0.104	2.336	0.005
Approx	1131.002	0.532	2.783	0.000
Corr	624.663	0.874	<u>2.213</u>	1.000

Underlined values denote the lowest loss functions for each setting. RC identifies the predictions from a VAR model directly on realized covariances, without any transformation; Cholesky and logm refer to the Cholesky decomposition and matrix logarithm transformation respectively; Approx concerns predictions from the PSD approximant presented in Sect. 2.3, while Corr refers to the parametrization proposed by Archakov and Hansen (2021). The p -value refers to the probability of being included in the SSM over 10000 block bootstrap replicates (in bold values with a p_{MCS} greater than 0.75). T_2^* is the number of times the VAR applied to $\text{vech}(RC_t)$ fails to predict a PSD matrix

Table 9 Average loss functions for a different number of assets and $h = 5$ in the simulation study from a DCC model

Model	Frobenius	p_{MCS}	Procrustes	p_{MCS}
$n = 5$				
$T_2 = 996, T_2^* = 15$				
RC	172.768	0.051	1.968	0.000
Cholesky	163.171	0.841	1.796	0.200
logm	163.377	0.017	1.790	0.003
Approx	172.692	0.787	1.968	0.000
Corr	<u>162.032</u>	1.000	<u>1.775</u>	1.000
$n = 10$				
$T_2 = 996, T_2^* = 156$				
RC	3813.088	0.234	3.954	0.010
Cholesky	919.405	1.000	3.148	0.495
logm	857.991	1.000	3.106	0.943
Approx	3798.891	0.237	3.944	0.084
Corr	<u>848.520</u>	1.000	<u>3.054</u>	1.000
$n = 15$				
Full sample - $T_2 = 996, T_2^* = 306$				
RC	65039.131	0.489	9.717	0.000
Cholesky	1302.355	0.439	3.178	0.000
logm	1015.444	0.592	2.994	0.000
Approx	64990.408	0.016	9.709	0.000
Corr	<u>1002.991</u>	1.000	<u>2.942</u>	1.000

Underlined values denote the lowest loss functions for each setting. RC identifies the predictions from a VAR model directly on realized covariances, without any transformation; Cholesky and logm refer to the Cholesky decomposition and matrix logarithm transformation respectively; Approx concerns predictions from the PSD approximant presented in Sect. 2.3, while Corr refers to the parametrization proposed by Archakov and Hansen (2021). The p -value refers to the probability of being included in the SSM over 10000 block bootstrap replicates (in bold values with a p_{MCS} greater than 0.75). T_2^* is the number of times the VAR applied to $\text{vech}(RC_t)$ fails to predict a PSD matrix

Table 10 Average loss functions for a different number of assets and $h = 22$ in the simulation study from a DCC model

Model	Frobenius	p_{MCS}	Procrustes	p_{MCS}
$n = 5$				
$T_2 = 979, T_2^* = 15$				
RC	187.289	0.055	2.096	0.001
Cholesky	175.187	0.469	1.942	0.079
logm	169.001	0.841	1.859	1.000
Approx	187.202	0.039	2.096	0.008
Corr	<u>168.191</u>	1.000	<u>1.854</u>	0.967
$n = 10$				
Full sample - $T_2 = 979, T_2^* = 154$				
RC	4070.829	0.262	4.297	0.006
Cholesky	1020.997	0.546	3.464	0.003
logm	881.021	0.060	3.187	0.001
Approx	4054.282	0.489	4.287	0.008
Corr	<u>875.804</u>	1.000	<u>3.163</u>	1.000
$n = 15$				
Full sample - $T_2 = 979, T_2^* = 294$				
RC	66489.107	0.484	10.147	0.000
Cholesky	1480.359	0.293	3.396	0.000
logm	1042.834	0.563	3.040	0.000
Approx	66439.507	0.001	10.127	0.000
Corr	<u>1038.839</u>	1.000	<u>3.017</u>	1.000

Underlined values denote the lowest loss functions for each setting. RC identifies the predictions from a VAR model directly on realized covariances, without any transformation; Cholesky and logm refer to the Cholesky decomposition and matrix logarithm transformation respectively; Approx concerns predictions from the PSD approximant presented in Sect. 2.3, while Corr refers to the parametrization proposed by Archakov and Hansen (2021). The p -value refers to the probability of being included in the SSM over 10000 block bootstrap replicates (in bold values with a p_{MCS} greater than 0.75). T_2^* is the number of times the VAR applied to $\text{vech}(RC_t)$ fails to predict a PSD matrix.

Appendix C: Heatmap of RMSE for the empirical application for $h = 1$

See Figs. 4, 5, 6, 7, and 8.



