Neural Computing and Applications
https://doi.org/10.1007/s00521-024-09574-5

ORIGINAL ARTICLE

On the empirical performance of different covariance-matrix forecasting methods

Rafael Torres[1] · Marcelo Villena[2]

Received: 24 September 2023 / Accepted: 22 January 2024

Abstract

In the study of financial time series, covariance/correlation matrices play a central role in risk-related applications, including financial contagion and portfolio selection. Different methodologies have been used in their prediction, from methods based on Financial Econometrics DCC-GARCH (Engle in J Bus Econ Stat 20(3):339–350, 2002), to others linked to Ecophysics like Random Matrix Theory (Wang et al. in Comput Econ 51:607–635, 2018), and more recently to Machine Learning (Fiszeder and Orzeszko in Appl Intell 51(10):7029–7042, 2021). Despite these developments, there is no state-of-the-art study that compares all these methods and assesses their predictive power in an out-of-sample setting. Indeed, in this work, we focus on measuring the out-of-sample predictive power of correlation matrices of these different statistical methods, in particular from three different fields that have converged in recent years in the analysis of financial data: Econometrics, Econophysics, and Machine Learning. Thus, using a moving window scheme, we studied the correlation matrices of 29 stock market indexes from different latitudes of the world. Among our findings, we see the relationship between the measures of Eigen Entropy found in the market, with the error found in the forecast of each method in the form of Square Forecast Error. We find that in the period from 2008 to 2022, considering 2608 moving windows, the out-of-sample error tends to converge between the different methods, highlighting the performance of DCC-GARCH.

Keywords Covariance-matrix forecasting · Random matrix theory · DCC-GARCH · Support vector regression · Random forest

1 Introduction

In the economic system, one of the most relevant issues is the allocation of resources and how economic agents make decisions in this respect, with the quantification of the risk associated with these decisions being a central issue.

In particular, it is in the financial system that transactions are undertaken that make it possible to mobilize resources from one place to another, from areas with surpluses to areas with deficits. Thus, the assessment of covariance/correlation matrices has become an area of increasing interest in the context of measuring systemic risk, financial contagion, and portfolio selection, among others [56].

Over time, various approaches have emerged from fields such as econometrics, econophysics, stochastic modeling, and more recently, Machine Learning.

Time series econometrics has explored phenomena in the behavior of financial return series, such as the presence of heavy tails associated with abnormal events, leptokurtosis (volatility clustering), asymmetries in positive versus negative returns, and asset dependencies. As a result, models have been proposed to study both returns and volatility. In particular, the DCC-GARCH model has studied the dynamics of correlations between assets over time and has taken into consideration the behavior of returns and their volatility [19].

On the other hand, econophysics has proposed to use statistical mechanics techniques to study correlation matrices. We can see applications of random matrix theory and information theory that account for noise in empirical

✉ Marcelo Villena
marcelo.villena@usm.cl

1 Universidad Adolfo Ibañez, Santiago, Chile

2 Universidad Técnica Federico Santa María, Santiago, Chile

Published online: 05 March 2024

estimations, as well as market phenomena related to the behavior of eigenvalues and eigenvectors present in these estimations [15, 16].

Finally, with the arrival of large amounts of data and high computational capacity, the option of using algorithms from the field of artificial intelligence, in particular Machine Learning, has emerged, adding the ability to detect patterns associated with the nonlinearity and chaotic behavior of data that live and interact within a complex system such as the financial system [31]. Thus, in the context of financial data, we can see applications related to the prediction of financial asset prices [48], and the selection of trading strategies [53]. Proposals have been made in the field of obtaining correlation matrices and their forecasting. The use of applications based on support vector regression has made it possible to model the elements of correlation matrices dynamically, allowing their forecast [22].

Despite these developments, there is no state-of-the-art study that compares all these methods and assesses their predictive power in an out-of-sample setting. Indeed, in this work, we focus on measuring the out-of-sample predictive power of correlation matrices of these different statistical methods, in particular from three different fields that have converged in recent years in the analysis of financial data: Econometrics, Econophysics, and Machine Learning.

It is important to notice that despite the availability of different ways of obtaining correlation matrices, several factors affect their performance and the accuracy of the estimates. For example, in the field of univariate time series estimation, the modeling of economic growth has proven that this task is dependent on variables such as the lag and the length of the time windows considered when training the models, as well as on the explanatory variables selected [54]. In this context, our approach to evaluating the performance of models from different fields is based on a fair comparison of each method in terms of out-of-sample prediction, considering 2608 moving windows of 120 days over 14 years (2008–2022).

Furthermore, given that the main focus of the study is the comparison, it is worth noting that some of the methodologies have common roots in the hypothesis behind them. For example, econometric, Random Matrix, and stochastic models share some statistical assumptions, such as the normal nature of returns. Machine Learning-based approaches, on the other hand, are defined as data-driven, which allows them to avoid the use of distributional assumptions and to be generalizable.

In relation to economic assumptions these methods follow the Efficient Market Hypothesis (EMH), which indicates in its strong form that all the information is already contained in the prices of assets; however, this pattern could be broken in periods of crises [13]. In this line, one of our findings is related to how market risk affects the out-of-sample forecast error. We see that the main errors occurred in the year 2008 period related to the post-subprime crisis, and then in 2020, related to the crisis generated by the pandemic.

In this context, some of the points made by this research are:

i. We introduced a method that allows us to choose the size of the window in which the out-of-sample error is stabilized.
ii. In order to have a fair assessment of the predictive power of the methods under analysis in predicting correlation matrices, we introduced a scheme analogous to Train/Test/Validate, where the model is trained and predicted within the window, and the matrices are evaluated against the realized matrix immediately adjacent to the moving window.
iii. Within this methodological framework, we studied methods for estimating correlation matrices from three different fields: Econometrics, Econophysics, and Machine Learning.
iv. We evaluated the relationship between market uncertainty, using the Eigen Entropy indicator as a proxy, and the out-of-sample error of correlation matrix estimation, measured as the square of the forecast error.

This article is structured following this schema. Firstly, a literature review of the different fields that have proposed methods to obtain correlation matrices is developed. Secondly, the data and the explored methods are presented, Finally, the main results and discussion are shown.

2 Literature review

The measurement of risk has a long history in humanity, with its first evidence in ancient civilizations such as the Egyptians, who measured the level of the Nile in agreement with the corresponding taxes that the farmers had to pay to the state. Implicit in this was the idea of having a way of predicting the future and, based on this, planning the amount of taxes to be levied. However, it was not until the application of mathematics to the study of probabilities that the formal study of modern risk began [3].

In this section, we will cover some of the methods associated with the estimation of an object that has been used as a measure of systemic risk [11, 45, 56], the correlation matrix.

2.1 Econometric methods

During the last decades, the volatility associated with the price fluctuation of assets has been studied by the econometric field generating a series of applications in fields like Derivatives, Trading, and Risk modeling. Thus some of the main stylized facts found in the empirical literature, such as Leptokurtosis and volatility clustering, have been explored from the perspective of time series econometrics beginning with the autoregressive conditional heteroskedasticity (ARCH) model [17] and then with the Generalized ARCH model, GARCH [5], models based on a generative process that permits to model a changing variance in a univariate time series of a financial asset.

Since then, the Generalized Autoregressive Conditional Heteroskedasticity (GARCH) models have been used for the modeling of volatility in financial returns, showing good performance in the capture of the properties of financial time series as volatility clustering, asymmetry, and fat tails. An important characteristic of the GARCH model is the way that manages both, the unconditional and the conditional volatility, determined by the entire sample and for the most recent observation correspondingly [14].

However, the different assets in a market are not alone and have relationships that need to be properly captured. In the case of multivariate time series, the Multivariate GARCH (MGARCH) models are of great importance, but their solutions require the calculation of some parameters that grow with the number of assets, and for this reason, their specifications are considered as a series of constraints. One of the most important MGARCH models is the Dynamic Conditional Correlation GARCH (DCC-GARCH), which allows the estimation of correlation matrices that vary over time. This model is estimated using maximum log-likelihood estimation. As mentioned above, time-varying correlation matrices are a key component of portfolio management and portfolio optimization [19], which has fostered the empirical investigation of these types of econometric techniques.

2.2 Econophysics methods

In the study of large particle systems, Random Matrix Theory (RMT) arises, based on the statistical properties of matrices whose elements are random variables, describing the possible microscopic states of a system [35]. This tool has different applications and has been adopted by Econophysics, an interdisciplinary field that mixes elements of Physics and Economics, for the study of economic systems.

In the field of finances, the RMT has been used to understand the interactions between stock indexes at the system level by the study of the statistical behavior of the eigenvalues from empirical correlation matrices and the main factors that influence it. The RMT has shown that a big percentage of eigenvalues are below the limits predicted by it and has shown many of them are dominated by the noise, and only a few of them are statistically significant and would really contain information [32].

Besides this object, the correlation matrix has been used in financial models like Portfolio optimization [37] and has been studied because of its changing nature in time [21] implying that models that use it need to consider this changes.

However, there are two aspects of this element that make its evaluation difficult, firstly its unobservable nature and secondly the existing noise in the sample correlation matrices [34]. Considering these issues, different approaches have been proposed to treat and obtain correlation matrices in the context of market behavior.

Following this line, the changes in the returns of the top 1000 stocks in the US market over a two-year period were studied, finding that the correlation matrices of the series have universal properties of random matrices and that the components of the eigenvectors vary over time [40].

Regarding the distributions of the correlation coefficients of the correlation matrix of the market indexes, the DJIA and DAX indexes (both indexes composed of 30 stocks) were studied at the same time, and it was observed that the distribution of the correlations showed a normal distribution behavior. Likewise, it was observed that the distribution of the DJIA index had a lower variance, as well as a smaller separation between the highest eigenvalue and the lowest ones, giving a possible explanation that the DAX index is more influenced by the market (represented in external news text quoted left), which makes the components evolve more coherently. They also observed that drawdowns and draw-ups behave differently; the former are dominated by a strong collective eigenvector with large eigenvalues, the opposite for the other case where the decrease of the larger eigenvalues is simultaneous with the increase of the smaller eigenvalues (so that the trace is preserved) [15].

Continuing with the study of the correlation matrices coming from financial series by means of RMT, it was proposed to clean these matrices from noise by subtracting the effect of the largest of the eigenvalues, recognized as the one that projects the effect of the market on the assets. The authors realize that the use of correlation matrices without clean leads to an underestimation of risk since it is over-invested in eigenvectors with an artificial low risk [33].

In reference to the analysis of the information contained in the eigenvectors associated with the significant eigenvalues, which follow the market eigenvector, it has been shown that these contain information that allows classifying the assets in such a way that they group in different sectors of the economy, and that in addition, these structures are stable over time [25].

In addition, in the study of eigenvectors from correlation matrices, the influence of the market over the assets that make up a stock index has also been analyzed, and it has been observed that the largest of them represents this factor. In this way after the removal of this effect was analyzed the composition of eigenvectors, showing their contribution to several business sectors [26].

In an attempt to unravel the relationships between markets, the nonsynchronous information between them was studied, i.e., without corrections due to the time-zone difference, generating a taxonomy from their correlation coefficients, finding an important geographical component in the generation of the groupings [6].

Similarly, the effect of synchrony between two markets with time-zone lags on the correlation matrix was investigated and it was observed that if these lags are properly treated they make both markets behave as one (an eigenvalue emerges that dominates the rest), and that one of them has more powerful effects on the other [16].