Fig. 4 Heatmap of the RMSE for the one-step ahead predictions from the VAR on the realized covariances. RMSE for the lower triangular elements of the covariance matrix



Fig. 5 Heatmap of the RMSE for the one-step ahead predictions from the PSD approximant on the realized covariances. RMSE for the lower triangular elements of the covariance matrix



Fig. 6 Heatmap of the RMSE for the one-step ahead predictions from the Cholesky parametrization on the realized covariances. RMSE for the lower triangular elements of the covariance matrix

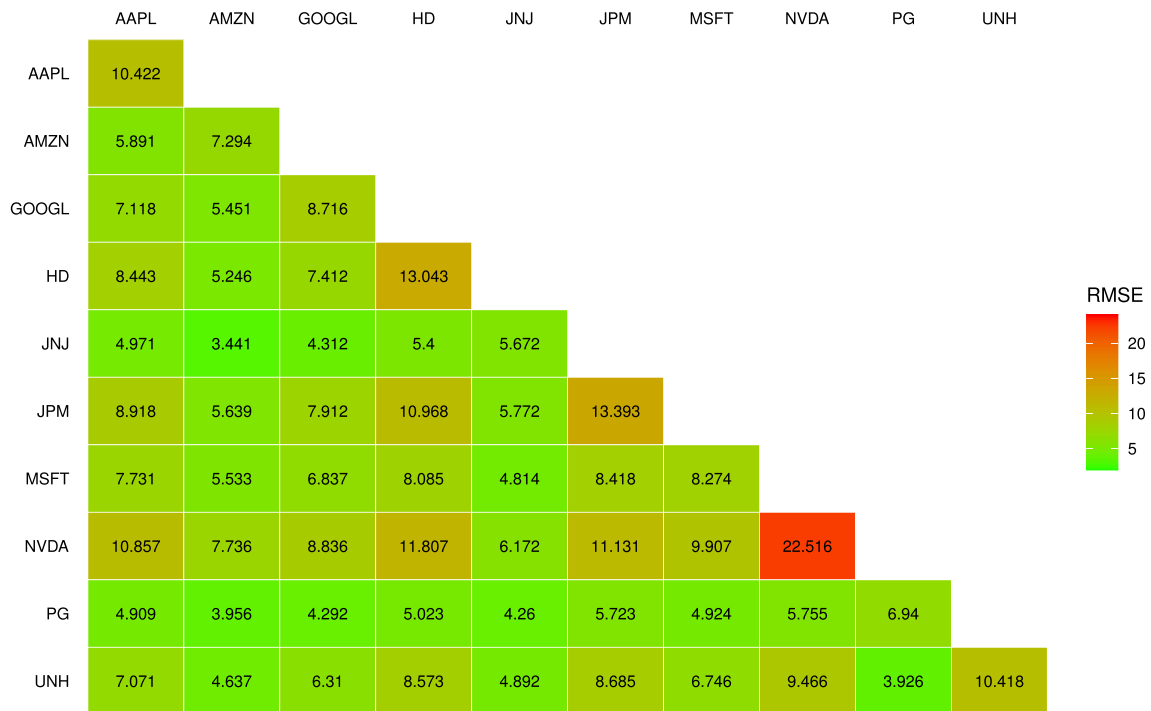


Fig. 7 Heatmap of the RMSE for the one-step ahead predictions from the matrix logarithm parametrization on the realized covariances. RMSE for the lower triangular elements of the covariance matrix