With respect to the behavior of the largest eigenvalues that deviate beyond the theoretical predictions of the RMT, it has been found that few of them are stable over time and that the significant components associated with them vary according to the time period, resulting in a dynamic behavior [41].

Regarding situations in which the RMT predictions are questionable, it was found that for covariance matrices generated from data produced by a nonsymmetric elliptic distribution with heavy tails, the Marchenko–Pastur law does not correctly detect those eigenvalues that represent noise by assigning an overestimated weight to the largest of them. This, however, can be corrected by substituting the sample covariance matrix with the spectral estimator [23].

On the other hand, it has been suggested that nonstationarity may affect the eigenvalues, a situation that has been associated with some of the stylized facts of financial series such as long-range dependencies and the change of covariance structure. Thus, the RMT would not be able to correct for these situations [38].

In spite of these observations, RMT has been shown to be a good way to select the optimal number of principal components in the context of multivariate volatility models giving to the problem the relevant information [46].

Associated knowledge has been developed from correlation structures, permitting to visualization of the relationship between assets from a Complex Network perspective. The use of Minimum Spanning Trees was proposed for first time generating a distance metric from correlation matrices [36].

Additionally, these relationships have been studied by the use of network methods in the context of RMT, permitting the evaluation of changes produced in a network before and during crises. In this way, changes were observed in topological measurements and the shapes of trees and hierarchical clustering generated from 20 global financial indices, before and after the 2008 crisis [30, 52].

From this brief review, we can see that RMT has been used in a series of applications being a fruitful field of study for quantitative finance.

2.3 Machine learning methods

The previous approaches considered, in some way, are models based on probability statistics theories and in a data generation process (DCC-GARCH) that has its main assumptions in the normal (symmetric) nature of the return stocks behavior, a questionable fact that has been a remarkable state in the named "stylized facts" of financial time series.

These stylized empirical facts are common properties that can be observed "across many instruments, markets, and time periods" [13]. Between those associated with asset returns, we can appreciate (i) Absence of correlations, (ii) Heavy tails, (iii) Gain/loss asymmetry, (iv) Volatility clustering. The importance of these patterns points to the selection of parametric distributions that usually models use to characterize the data behavior to obtain accurate forecasts.

One way to avoid this parametric selection is the use of models from the field of Artificial Intelligence. In the context of financial data, mainly have been used supervised Machine Learning models, whose characteristics aim at establishing a data-driven, self-adaptive, and nonparametric process [31]. The main objective of these models is to provide investors with predictions that allow them to consistently earn above-market returns by breaking the so-called efficient market hypothesis.

Artificial neural networks (ANNs) have been used for stock market prediction applications. Thus, by using a Feed-Forward Neural Network with an output layer, the authors evaluated the effect of selecting nodes within the hidden layer by training the network with a backpropagation algorithm. They observed high competition among the different alternatives, concluding that the parameterization of these models (i.e., the selection of the number of nodes and the number of hidden layers) is an entire topic in itself [43]. Alternatives to the training algorithm for ANNs have been proposed. In financial forecasting, an evolutionary virtual data position (EVDP) was proposed which reduces noise and enhances the accuracy of forecasts [39].

Applications that relate multivariate time series models and multivariate artificial neural networks (MANNs) also have been developed to forecast volatility. Thus, a hybrid DCC-GARCH(1,1)-MANN was used to calculate the volatility between five assets from different world stock markets [KSE-100 (Pakistan), BSESN(India), S &P500 (USA), FTSE-100 (UK), KLSE (Malaysia)], showing a competitive performance in relation with DCC-GARCH, while authors mention that are necessary additional testing to a further extensions [20]

In relation to the capacity of generalization of Machine learning models, the forecast of financial time series was evaluated between the Support Vector Machine and a Multilayer Perceptron, seeing that SVM has a better out-of-sample performance, besides a faster training [10].

In another application of ML related to the prediction of the change direction of financial assets was used a Random Forest algorithm, seeing that treating the problem of forecasting as a classification problem is more feasible than predicting exact prices because of the chaotic nature and the volatility of its [29].

Following with Random Forest application, in a near study was evaluated a Random Forest model vs XGBOOST in the classification problem related to the sign of the return. Besides was analyzed the dependence of the accuracy of models as function of the trading window, seeing that for bigger windows the accuracy of both models increases [1]. The use of Random Forest also has shown interesting results about the integration of Global Markets, showing that models that use other global market for the forecast of a particular global market improve the forecast of those models that considers autoregressive past lags and technical information [49].

Many applications of Machine Learning models focus on predicting prices, returns of financial assets, macroeconomic variables, and fundamental indicators. Support vector regression (SVR) is a model that has shown good performance in time series prediction as it can handle the nonlinear properties of time series. In particular, one way to treat and forecast multivariate series from covariance matrices is by allowing the obtaining of positive definite matrices from disjoint series. This is done by using a process that decomposes covariance matrices using the Cholesky decomposition, generating Cholesky factors for each entry of the covariance matrices. An autoregressive SVR model can then be applied which transforms the multivariate problem into a univariate problem [22].

3 Data and measures

3.1 World stock indices

We focus our investigation on the statistical relationship among 29 world stock Indices that include different geographical locations, such as Africa, Asia, Europe, Latin America (LATAM), Oceania, and North America. We use daily data considered the period 2008–2022; its description is presented in Table 5.

We calculated the log daily returns of each asset i, denoted r_i, which were obtained from their prices at time t, $P_i(t)$ by means of the relation:

$$R_i(t) = \ln P_i(t) - \ln P_i(t-1) \quad (1)$$

The descriptive statistics for the daily normalized returns of the indices are shown in Table 6. We can observe the differences in the individual indices statistics considered. Positive and negative values for the Average Returns, Skewness, and Kurtosis reflect a variety of different behaviors between them in the time span.

3.2 Moving windows selection

Before making comparisons between the different methods of estimating the correlation matrices, it was necessary to define the training and testing scheme from which they are obtained, choosing between a sliding window or an expanding window scheme [27]. This selection depends on the time series properties across time. The dynamics of correlation through time have been studied frequently using a sliding moving window schema.

In this way, correlation matrices have been studied for the problem of the dynamic of correlation over time [44]. Here the authors used a time sliding moving windows schema, where each considers 400 trading days across a period of 12 years, considering for the window indexes $[t, t+w-1]$, where t varies between $[1, T-w+1]$, with T being the total length of the series. In addition, it is necessary to take into consideration two other aspects that influence the forecast problem (i) the window size (w) selection and (ii) the periods to forecast ahead (K). This has been studied in a forecast problem in the Macroeconomics field showing its importance [54]. Also in the field of Finances have been analyzed the importance of size window selection to generate realized covariances [9].

In our work, according to [54] we separate the decision in two steps:

- Slide window size selection.
- Comparison of models in relation to the K periods ahead.

Thus, for each combination of window size and prediction horizon, the forecast error was evaluated using the mean absolute error (MAE) and mean squared error (MSE) metrics, finally selecting the combination where we see a stabilization of this error.

3.3 Statistics measurements

For each of the different correlation matrices obtained in each time window, the following indicators were calculated for the correlation coefficients [12]: Mean:

$$\bar{\rho} = \frac{2}{N(N-1)} \sum_{i<ij} \rho_{ij} \tag{2}$$

Variance:

$$\bar{\lambda_2} = \frac{2}{N(N-1)} \sum_{i<ij} (\rho_{ij} - \bar{\rho})^2 \tag{3}$$

Skewness:

$$\bar{\lambda_3} = \frac{2}{N(N-1)\lambda_2^{2/3}} \sum_{i<ij} (\rho_{ij} - \bar{\rho})^3 \tag{4}$$

Kurtosis:

$$\bar{\lambda_4} = \frac{2}{N(N-1)\lambda_2^{2}} \sum_{i<ij} (\rho_{ij} - \bar{\rho})^4 \tag{5}$$

4 Methods for the correlation matrix estimation

4.1 Pearson correlation matrix

Pearson's correlation is an indicator that arises to understand the strength of linear association between two variables. Thus for two series of returns $R_i(t)$ and $R_j(t)$ it will be calculated according to:

$$c_{ij} = \frac{\langle [R_i(t) - \langle R_i(t) \rangle][R_j(t) - \langle R_j(t) \rangle] \rangle}{\sigma_i \sigma_j} \tag{6}$$

where σ_i and σ_j correspond to the standard deviations of the returns of two ETFs i and j at time t, and $\langle \cdot \rangle$ to the mean. Thus for a matrix with N series of ETFs, the total number of elements ρ_{ij} to calculate will be $N(N-1)/2$, which will populate the correlation matrix $\mathbf{C}$.

4.2 Random matrix theory

The Random Matrix Theory is a tool that permits the separation of the information present in correlation matrices from noise. The hypothesis that assumes the RMT is the zero information hypothesis [47].

Thus, for random matrices with entries from a normal distribution, where the number of rows (T) and the number of columns (N) tend to infinity, and the ratio between them $Q^* = T/N \geq 1$, the probability law for their eigenvalues was defined the Marchenko–Pastur law [33]:

$$\Pr(\lambda) = \frac{Q^*}{2\pi\sigma^2} \frac{\sqrt{(\lambda_{\max} - \lambda)(\lambda - \lambda_{\min})}}{\lambda} \tag{7}$$

modifying the limits $\lambda_{(\max/\min)}$ as:

$$\lambda'_{(\max/\min)} = 1 + \frac{1}{Q^*} \pm 2\sqrt{\frac{1}{Q^*}} \tag{8}$$

Regarding the rule to select the significant eigenvalues, have been proposed other methods that target those eigenvalues with economical meaning [46] and methods related to information theory [28]. In accordance with the results of previous work, there were selected those matrices that consider the three bigger eigenvalues [50]. In this way, the top three eigenvalues were selected and the correlation matrices were filtered as shown in [47].

Thus, to obtain a filtered correlation matrix, this is diagonalized and their eigenvalues are ordered by selecting the S largest than the value $\lambda_{\max}$ previously determined by Eq. (7).

Then, to keep the part of the correlation matrix associated with these eigenvalues, a filtered diagonal matrix is constructed, whose elements are as follows:

$$\Lambda' = \begin{cases} 0, & \text{if } 1 \leqslant i < N - S \\ 1, & \text{if } N - S \leqslant i \leqslant N \end{cases} \tag{9}$$

In this work, the case for $S = 3$. Finally, a correlation matrix $\mathbf{C}'$ is obtained by applying the transformation Λ', and the elements of the diagonal are set to one.

4.3 DCC-GARCH

The Generalized Autoregressive Conditional Heteroscedasticity models (GARCH) are supported by the definition of two equations: to both, (i) the returns of assets () and (ii) the behavior of its volatility ().

$$r_t = E[r_t | I_{t-1}] + \epsilon_t = \mu_t + \epsilon_t \tag{10}$$

$$\sigma_t^2 = E[r_t^2 | I_{t-1} - \mu_t^2] = E[\epsilon_t^2 | I_{t-1}] \tag{11}$$

For the multivariate version, DCC-GARCH model [19, 46] (dynamical conditional correlation model) is the assumption of normality of the considered assets with a covariance matrix V_t. Then, the parameter estimation consists of two steps. In the first step, the different returns associated with each asset are modeled as a univariate GARCH(1,1) model obtaining its individual parameters. In the second step, the standardized residuals obtained from the first step are used to estimate the conditional covariance matrices.