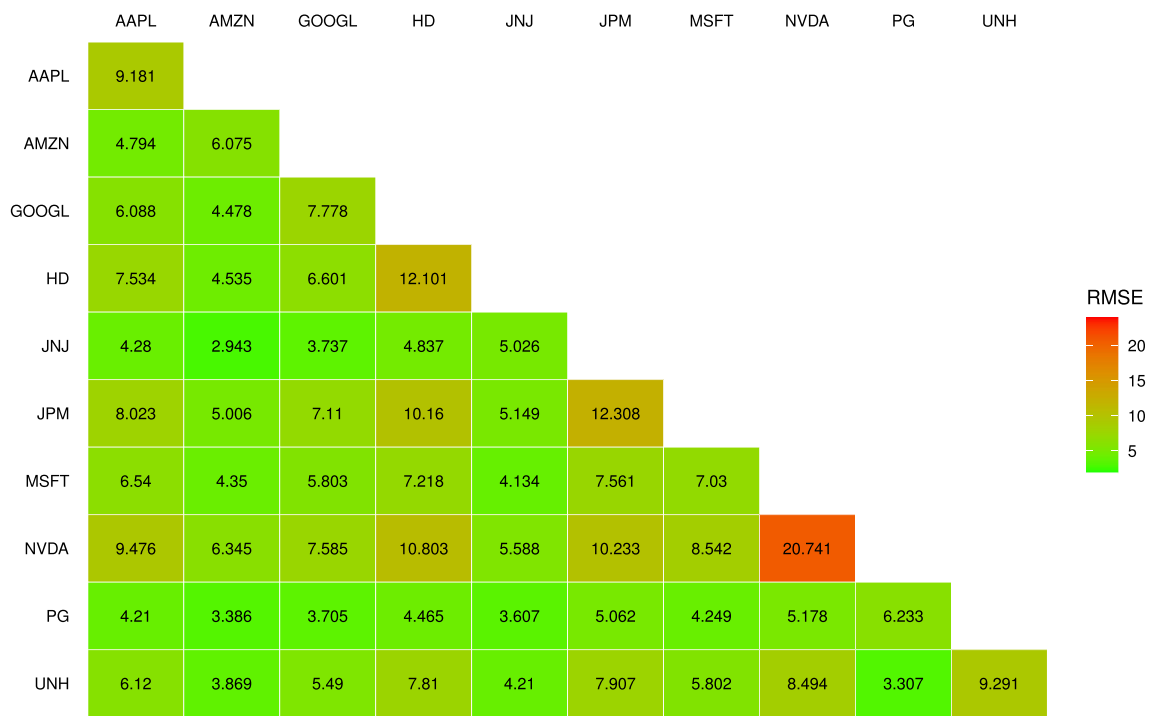


Fig. 8 Heatmap of the RMSE for the one-step ahead predictions from parametrization proposed by Archakov and Hansen (2021) on the realized covariances. RMSE for the lower triangular elements of the covariance matrix