The covariance matrix V_t is decomposed as:

$$V_t = D_t R_t D_t \tag{12}$$

with $D_t = \mathrm{diag}(\sigma_{it}, \ldots, \sigma_{Nt})$ as the diagonal matrix that contains the standard deviation of asset returns. R_t is the correlation matrix. The elements of D_t, σ_{it} are modeled as a univariate GARCH process that in our study was a GARCH(1,1) given by:

$$\sigma_{it}^2 = \alpha_{0i} + \alpha_{1i}\epsilon_{it-1}^2 + \beta_{1i}\sigma_{it-1}^2 \tag{13}$$

The standardized residuals from GARCH(1,1) estimation are used to estimate the correlation specification of DCC-GARCH(1,1):

$$R_t = Q_t^{*-1} Q_t Q_t^{*-1} \tag{14}$$

and

$$Q_t = (1 - \lambda_1 - \lambda_2)\bar{Q} + \lambda_1 e_{t-1} e_{t-1}^{\mathsf{T}} + \lambda_2 Q_{t-1} \tag{15}$$

where $\bar{Q} = E[e_t e_t^{\mathsf{T}}]$ is the unconditional covariance and $e_t = D_t^{-1}\epsilon_t$ are the standardized errors, and Q^* is a diagonal matrix that contains square root of the diagonal elements of Q_t, $Q^* = (\mathrm{diag}(Q_t))^{1/2}$.

In our study, DCC-GARCH simulations were made by the use of R programming, particularly the *rmgarch* package [24].

4.4 Univariate modeling using Cholesky decomposition

In this work, we use the methodological approach used in [22] to test the behavior of models that do have not a multivariate specification to the estimation of correlation matrices. Thus, the Cholesky factors obtained from the Correlation Matrix Cholesky Decomposition were modeled as univariate time series using this model:

- Autoregressive.
- Autoregressive moving average (ARMA).
- Support vector regression (SVR).
- Feed-forward neural networks (AANN).

The general procedure to treat Correlation Matrices and Forecast their elements in an out-of-sample schema considers the following steps:

1. Decomposition of covariance matrices of returns into Cholesky factors
2. Forecast the univariate series of entries of Cholesky factors using a selected model {ARMA, Brownian Motion, SVR, Recurrent Neural Network }
3. Reconstruct covariance matrix from forecast using a reverse operation of Cholesky decomposition.

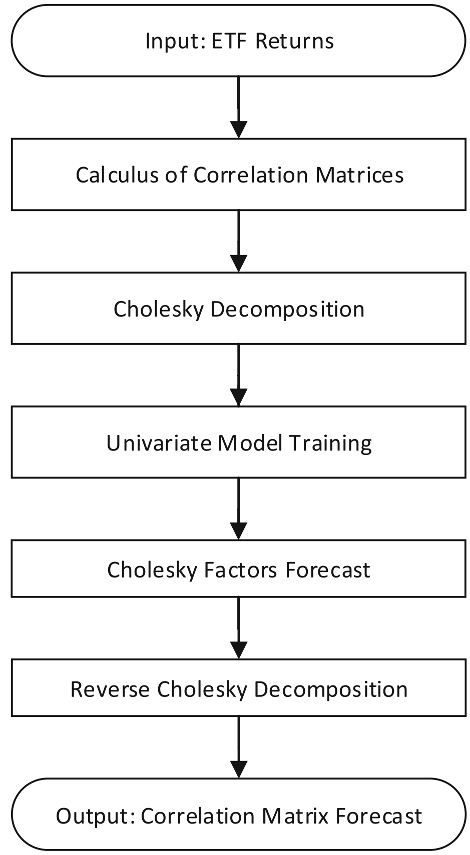


Fig. 1 The algorithm used to test the univariate models to forecast Correlation Matrices of returns

We can visualize the process in Fig. 1.

4.4.1 Mixed autoregressive-moving average models (ARMA)

The study of time series contemplates the study of the dependence of data in relation to past occurrences. Autoregressive models are a stochastic model that considers current observation as a linear function of past p observations, taken a denomination of autoregressive process of "p" order, AR(p) defined as:

$$y_t = \phi_1 y_{t-1} + \phi_1 y_{t-2} + \cdots + \phi_p y_{t-p} + z_t \tag{16}$$

where the term z_t is a random error. In the same way have been defined Moving Average(MA) models where the dependent variable has a dependence on the error terms, z_t

$$y_t = z_t - \theta_1 z_{t-1} - \theta_2 z_{t-2} - \cdots - \theta_q z_{t-q} \tag{17}$$

In this way, the ARMA model considers both processes, (16) and (17), to have a more flexible model for the current time series value, y_t [7]. This model is defined for a stationary time series.

4.4.2 Support vector regression

The support vector regression is an extension of the Support Vector Machine method, used to regression problems [51]. It has been applied to forecast financial series like stock indices, stock prices, and volatility indices. According to [22] we first define a regression model:

$$y = r(\mathbf{x}) + \delta \tag{18}$$

where $\mathbf{x}$ is a the vector of regressors and δ is additive zero-mean noise with variance σ^2, and $r(\mathbf{x})$ is a regression function. Considering a training dataset we want to approximate a regression function that deviates ϵ from the outputs. In the SVR the input $\mathbf{x}$ is mapped on a high-dimensional feature space using a nonlinear function, and then a linear model it is constructed in this feature space:

$$f(\mathbf{x}) = \sum_{i=1}^{d} \omega_i \psi_i(\mathbf{x}) + b \tag{19}$$

where ω_i are coefficients, $\psi_i(\mathbf{x})$ are transformations and b a bias term. To determine $f(\mathbf{x})$ it is necessary to determine the coefficients ω_i. To do it was proposed a ϵ-insensitive loss function L_ϵ. So, this L_{epsilon} will compare its output with the testing set y, and in those cases where the absolute difference will be greater that ϵ there will be a penalization, no penalty in the other case:

$$L_\epsilon = \begin{cases} 0, & |y - f(\mathbf{x})| \le \epsilon \\ |y - f(\mathbf{x})| - \epsilon, & \text{otherwise} \end{cases} \tag{20}$$

Finally it is calculated an optimal regression function considering:

$$\Psi(\omega, \xi) = \frac{1}{2}||\omega||^2 + C\sum_{t=1}^{n}(\xi_t + \xi_t^*) \tag{21}$$

where C is a constant and ξ_t and ξ_t^* are slack variables [22]. In this work was used a Radial Kernel. Simulations were made in the R programming language using the interface to libsvm in e1070 library [42].

4.4.3 Autoregressive neural networks

The artificial neural network model is a structure that connects computational nodes, or neurons, in a series of layers. It can be classified as a semi-parametric method. Particularly, in feedforward neural networks, the neurons in the input layer are connected to other nodes present in a hidden layer, and those connect to each other until they reach the unique node in the output layer. This model can be seen as an autoregressive nonlinear model when lagged values of the time series are used in its construction, also known as an NNAR model [27] (Fig. 2).

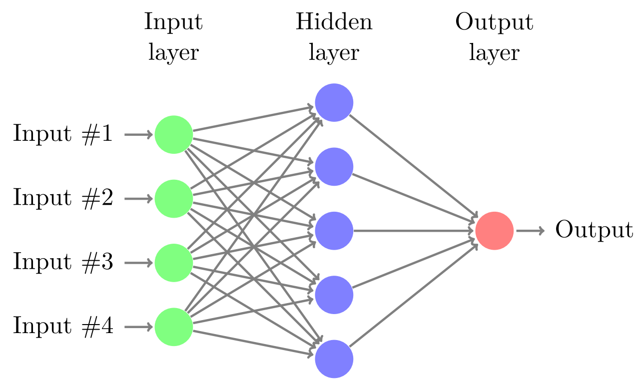


Fig. 2 The multilayer feed-forward neural network with one hidden layer

The inputs are multiplied by weights and pass to an activation function to produce an output. In this way, we can see it as a linear model where the output o can be written as:

$$o = f_o\left[\alpha_{0o} + \sum w_{jo}h_j\right] \tag{22}$$

where w_i represents the weight related to output node [4], being the $f_o(\cdot)$ a linear function. Simulations were made in the R programming language using the *nnetar* package [42].

4.4.4 Random forest

In Machine Learning, the concept of several methods working together to give a better recommendation has taken the name of ensemble learning, being the basis of procedures like the Random Forest. Particularly in the case of Random Forest, the unique model behind is the decision tree, however, the combination of prediction depends on how the different trees are structured, in this particular case from independent random vectors [8]. Among the characteristics of Random Forest are good performance prediction and flexibility, which would be used in both, classification or regression problems. The random forest classification consists of a collection of trees

$$\hat{h}(\cdot, \Theta_1), \ldots, \hat{h}(\cdot, \Theta_q) \tag{23}$$

where $\Theta_1 \ldots \Theta_q$ are i.i.d random variables. Then, the Random Forest predictor is obtained by the aggregation of the multiple Random Trees. For the particular case of a regression problem, the aggregation is defined as:

$$\hat{h}_{\text{RF}}(x) = \frac{1}{q}\sum_{l=1}^{q}(\hat{x}, \Theta_l) \tag{24}$$

Simulations were made in the R programming language using the *randomForest* package [42].

5 Out-of-sample evaluation

Once we select the moving window size, considering for the decision both, the size that stabilizes the error and a size that makes sense with the problem of the dynamic of the correlation, we need to make an estimation from each moving window and compare this forecast with an out-of-sample realized correlation, in this case the realized correlation matrix that considers the next data to the moving window equivalent to 30 days ahead.

For each method were estimated correlation matrices for the K ahead periods from the training period $[1, T - w + 1 - K]$.

Additionally, hyperparameter tuning was performed for the Machine Learning methods based on a fivefold cross-validation schema. Once the model parameters were selected, they were used to train the models during the previously mentioned training periods $[1, T - w + 1 - K]$.

6 Results

In this section, we present some of the main results obtained from the studied methods.

6.1 Moving window size

As we mention in Sect. 3.2, to have a scientific approach about the moving window size selection, it was measured the mean square error (MSE) and mean absolute error (MAE) as a function of moving window sizes (w) and forecast horizon (K).

Both metrics MSE and MAE between two matrices S and H with dimension $N \times N$ each are defined in [2]:

$$\mathrm{MSE} = \frac{1}{N^2} \mathrm{vec}(H - S)' \mathrm{vec}(H - S), \tag{25}$$

where $\mathrm{vec}(\cdot)$ operator represents column stacking operator.

$$\mathrm{MAE} = \frac{1}{N^2} abs(H - S)' \mathrm{vec}(H - S) \tag{26}$$

From Figs. 3 and 4 we can see that at the size of 120 days both errors are stabilized. When we refer to the stabilization of the error, we mean that the metrics used to evaluate this error, the difference between the current value and the previous value, decreases by less than 20%.

To give an example for MAE, if we look at the error for $K = 15$, for a time window of size $\mathrm{MW} = 45$ days $\mathrm{MAE} = 0.0628$, for MW of 60 days $\mathrm{MAE} = 0.0450$, for

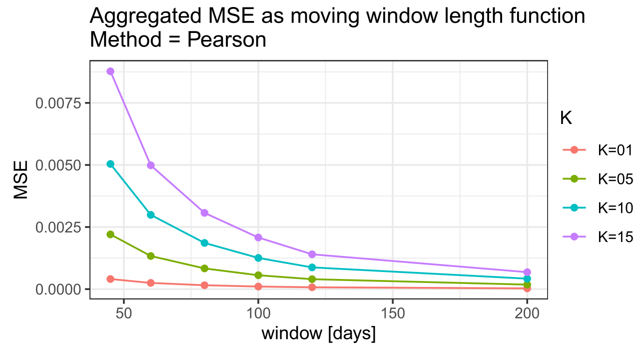


Fig. 3 The error behavior in the selection of moving window size—MSE

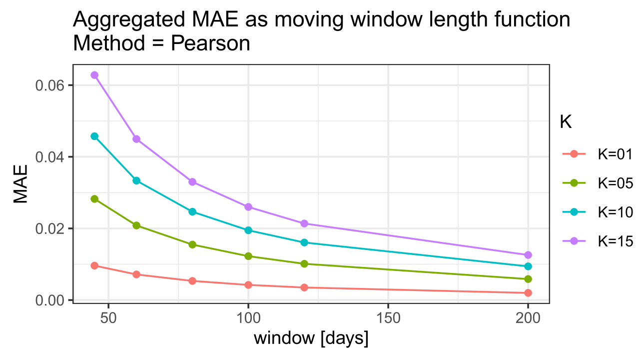


Fig. 4 The error behavior in the selection of moving window size—MAE

MW of 80 days $\mathrm{MAE} = 0.0330$, for MW 100 days $\mathrm{MAE} = 0.026$, for MW of 120 days $\mathrm{MAE} = 0.0214$ and for MW of 200 days $\mathrm{MAE} = 0.0126$. The first occasion, between $\mathrm{MW} = 45$ and $\mathrm{MW} = 60$, decreases 28.4%, between MW60 and MW80 decreases 26.6%, between MW80 and MW100 decreases 21%. The first time the decrease is less than 20% is between MW100 and MW120. We observe this for all the K ahead, Table 1.

Because of this result, the next experiments will consider only moving windows of this size. Besides the problem context makes sense to evaluate the dynamic of the correlation in the period between quarter and semester, understanding that correlations between indexes may change in this considered range of time.

6.1.1 Moving windows statistics

From the different statistical measures, we observe that the different methods used preserve the trends, even though they have variations among them. Mainly we see two areas where the average correlations increase, the first one between the periods 2012 and 2013, and the second one from 2020 onward where a peak is observed. The variance of the correlation coefficients decreases in relation to this peak of average correlation, which could suggest that a

Table 1 Stability of moving windows

Method	MW	KPI	$K = 01$	$K = 05$	$K = 10$	$K = 15$	d%_$K = 01$	d%_$K = 05$	d%_$K = 10$	d%_$K = 15$
Pearson	45	MAE	0.0096	0.0282	0.0457	0.0628				
Pearson	60	MAE	0.0072	0.0208	0.0333	0.0450	− 25.6	− 26.2	− 27.1	− 28.4
Pearson	80	MAE	0.0053	0.0155	0.0247	0.0330	− 25.4	− 25.6	− 26.0	− 26.6
Pearson	100	MAE	0.0042	0.0123	0.0195	0.0260	− 20.8	− 20.8	− 21.0	− 21.2
Pearson	120	MAE	0.0035	0.0101	0.0161	0.0214	− 17.1	− 17.4	− 17.5	− 17.7
Pearson	200	MAE	0.0020	0.0059	0.0094	0.0126	− 43.0	− 42.1	− 41.4	− 41.1

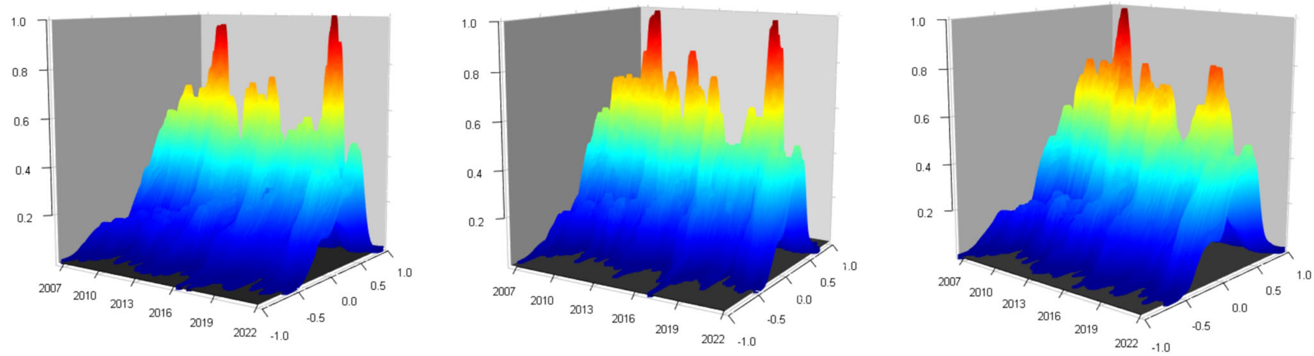
Fig. 5 Mean, variance, skewness, and kurtosis of correlation coefficients obtained from the moving windows of 120 days between 2008 and 2022 years

general increase in correlations makes these values more homogeneous. If we evaluate the skewness and kurtosis associated with this same zone where we see the mean correlation peak, we see that the skewness moves away from zero (associated with a normal distribution) and the kurtosis values become positive, indicating that the coefficients are more concentrated. Thus we see that in general, the correlation coefficients are far from normal, Fig. 5.

6.2 Dynamics of correlation matrices

One of the problems associated with the correlations is its dependence on time, which affects between others things the different models that use it as input. In a related area has been studied this behavior by the use of statistical analysis produced from correlation matrices, showing that the correlation increases in those periods of crisis [44].

In this way, using the different correlation matrices we construct from each window a probability density function (PDF) to see how they behave on time.



(a) RMT (b) SVR (c) DCC GARCH

Fig. 6 Dynamics of correlation coefficient distribution for the different models. The distribution was calculated from 120-day moving windows of returns

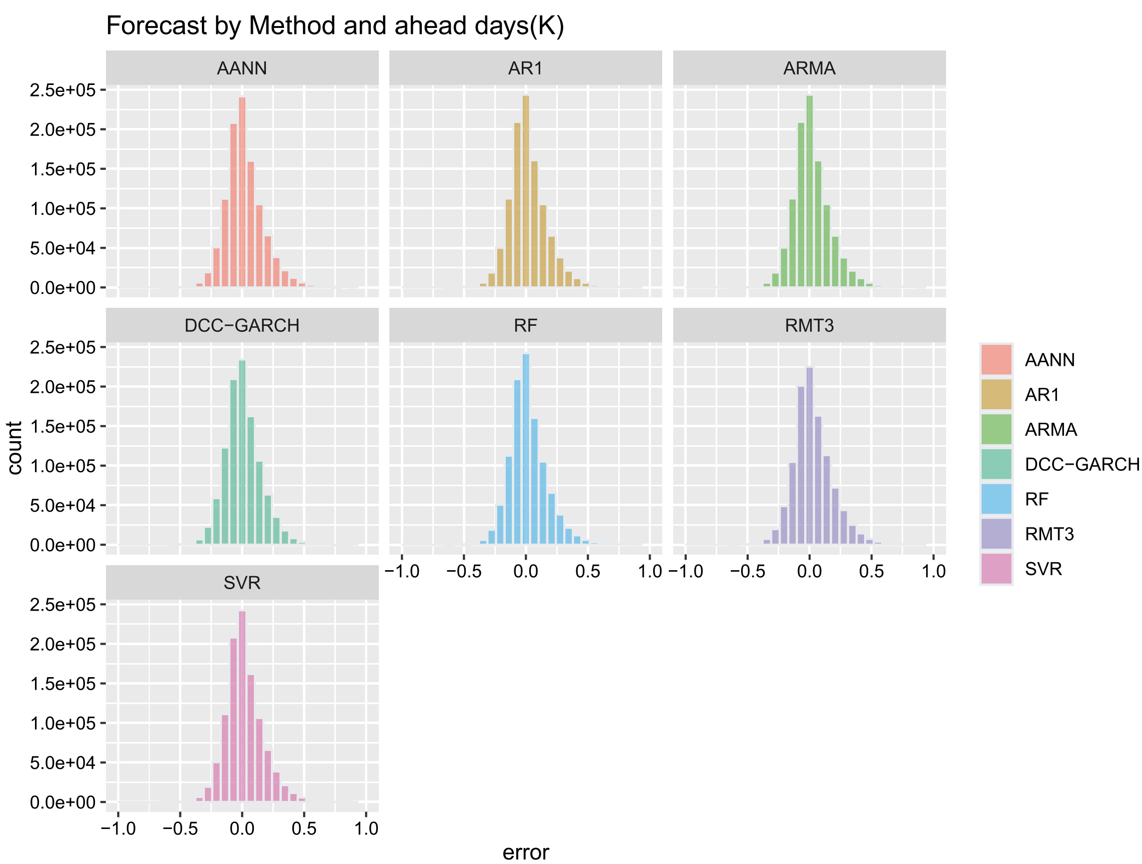


Fig. 7 Histogram with the out-of-sample error between Correlations components of estimated correlations against the realized correlation of next period. Include all ranges of time, between the years 2008 and 2022

As shown in Fig. 6a–c, the results indicate that the models differ in the way to sense the crisis periods associated with the levels of correlations.

6.3 Forecasting errors

Once we observed the difference in how methods see the correlation among indices, a question arises about how methods behave in relation to forecast error as a function of the periods ahead (K). In this way, each model forecast was compared with the realized volatility in the whole period (2008–2022) to have a first view of the error distribution.

In Fig. 7 we appreciate that there are differences, not major at least for $K = 1$, however considering the previous behavior of the correlation distribution (Fig. 6) we want to