References

- Andersen, T.G., Bollerslev, T., Diebold, F.X., Ebens, H.: The distribution of realized stock return volatility. *J. Financ. Econ.* **61**, 43–76 (2001)
- Andersen, T.G., Bollerslev, T., Diebold, F.X., Labys, P.: The distribution of exchange rate volatility. *J. Am. Stat. Assoc.* **96**, 42–55 (2001)
- Andersen, T.G., Bollerslev, T., Diebold, F.X., Labys, P.: Modeling and forecasting realized volatility. *Econometrica* **71**, 579–625 (2003)
- Archakov, I., Hansen, P.R.: A new parametrization of correlation matrices. *Econometrica* **89**(4), 1699–1715 (2021). <https://doi.org/10.3982/ECTA16910>
- Asai, M., Chang, C.-L., McAleer, M.: Realized matrix-exponential stochastic volatility with asymmetry, long memory and higher-moment spillovers. *J. Econom.* **227**(1), 285–304 (2022)
- Barndorff-Nielsen, O.E., Shephard, N.: Econometric analysis of realised volatility and its use in estimating stochastic volatility models. *J. R. Stat. Soc.* **64**(2), 253–280 (2002)
- Barndorff-Nielsen, O.E., Shephard, N.: Estimating quadratic variation using realised variance. *J. Appl. Econom.* **17**, 457–477 (2002)
- Barndorff-Nielsen, O.E., Shephard, N.: Econometric analysis of realized covariation: high frequency based covariance, regression, and correlation in financial economics. *Econometrica* **72**(3), 885–925 (2004)
- Bauer, G.H., Vorkink, K.: Forecasting multivariate realized stock market volatility. *J. Econom.* **160**(1), 93–101 (2011)
- Bollerslev, T.: Modelling the coherence in short-run nominal exchange rates: a multivariate generalized arch model. *Rev. Econ. Stat.* **72**(3), 498–505 (1990)
- Bucci, A.: Cholesky-ANN models for predicting multivariate realized volatility. *J. Forecast.* **39**(6), 865–876 (2020)
- C.Čech, F., & Barunk, J.: On the modelling and forecasting of multivariate realized volatility: generalized heterogeneous autoregressive (GHAR) model. *J. Forecast.* **36**(2), 181–206 (2017). <https://doi.org/10.1002/for.2423>
- Chiu, T.Y.M., Leonard, T., Tsui, K.-W.: The Matrix-Logarithmic Covariance Model. *J. Am. Stat. Assoc.* **91**(433), 198–210 (1996). <https://doi.org/10.2307/2291396>
- Christiansen, C., Schmeling, M., Schrimpf, A.: A comprehensive look at financial volatility prediction by economic variables. *J. Appl. Econom.* **27**(6), 956–977 (2012)
- Davis, R.A., Zang, P., Zheng, T.: Sparse vector autoregressive modeling. *J. Comput. Graph. Stat.* **25**(4), 1077–1096 (2016). <https://doi.org/10.1080/10618600.2015.1092978>
- De Pooter, M., Martens, M., Van Dijk, D.: Predicting the Daily Covariance Matrix for S&P 100 Stocks Using Intraday Data - But Which Frequency to Use? *Econom. Rev.* **27**, 199–229 (2008)
- Dryden, I.L., & Mardia, K.V.: Procrustes analysis. In: *Statistical Shape Analysis, with Applications in R* (pp. 125–173) (2016). Wiley. <https://doi.org/10.1002/9781119072492.ch7>
- Dufour, J.-M.: Unbiasedness of predictions from estimated vector autoregressions. *Econom. Theory* **1**(3), 387–402 (1985)
- Engle, R.F.: Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica* **50**, 987–1007 (1982)
- Engle, R.F.: Dynamic conditional correlation: a simple class of multivariate GARCH models. *J. Bus. Econ. Stat.* **20**, 339–350 (2002)
- Engle, R.F., Kroner, K.F.: Multivariate simultaneous generalized ARCH. *Econom. Theor.* **11**, 122–150 (1995)
- Fan, J., Li, Y., Yu, K.: Vast volatility matrix estimation using high-frequency data for portfolio selection. *J. Am. Stat. Assoc.* **107**(497), 412–428 (2012). <https://doi.org/10.1080/01621459.2012.656041>
- George, E.I., Sun, D., Ni, S.: Bayesian stochastic search for VAR model restrictions. *J. Econom.* **142**(1), 553–580 (2008)
- Guhaniyogi, R., Dunson, D.B.: Bayesian compressed regression. *J. Am. Stat. Assoc.* **110**(512), 1500–1514 (2015). <https://doi.org/10.1080/01621459.2014.969425>
- Halbleib-Chiriac, R., Voev, V.: Modelling and forecasting multivariate realized volatility. *J. Appl. Econom.* **26**, 922–947 (2011)
- Han, C., Park, F.C.: A geometric framework for covariance dynamics. *J. Bank. Finance* **134**, 106319 (2022)
- Hansen, P.R., Lunde, A., Nason, J.M.: The model confidence set. *Econometrica* **79**(2), 435–497 (2011)
- Heiden, M.: Pitfalls of the Cholesky Decomposition for Forecasting Multivariate Volatility (Heiden, Moritz, Pitfalls of the Cholesky Decomposition for Forecasting Multivariate Volatility (July 23, 2015). Available at SSRN: <https://ssrn.com/abstract=2686482> or <https://doi.org/10.2139/ssrn.2686482> (2015)
- Heston, S.L.: A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Rev. Financ. Stud.* **6**(2), 327–343 (1993). <https://doi.org/10.1093/rfs/6.2.327>
- Higham, N.J.: Computing the polar decomposition-with applications. *SIAM J. Sci. Stat. Comput.* **7**(4), 1160–1174 (1986)
- Higham, N.J.: Computing a nearest symmetric positive semidefinite matrix. *Linear Algebra Appl.* **103**, 103–118 (1988)
- Laurent, S., Rombouts, J.V., Violante, F.: On loss functions and ranking forecasting performances of multivariate volatility models. *J. Econom.* **173**(1), 1–10 (2013)
- Marron, J., Dryden, I.L.: *Object Oriented Data Analysis*. Chapman and Hall, London (2021)
- Pástor, L., Veronesi, P.: Uncertainty about government policy and stock prices. *J. Finance* **67**(4), 1219–1264 (2012). <https://doi.org/10.1111/j.1540-6261.2012.01746.x>
- Pinheiro, J.C., Bates, D.M.: Unconstrained parametrizations for variance covariance matrices. *Stat. Comput.* **6**(3), 289–296 (1996). <https://doi.org/10.1007/BF00140873>
- Reiss, M., & Winkelmann, L.: Inference on the maximal rank of time-varying covariance matrices using high-frequency data. *arXiv* (2021)
- Rousseeuw, P.J., Molenberghs, G.: Transformation of non positive semidefinite correlation matrices. *Commun. Stat. Theory Methods* **22**(4), 965–984 (1993). <https://doi.org/10.1080/03610928308831068>
- Zhang, L., Mykland, P.A., Ait-Sahalia, Y.: A tale of two time scales. *J. Am. Stat. Assoc.* **100**(472), 1394–1411 (2005). <https://doi.org/10.1198/016214505000000169>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.