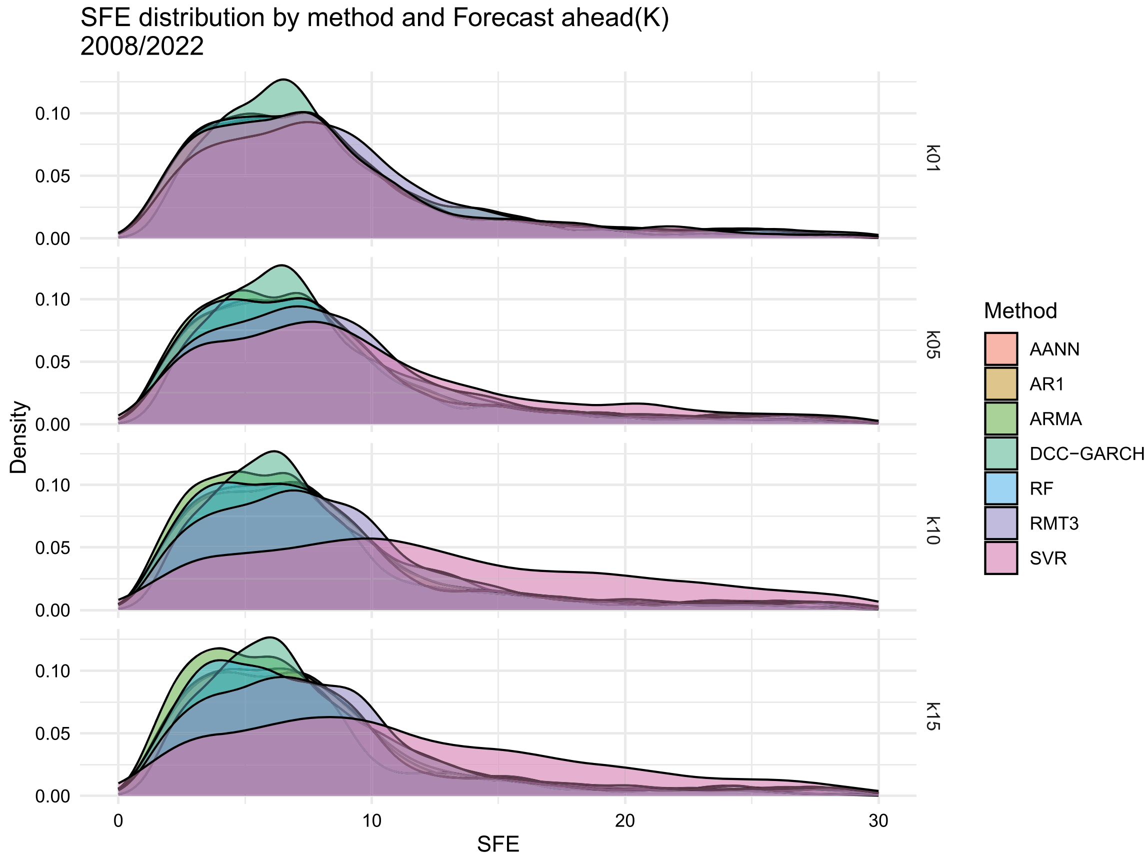


Fig. 8 The square forecast error (SFE) of different combinations between method/forecast ahead in the out-of-sample experiment

understand if there are factors such as the time period, or the K ahead considered that drive this behavior (Fig. 8).

6.3.1 SFE in the whole period

To determine the precision of the forecast and compare the behavior between selected methods we measure the out-of-sample square forecast error (SFE) for the elements σ_{ij}. In the moving window was determined the ahead forecast for k days ahead ($k = 1, 5, 10, 15$ days) that were compared with the realized correlation of the next period. In this way, as is detailed in [55] we define the indicator for each moving window mw as:

$$\mathrm{SFE}_{mw} = \sum_{i=1}^{n} \sum_{j=1}^{i} \sigma_{ij,mw} - \hat{\sigma}_{ij,mw} \tag{27}$$

With the purpose of visualizing the dispersions of the SFE error generated in the estimations, we proceeded to generate a boxplot of SFE vs the method and k ahead estimation, Fig. 9. In order to make it comparable with the previous figure, the same scale of values was considered, that is, the SFE values are between 0 and 30. With the exception of SVR, we see that the mean of the methods for all the K ahead is similar; also the bulk of values shows to have a similar dispersion. These two previous graphs give us the basis to incorporate the temporal variable in the following analysis.

Besides having a view of how a particular period of time was behaving in relation to its forecast performance, we compute the mean squared forecast error (MSFE) in the different years in which each moving window belongs in this way:

$$\mathrm{MSFE}_{\mathrm{year}} = \frac{1}{M} \sum_{y=1}^{M} \mathrm{SFE}_y \tag{28}$$

where M is the particular index year at which the moving window belongs (Fig. 10).

In Fig. 11a we appreciate levels of errors that vary from 10.47 to 8.4%, where DCC-GARCH over outperforms the other methods in the out-of-sample error. In Fig. 11b we can appreciate that between different methods there are differences in the performance of forecast through time, with similar levels; however, SVR shows a decrease in its performance for the forecast ahead $K = 10$ and $K = 15$.

In a dynamic visualization about the behavior of forecast error of methods in the different years, we can see Fig. 12. Interestingly we can appreciate that for DCC-GARCH, the MSFE values are similar to other methods for all periods

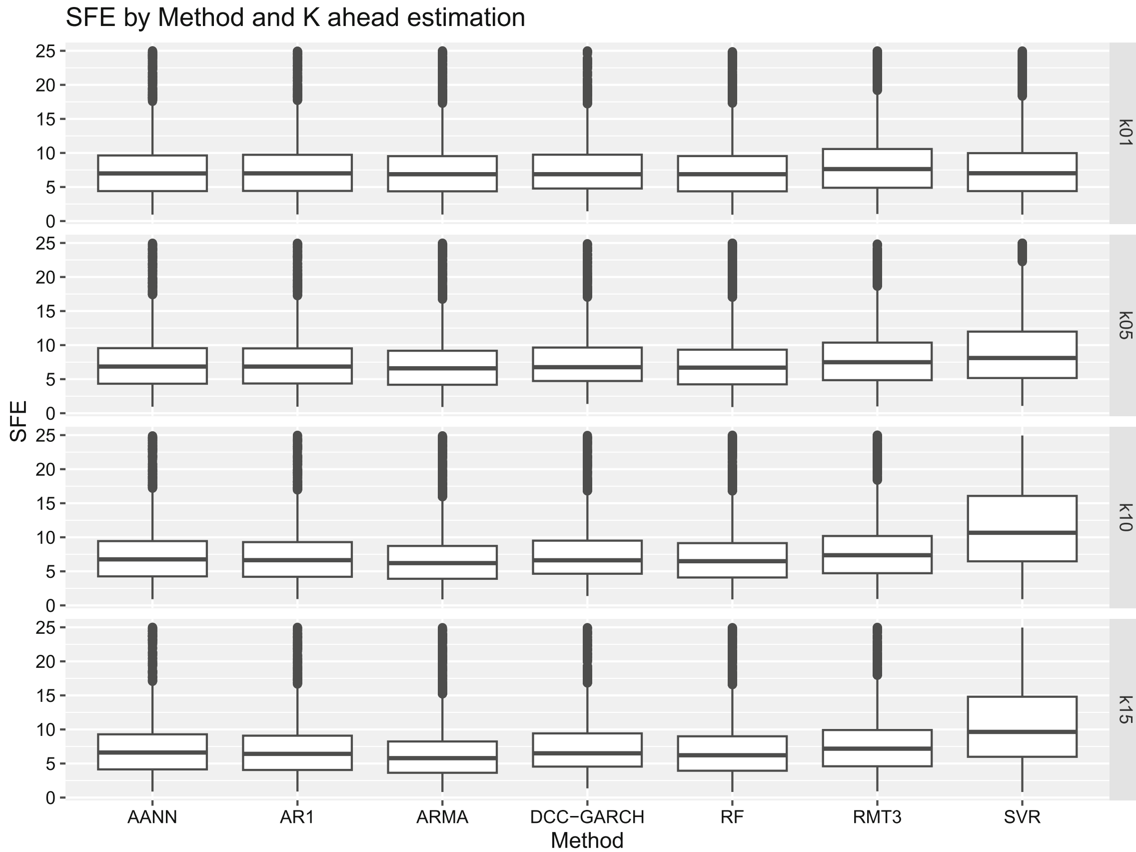
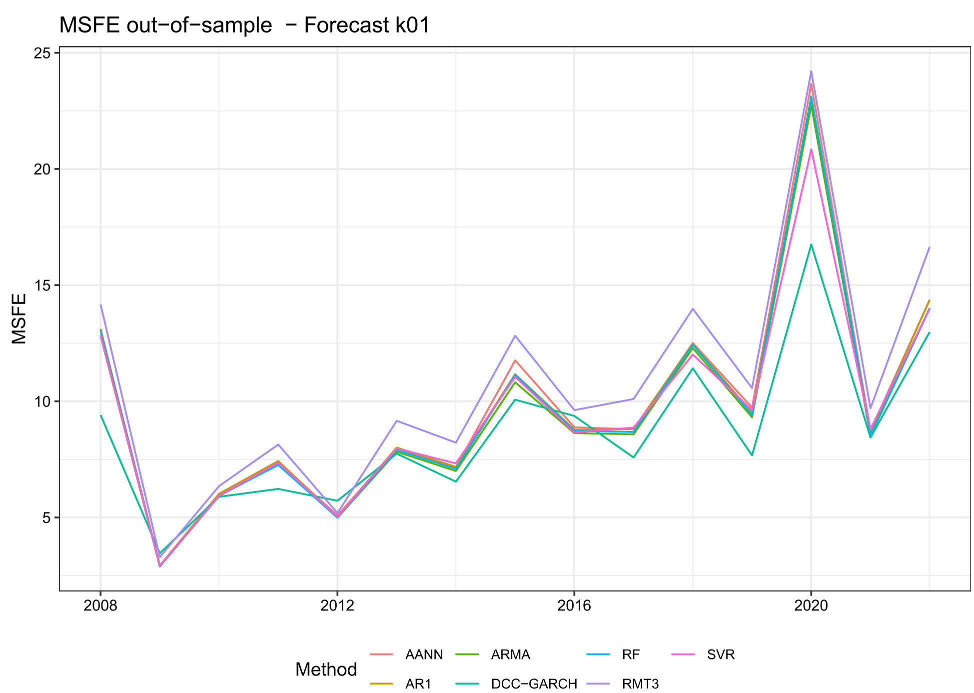


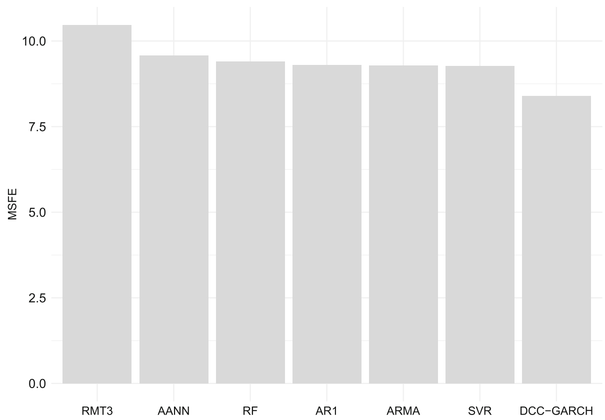
Fig. 9 Boxplots for the SFE error for different combinations of forecast horizon (K) and methods

Fig. 10 The mean square forecast error (MSFE) of different methods for a one ahead forecast ($k = 1$) in the out-of-sample experiment



with two exceptions, years 2008 and 2020; both years were the last two main crises at a global level.

After these observations, our analysis moves to understand what factors can drive these differences at the level of



(a)

metodo	k01	k05	k10	k15
AANN	9.58	9.63	9.64	9.64
AR1	9.30	8.94	8.54	8.17
ARMA	9.29	8.94	8.56	8.41
DCC-GARCH	8.39	8.30	8.16	8.01
RF	9.40	9.32	9.22	9.03
RMT3	10.47	10.39	10.28	10.18
SVR	9.27	16.44	64.39	44.01

(b)

Fig. 11 **a** The mean square forecast error (MSFE) of the different methods for 2608 moving windows between 2008 and 2022 and one ahead period estimation, $k = 1$. **b** Table with the different values of MSFE by K ahead

the decomposition of the estimated correlation matrices for the analysis of eigenvectors and eigenvalues.

6.3.2 Eigenvalues and eigenvectors analysis

In order to understand the forces driving the correlation, we studied the behavior of both the eigenvalues and the eigenvectors of the correlation matrices. In the following, we observe the Eigen Entropy index related to the eigenvalues of the decomposition of the correlation matrices $\Sigma = V\Lambda V^{\intercal}$. This is a systemic risk measurement based on correlation [11].

$$\text{Eigen Entropy} = -\sum_{i} p_i \log p_i \tag{29}$$

We can see that Eigen Entropy related to the DCC-GARCH is strongly decoupled from the other methods in 2020, but in general, we observe that the index has a similar behavior.

6.3.3 Out-of-sample error significance

Once we have analyzed the behavior of SFE distributions and the Eigenvector/Eigenvalue, we want to evaluate if the measure of out-of-sample error has a relation with measures that come from the Eigenvector. In this way, we evaluate the relation between these two variables using a linear regression. So in this way we set:

$$\log(\text{RealizedEntropy}) = \log(\text{SFE}) \tag{30}$$

We found significance for the coefficients associated with the out-of-sample forecast error (SFE) and show us are explaining. This is interesting because it provides us with an economic measure of the percent change in the forecast in relation to market conditions, which are incorporated differently by the different methods. Thus we find that the method that is least affected by the market risk condition is DCC-GARCH.

6.3.4 Temporal MST

Finally, in order to observe the changes through the different time windows, Minimum Spanning Trees were constructed from Pearson's Correlation matrices. There were considered three periods that contain the moving windows and were calculated the average correlation. The periods selected considered years 2008–2012, 2013–2017, and 2018–2022. These correlations were transferred to a measure that allowed them to be constructed according to a distance [36].

Thus to convert the distance correlation into an abstract space, we define a metric $d(i,j) = 1 - \rho^2$. This measure satisfies the three criteria for a metric:

1. $d(i,j) = 0$, if $i = j$
2. $d(i,j) = d(j,i)$
3. $d(i,j) \leq d(i,k) + d(k,j)$

From the MST generated we observe that the last period has two particularities: The first, it has a lower variability, considering it as the mixture of geographical indexes; the second, we see the reordering of the Asian components in a single united cluster; this highlights the effects of the crisis of the last period where the correlation takes the maximum value in the year 2020 (Figs. 13 and 14).

7 Concluding remarks and future research

In this research, we rigorously evaluate the predictive power of various statistical techniques for predicting correlation matrices, some older, such as econometrics, and some newer, such as those proposed by econophysics and machine learning. We also look at the conditions under

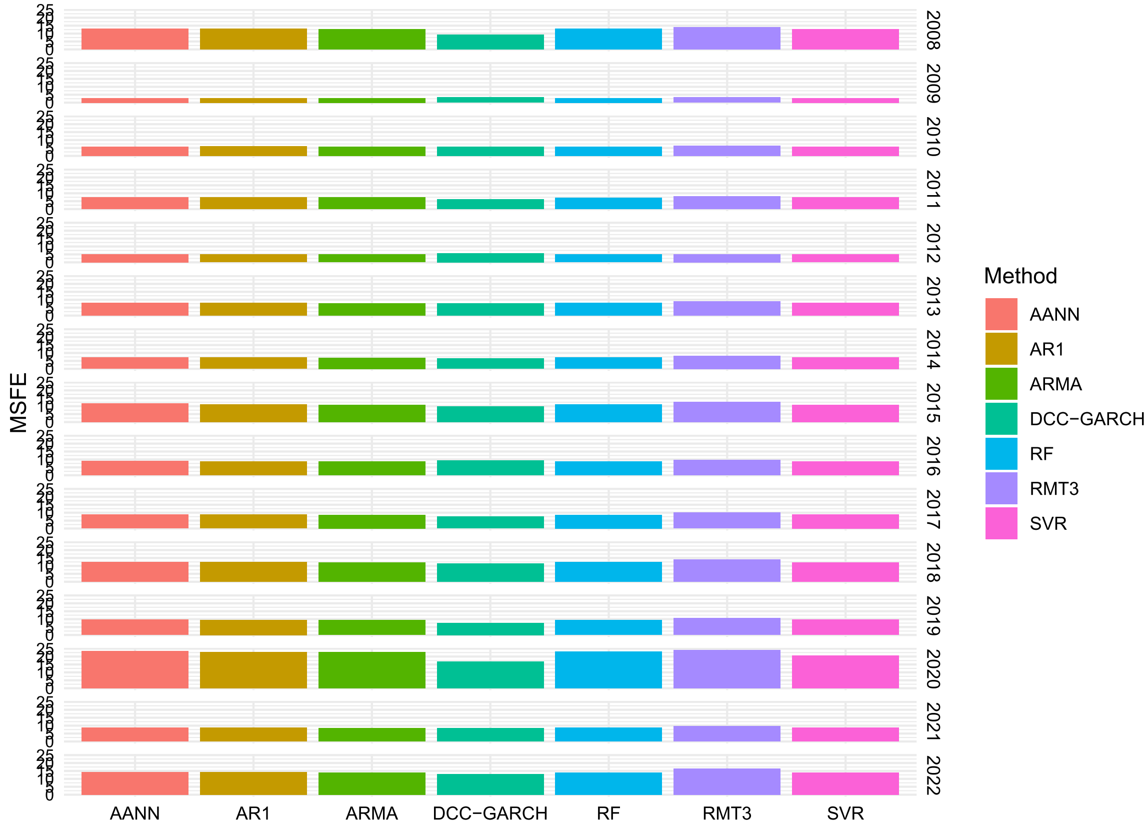
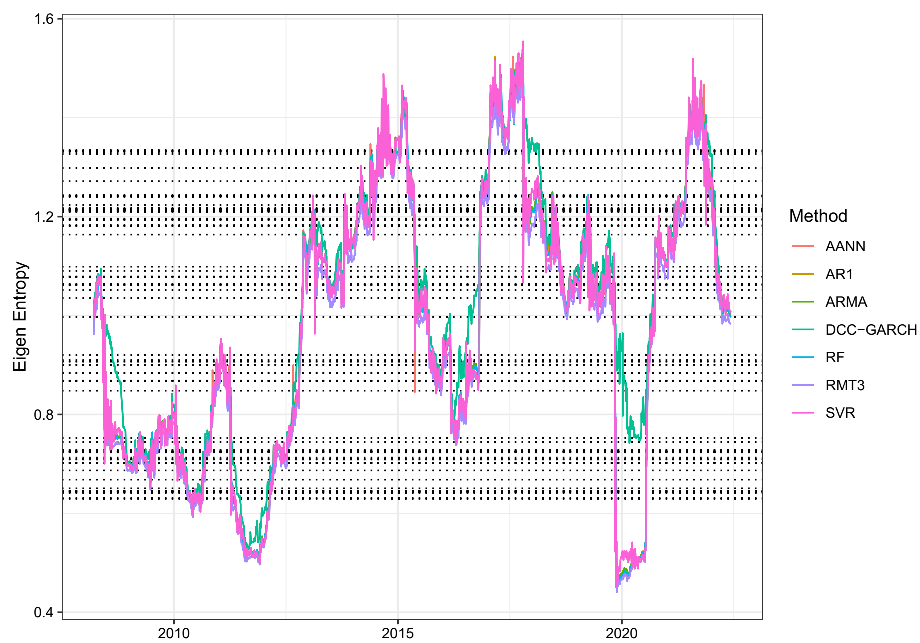
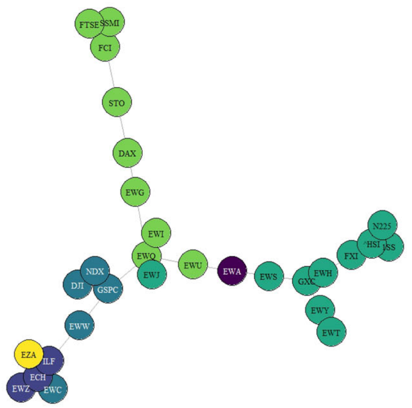


Fig. 12 The mean square forecast error (MSFE) of different methods for the 2608 moving windows between 2008 and 2022. In this case, each moving window belongs to a particular period

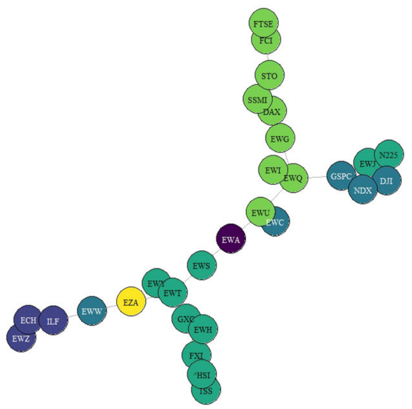
Fig. 13 Eigen Entropy calculated from the eigenvalues of correlation matrices from moving windows of 120 days



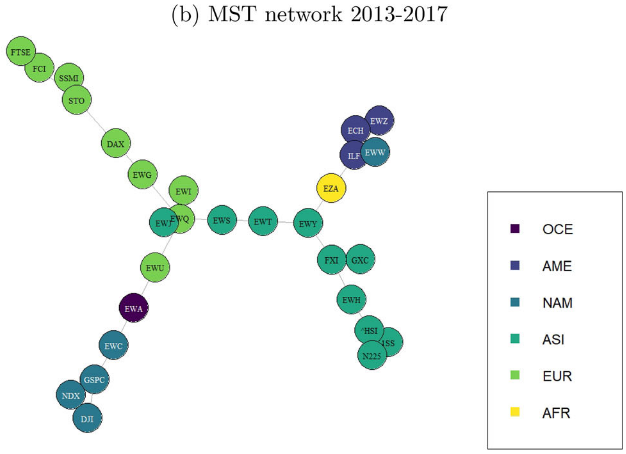
which each method works best, and the conditions under which it makes more or less the same difference which method is used. On the one hand, we test the predictive power of RM and compare it with very powerful techniques such as ANN, Random Forest, and SVM. Finally, we have seen the effectiveness of the DCC GARCH models. All these issues are undoubtedly very



(a) MST network 2008-2012



(b) MST network 2013-2017



(c) MST network 2018-2022

Fig. 14 Dynamics of MST trees generated from the moving windows divided by periods, 2008–2012 | 2013–2017 | 2018–2022. The abbreviations present in the legend should be read as follows: OCE: Oceania, AME: South America/Latin America, NAM: North America, ASI: Asia, EUR: Europe and AFR: Africa

relevant to financial forecasters who are keen to reduce the prediction error of correlation matrices.

There are also some new and interesting conclusions to be drawn from this analysis:

- At the methodological level, the window size was selected empirically. This is a problem that is not always addressed in the literature that forecast correlation matrices over moving windows. In the current literature on correlation matrix forecasting, there are many papers where temporal windows are used without describing the rationale for the choice of size. In this paper, this issue has been taken seriously in order to provide a fair and clear assessment and comparison of methods.
- Another point made in the paper is the stability of the estimates obtained with these methods. At the level of the SFE distribution, in the period 2008–2022 (Fig. 8), we see that for the forecast periods $K = 1$ and $K = 2$ the behaviors look similar for all methods under analysis. Nevertheless, for $K = 10$ and $K = 15$ there are methods whose errors have wider tails (SVR). This is undoubtedly an interesting topic for future research.
- In relation to the dynamics of the MSFE Fig. 12, we can see that except for the years 2008 and 2020, the errors have very similar levels between the different methods. However, in those years, for DCC-GARCH, the MSFE values are notoriously lower. This could be shedding light on a structural component (nonlinearity) of DCC-GARCH that is capturing the risk signal in a better way.
- Regarding the study of the properties of eigenvalues and eigenvectors from the decomposition of correlation matrices, the Eigen Entropy's indicator [11] has been used as a systematic risk indicator. What is known about eigenvalues and how they have been interpreted from RMT studies is that, generally, the largest of them represents the strength of the market. When there is a crisis the whole market falls and becomes dominant. If all the eigenvalues were equal we would have no correlation among the elements of the correlation matrix (i.e., a matrix with entries from a normal distribution), if we had only one eigenvalue the Eigen Entropy would be minimal.

An important finding of our research is that, from a statistical point of view, these eigenvalues explain the variance of the system. How do we understand the relationship between Eigen Entropy and out-of-sample error, expressed as SFE? A percentage change in the estimation error can be explained by the variability of the eigenvalues measured by the Eigen Entropy in the realized correlation, which explains the risk present in the market. In this case, we see that the lowest prediction variability is observed for DCC-GARCH. Thus, a 1% change in the Eigen Entropy measured on the data, the environmental state of the system captured by the chosen index (the realized correlation matrix), is associated with an increase in the estimation error of

Table 2 Linear regression summary for the relationship between market entropy and the forecast error (SFE) for the different methods

	Dependent variable						
	log(Entropy)						
	AANN (1)	AR1 (2)	ARMA (3)	DCC-GARCH (4)	RF (5)	RMT3 (6)	SVR (7)
log(sfe)	0.205***	0.207***	0.207***	0.161***	0.208***	0.230***	0.210***
	(0.007)	(0.007)	(0.007)	(0.009)	(0.007)	(0.006)	(0.007)
Constant	− 0.445***	− 0.444***	− 0.444***	− 0.357***	− 0.448***	− 0.519***	− 0.452***
	(0.014)	(0.014)	(0.014)	(0.018)	(0.014)	(0.014)	(0.014)
Observations	2608	2608	2608	2608	2608	2608	2608
R^2	0.263	0.263	0.262	0.116	0.268	0.327	0.268
Adjusted R^2	0.263	0.262	0.261	0.116	0.268	0.327	0.268
Residual Std. Error ($df = 2606$)	0.286	0.286	0.286	0.313	0.285	0.273	0.285

*$p<0.1$; **$p<0.05$; ***$p<0.01$

0.161***, 0.205***, 0.207***, 0.207***, 0.208***, 0.210***, and 0.230*** for DCCGARCH, AANN, AR(1), ARMA(1,1), RF, SVR, and RMT3 correspondingly, Table 2.

Looking at the MST trees for the average correlations corresponding to the 2008–2012, 2013–2017, and 2018–2022 period windows we appreciate differences in the shape and manner in which the markets are connected. We also see how the relationship of each asset with respect to geography becomes stronger in the period containing the year with the highest correlation, 2020. Thus, all Asian countries are linked in the same grouping. A very interesting future research question that emerges from this graphical analysis is the extent to which the correlation matrices are autoregressive and/or stationary. This line of research has not yet been explored.

Appendix

Appendix A.1: Hypertuning of machine learning models

We performed hyperparameter tuning for the Machine Learning models, that is, autoregressive neural network (ANN), random forest (RF), and support vector regression (SVR), using a cross-validation scheme, where the data were first divided into Training and Testing sets in an 80%/20% proportion, doing validation/hyperparameter selection in the Training set and then testing the models on the Testing set.

So, for each rolling 120-day time window we generated a set of related Cholesky factors, generating series for each of the Cholesky factors from the correlation matrices of $N \times N$, a total of 435 factors $(N + (N - 1))/2 + N$, with 2608 observations each.

For the ANN model we did a grid search for hidden units and num networks, considering consecutive values hidden unit = (1, 27) and num networks = (20, 50).

For the case of RF we did a grid search for mtry and trees, considering consecutive values mtry = (1, 30) and trees = (1, 3000).

Finally for the case of SVR we did a grid search for cost, sigma, and margin considering consecutive values cost = (1, 5) rbf_sigma = (0.1, 0.5), and margin = (0.1, 1).

For the three methods, a fivefold cross-validation scheme was used.

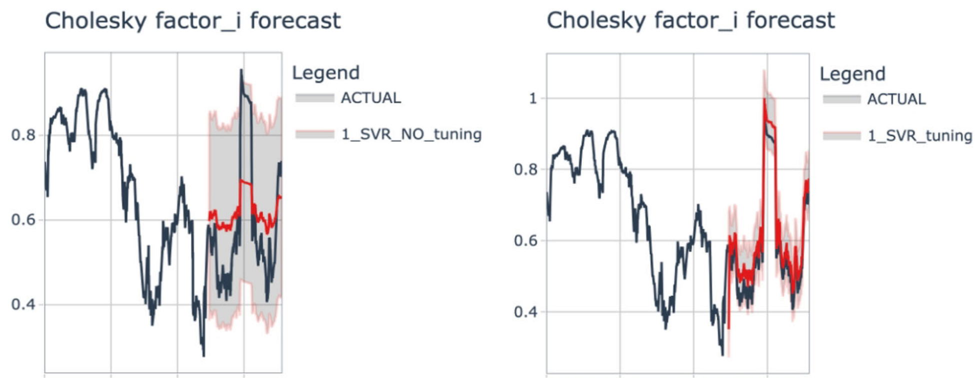
Since we had 435 series, we take a sample of 29 series, and tune the hyperparameters for one series each, and then execute the found parameters for the whole set of series. Then we evaluate if performance metric improved globally compared to the default parameter.

For the cases of RF and SVR we did not observe a significant gain. For the case of ANN we saw differences between defaults values used previously and the new hyperparameters tuned that were *hidden_unit* = 5 and *num_networks* = 22, that generate a significant gain in MAE = 0.265 for default NNAR(1, 1, 1) versus MAE = 0.183 for NNAR(1, 1, 5).

Here we show as example how the tuning for the SVR before and after the tuning of a particular series of Cholesky factor work.

Table 3 Evaluation of models with default parameters versus parameters selected by hypertuning

Model	Setting	Cost	rbf_sigma	Margin	mtry	Trees	Hidden_units	nnetworks	type_data	N_Series	MAE	RMSE
SVR	Default	1	0.2	0.1	–	–	–	–	Multivariate	29.00	0.15	0.20
SVR	Tuned	23	0.0132	0.179	–	–	–	–	Multivariate	29.00	0.17	0.20
RF	Default	–	–	–	10	2000	–	–	Multivariate	29.00	0.15	0.18
RF	Tuned	–	–	–	15	2581	–	–	Multivariate	29.00	0.16	0.20
ANN	Default	–	–	–	–	–	1	20	Multivariate	29.00	0.16	0.20
ANN	Tuned	–	–	–	–	–	5	22	Multivariate	29.00	0.14	0.17



(a) SVR default parameters (b) SVR tuned parameters

However if we use those parameters in the rest of series, there are no a improvement, no at least in SVR and RF model, yes in the case of ANN model, Table 3.

Note that in the context of univariate time series the metric mean absolute error and root-mean-square error formulas take the form:

$$\mathrm{MAE} = \frac{1}{n}\sum_{t=1}^{n} |y_t - \hat{y}_t| \tag{31}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n} (y_t - \hat{y}_t)^2} \tag{32}$$

where y_t: real value at time t, $\hat{y}_t$: predicted value at time t, and n: number of predictions.

Appendix A.2: Note about Sect. 6.3.3

In relation to Sect. 6.3.3, we additionally placed the regression in the opposite direction of the variables. With

Table 4 Linear regression summary for the relationship between forecast error (SFE) and market entropy for the different methods

	Dependent variable						
	log(sfe)						
	AANN (1)	AR1 (2)	ARMA (3)	DCC-GARCH (4)	RF (5)	RMT3 (6)	SVR (7)
log(Entropy)	1.283***	1.269***	1.266***	0.721***	1.288***	1.419***	1.278***
	(0.042)	(0.042)	(0.042)	(0.039)	(0.042)	(0.040)	(0.041)
Constant	2.011***	1.990***	1.989***	1.974***	1.998***	2.122***	2.000***
	(0.014)	(0.014)	(0.014)	(0.013)	(0.014)	(0.013)	(0.014)
Observations	2608	2608	2608	2608	2608	2608	2608
R^2	0.263	0.263	0.262	0.116	0.268	0.327	0.268
Adjusted R^2	0.263	0.262	0.261	0.116	0.268	0.327	0.268
Residual Std. Error ($df = 2606$)	0.715	0.709	0.709	0.663	0.709	0.678	0.704

$^*p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

this we show that regardless of the direction, DCC-GARCH remains as the least sensitive method (Table 4).

$$\log(\mathrm{SFE}) = \log(\text{Realized Entropy}) \tag{33}$$

Appendix A.3: Data availability

The data used in this research are freely available in the site https://finance.yahoo.com/. The data description is presented in detail in Table 5 (Table 6).

Table 5 Index selection from different regions of the World—contain data between 2008 and 2022

Country	Ticker	Region	Economy	Description
South Africa	EZA	Africa	Developing	iShares MSCI South Africa ETF
China	GXC	Asia	Developing	SPDR S &P China ETF
China	FXI	Asia	Developing	iShares China Large-Cap ETF
China	000001.SS	Asia	Developing	Shanghai Stock Exchange SSE 50 A Share Index
Hong Kong—China	EWH	Asia	Developed	iShares MSCI Hong Kong ETF
Hong Kong—China	^HSI	Asia	Developed	HANG SENG INDEX
Japan	EWJ	Asia	Developed	iShares MSCI Japan ETF
Japan	^N225	Asia	Developing	Nikkei 225
Singapore	EWS	Asia	Developing	iShares MSCI Singapore ETF
South Korea	EWY	Asia	Developed	iShares MSCI South Korea ETF
Taiwan	EWT	Asia	Developing	iShares MSCI Taiwan ETF
France	EWQ	Europe	Developed	iShares MSCI France ETF}
France	^FCHI	Europe	Developed	CAC 40
Germany	EWG	Europe	Developed	iShares MSCI Germany ETF
Germany	^GDAXI	Europe	Developed	DAX PERFORMANCE-INDEX
Italy	EWI	Europe	Developed	iShares MSCI Italy ETF
Switzerland	^SSMI	Europe	Developed	SMI PR
UK	EWU	Europe	Developed	iShares MSCI UK ETF
UK	^FTSE	Europe	Developed	FTSE 100
Euro Zone	^STOXX50E	Europe	Developed	ESTX 50 PR.EUR
LATAM	ILF	LATAM	Developing	iShares Latin America 40 ETF
Canada	EWC	North America	Developed	iShares MSCI Canada ETF
Mexico	EWW	North America	Developing	iShares MSCI Mexico ETF
USA	^DJI	North America	Developed	Dow Jones Industrial Average
USA	^GSPC	North America	Developed	S &P 500
USA	^NDX	North America	Developed	NASDAQ 100
Australia	EWA	Oceania	Developed	iShares MSCI Australia ETF
Brazil	EWZ	South America	Developing	iShares MSCI Brazil ETF
Chile	ECH	South America	Developing	iShares MSCI Chile ETF

Table 6 Descriptive statistics of daily returns of ETFs considered in the study

Country	Ticker	Mean (%)	Std. Dev. (%)	Maximum (%)	Minimum (%)	Skewness	Kurtosis
Australia	EWA	− 0.01	1.88	13.24	− 17.56	− 0.68	10.85
Brazil	EWZ	− 0.03	2.53	16.23	− 26.26	− 1.11	11.93
Canada	EWC	0.01	1.51	12.10	− 14.30	− 1.06	13.15
Chile	ECH	0.00	1.76	11.32	− 16.97	− 0.78	10.45
China	GXC	− 0.01	1.99	17.96	− 16.71	0.18	9.38
China	FXI	− 0.03	2.12	19.26	− 16.07	0.25	9.71
China	000001.SS	− 0.02	1.50	9.03	− 8.87	− 0.39	5.56
France	EWQ	0.01	1.73	10.70	− 13.56	− 0.77	8.38
France	^FCHI	0.01	1.45	9.22	− 13.10	− 0.36	7.90
Germany	EWG	− 0.00	1.71	10.22	− 13.57	− 0.88	8.06
Germany	^GDAXI	0.01	1.41	10.41	− 13.05	− 0.53	7.16
Hong Kong	EWH	− 0.01	1.53	10.90	− 13.12	− 0.36	8.64
Hong Kong	^HSI	− 0.02	1.45	12.06	− 8.66	0.19	6.50
Italy	EWI	− 0.01	1.99	11.08	− 17.01	− 0.88	8.02
Japan	EWJ	0.00	1.32	7.60	− 10.99	− 0.74	7.81
Japan	^N225	0.02	1.49	9.49	− 12.11	− 0.68	7.46
LATAM	ILF	− 0.02	2.19	12.76	− 21.65	− 1.08	11.52
Mexico	EWW	0.01	1.78	9.02	− 16.55	− 0.89	7.82
Singapore	EWS	− 0.02	1.53	10.37	− 11.76	− 0.56	8.48
South Africa	EZA	0.01	2.31	17.26	− 22.42	− 0.81	8.93
South Korea	EWY	0.01	2.01	19.94	− 17.20	− 0.19	13.62
Switzerland	^SSMI	0.01	1.12	6.78	− 10.13	− 0.68	8.85
Taiwan	EWT	0.01	1.61	13.24	− 11.51	− 0.35	7.02
UK	EWU	− 0.01	1.56	10.93	− 12.77	− 1.19	10.80
UK	^FTSE	− 0.01	1.20	8.67	− 11.51	− 0.53	10.06
USA	^DJI	0.03	1.26	10.76	− 13.84	− 1.00	14.91
USA	^GSPC	0.03	1.32	8.97	− 12.77	− 0.99	11.62
USA	^NDX	0.04	1.45	9.60	− 13.00	− 0.67	6.86
ZonaEuro	^STOXX50E	0.01	1.47	9.85	− 13.24	− 0.41	7.51

The sample covers the period between 2008 and 2022

Funding No funds were received in the development of this research.

Data availability Data will be made available on request.

Declarations

Conflict of interest The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Basak S, Kar S, Saha S, Khaidem L, Dey SR (2019) Predicting the direction of stock market prices using tree-based classifiers. N Am J Econ Finance 47:552–567
2. Becker R, Clements AE, Doolan MB, Hurn AS (2015) Selecting volatility forecasting models for portfolio allocation purposes. Int J Forecast 31(3):849–861
3. Bernstein PL (1996) Against the gods: the remarkable story of risk. John Wiley & Sons, New York
4. Bloznelis D (2018) Short-term salmon price forecasting. J Forecast 37(2):151–169
5. Bollerslev T (1986) Generalized autoregressive conditional heteroskedasticity. J Econom 31(3):307–327
6. Bonanno G, Vandewalle N, Mantegna RN (2000) Taxonomy of stock market indices. Phys Rev E 62(6):R7615–R7618. https://doi.org/10.1103/PhysRevE.62.R7615
7. Box GE, Jenkins GM, Reinsel GC, Ljung GM (2015) Time series analysis: forecasting and control. John Wiley & Sons, New York
8. Breiman L (2001) Random forests. Mach Learn 45:5–32
9. Caldeira JF, Moura GV, Perlin MS, Santos AA (2017) Portfolio management using realized covariances: evidence from Brazil. EconomiA 18(3):328–343

10. Cao L, Tay FE (2001) Financial forecasting using support vector machines. Neural Comput Appl 10(2):184–192
11. Civitarese J (2016) Volatility and correlation-based systemic risk measures in the US market. Phys A Stat Mech Appl 459:55–67
12. Coelho R, Gilmore CG, Lucey B, Richmond P, Hutzler S (2007) The evolution of interdependence in world equity markets-evidence from minimum spanning trees. Phys A Stat Mech Appl 376:455–466
13. Cont R (2001) Empirical properties of asset returns: stylized facts and statistical issues. Quant Finance 1(2):223
14. Danielsson J (2011) Financial risk forecasting: the theory and practice of forecasting market risk with implementation in R and MATLAB. John Wiley & Sons, New York
15. Drożdż S, Grümmer F, Górski AZ, Ruf F, Speth J (2000) Dynamics of competition between collectivity and noise in the stock market. Phys A Stat Mech Appl 287(3):440–449. https://doi.org/10.1016/S0378-4371(00)00383-6
16. Drożdż S, Grümmer F, Ruf F, Speth J (2001) Towards identifying the world stock market cross-correlations: DAX versus Dow Jones. Phys A Stat Mech Appl 294(1):226–234. https://doi.org/10.1016/S0378-4371(01)00119-4
17. Engle RF (1982) Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. Econ J Econom Soc 987–1007
18. Engle R (2002) Dynamic conditional correlation: a simple class of multivariate generalized autoregressive conditional heteroskedasticity models. J Bus Econ Stat 20(3):339–350
19. Engle RF III, Sheppard K (2001) Theoretical and empirical properties of dynamic conditional correlation multivariate GARCH. National Bureau of Economic Research, Cambridge
20. Fatima S, Uddin M (2022) On the forecasting of multivariate financial time series using hybridization of DCC-GARCH model and multivariate ANNs. Neural Comput Appl 34(24):21911–21925
21. Fenn DJ, Porter MA, Williams S, McDonald M, Johnson NF, Jones NS (2011) Temporal evolution of financial-market correlations. Phys Rev E 84(2):026109
22. Fiszeder P, Orzeszko W (2021) Covariance matrix forecasting using support vector regression. Appl Intell 51(10):7029–7042
23. Frahm G, Jaekel U (2005) Random matrix theory and robust covariance matrix estimation for financial data. arXiv:physics/0503007. Retrieved 12 Jan 2019
24. Galanos A (2022) rmgarch: Multivariate garch models, [Computer software manual] R package version 1.3-9
25. Gopikrishnan P, Rosenow B, Plerou V, Stanley HE (2000) Identifying business sectors from stock price fluctuations. arXiv:cond-mat/0011145. Retrieved 19 March 2019
26. Gopikrishnan P, Rosenow B, Plerou V, Stanley HE (2001) Quantifying and interpreting collective behavior in financial markets. Phys Rev E. https://doi.org/10.1103/PhysRevE.64.035106
27. Hyndman RJ, Athanasopoulos G (2018) Forecasting: principles and practice. OTexts
28. Izrailev F. M (1990) Simple models of quantum chaos: spectrum and eigenfunctions. Phys Rep 196(5):299–392. https://doi.org/10.1016/0370-1573(90)90067-C
29. Khaidem L, Saha S, Dey SR (2016) Predicting the direction of stock market prices using random forest. arXiv preprint arXiv:1605.00003
30. Kumar S, Deo N (2012) Correlation and network analysis of global financial indices. Phys Rev E 86(2):026101. https://doi.org/10.1103/PhysRevE.86.026101
31. Kumbure MM, Lohrmann C, Luukka P, Porras J (2022) Machine learning techniques and data for stock market forecasting: a literature review. Expert Syst Appl 197:116659
32. Laloux L, Cizeau P, Bouchaud J-P, Potters M (1999) Noise dressing of financial correlation matrices. Phys Rev Lett 83(7):1467
33. Laloux L, Cizeau P, Potters M, Bouchaud J-P (2000) Random matrix theory and financial correlations. Int J Theor Appl Finance 3(03):391–397
34. Ledoit O, Wolf M (2004) Honey, I shrunk the sample covariance matrix. J Portfolio Manag 30(4):110–119
35. Livan G, Novaes M, Vivo P (2018) Introduction to random matrices: theory and practice. Springer, Berlin
36. Mantegna RN (1999) Hierarchical structure in financial markets. Eur Phys J B Condens Matter Complex Syst 11(1):193–197. https://doi.org/10.1007/s100510050929
37. Mantegna RN, Stanley HE (1999) Introduction to econophysics: correlations and complexity in finance. Cambridge University Press, Cambridge
38. Martins ACR (2007) Non-stationary correlation matrices and noise. Phys A Stat Mech Appl 379(2):552–558. https://doi.org/10.1016/j.physa.2006.12.020
39. Nayak SC, Misra BB, Behera HS (2019) Efficient financial time series prediction with evolutionary virtual data position exploration. Neural Comput Appl 31:1053–1074
40. Plerou V, Gopikrishnan P, Rosenow B, Amaral LAN, Stanley HE (1999) Universal and nonuniversal properties of cross correlations in financial time series. Phys Rev Lett. 83(7):1471
41. Plerou V, Gopikrishnan P, Rosenow B, Amaral LN, Guhr T, Stanley HE (2002) Random matrix approach to cross correlations in financial data. Phys Rev E 65:6066126. https://doi.org/10.1103/PhysRevE.65.066126
42. Posit team (2023) RStudio: Integrated Development Environment for R [Computer software manual]. Boston. http://www.posit.co/
43. Rajihy Y, Nermend K, Alsakaa A (2017) Back-propagation artificial neural networks in stock market forecasting. An application to the Warsaw stock exchange WIG20. Aestimatio IEB Int J Finance 15:88–99
44. Ren F, Zhou W-X (2014) Dynamic evolution of cross-correlations in the Chinese stock market. PLoS ONE 9(5):e97711
45. Rodríguez-Moreno M, Peña JI (2013) Systemic risk measures: the simpler the better? J Bank Finance 37(61):817–1831
46. Rosenow B (2008) Determining the optimal dimensionality of multivariate volatility models with tools from random matrix theory. J Econ Dyn Control 32(1):279–302
47. Rosenow B, Plerou V, Gopikrishnan P, Stanley HE (2002) Portfolio optimization and the random magnet problem. EPL (Europhys Lett) 59(4):500
48. Soni P, Tewari Y, Krishnan D (2022) Machine Learning approaches in stock price prediction: a systematic review. J Phys Conf Ser 2161:012065
49. Thenmozhi M, Sarath Chand G (2016) Forecasting stock returns based on information transmission across global markets using support vector machines. Neural Comput Appl 27:805–824
50. Torres R, Villena M (2023) On the predictive power of RMT correlation matrices. Manuscript
51. Vapnik V (1999) The nature of statistical learning theory. Springer, Berlin
52. Wang G-J, Xie C, Stanley HE (2018) Correlation structure and evolution of world stock markets: evidence from Pearson and partial correlation-based networks. Comput Econ 51:607–635
53. Wang J, Zhuang Z, Feng L (2022) Intelligent optimization based multi-factor deep learning stock selection model and quantitative trading strategy. Mathematics 10(4):566

54. Yang PR (2020) Using the yield curve to forecast economic growth. J Forecast 39(7):1057–1080
55. Zakamulin V (2015) A test of covariance-matrix forecasting methods. J Portfolio Manag 41(3):97–108
56. Zheng Z, Podobnik B, Feng L, Li B (2012) Changes in cross-correlations as an indicator for systemic risk. Sci Rep 2:888