REVISTA INNOVAR JOURNAL

A New Forecasting Combination System for Predicting Volatility

Johanna M. Orozco
Departamento de Ciencias de la Computación y la Decisión,
Universidad Nacional de Colombia, Sede Medellín.
Correo electrónico: jmorozcoc@unal.edu.co

Juan D. Velásquez
Departamento de Ciencias de la Computación y la Decisión,
Universidad Nacional de Colombia, Sede Medellín.
Correo electrónico: jdvelasq@unal.edu.co
Facultad de Minas, Universidad Nacional de Colombia, A.A. 1027, Medellín, Colombia

UN NUEVO SISTEMA DE COMBINACIÓN DE PRONÓSTICOS PARA LA PREDICCIÓN DE LA VOLATILIDAD

RESUMEN: Los modelos para la combinación de pronósticos han sido ampliamente estudiados, y de uso frecuente en el mejoramiento de la exactitud de predicciones. En este artículo se presenta un nuevo modelo compuesto no lineal, para la predicción de la volatilidad de activos. Dicho modelo esta compuesto de una serie de modelos GARCH, anclados a un conjunto de datos de series de tiempo, que emplean diferentes funciones de pérdida, los cuales tienen el objetivo de capturar diferentes características de la dinámica propia de la volatilidad. Se combinan predicciones individuales, mediante el uso de la media aritmética simple, o de una red neuronal artificial. Este modelo propuesto se emplea para predecir la rentabilidad mensual de series de tiempo S&P500, llevando a concluir que el nuevo enfoque permite predecir la volatilidad con mayor exactitud que cada uno de los modelos GARCH considerados.

PALABRAS CLAVE: Volatilidad, modelos de predicción de la volatilidad, combinaciones de pronósticos

UN NOUVEAU SYSTÈME DE COMBINAISON DE PRONOSTICS POUR LA PRÉDICTION DE LA VOLATILITÉ

RÉSUMÉ : Les modèles pour la combinaison de pronostics ont été largement étudiés et ont été d'un usage fréquent pour l'amélioration de l'exactitude des prédictions. Dans cet article est présenté un nouveau modèle composé non linéaire pour la prédiction de la volatilité des actifs. Ce modèle est composé d'une série de modèles GARCH, ancrés à un ensemble de données de séries de temps, qui emploient différentes fonctions de perte, lesquelles ont pour objectif de de saisir plusieurs caractéristiques de la dynamique propre de la volatilité. Se combinent des prédictions individuelles, par l'usage soit de la moyenne arithmétique simple, soit d'un réseau neuronal artificiel. Ce modèle proposé est utilisé pour prédire la rentabilité mensuelle de séries de temps S&P500, ce qui mène à conclure que la nouvelle approche permet de prédire la volatilité avec plus d'exactitude que chacun des modèles GARCH considérés.

MOTS-CLÉS : Volatilité, modèles de prédiction de la volatilité, combinaisons de pronostics.

UM NOVO SISTEMA DE COMBINAÇÃO DE PROGNÓSTICOS PARA A PREDIÇÃO DA VOLATILIDADE

RESUMO: Os modelos para a combinação de prognósticos foram amplamente estudados, e de uso frequente no melhoramento da exatidão das predições. Neste artigo apresenta-se um novo modelo composto não lineal, para a predição da volatilidade de ativos. Este modelo está composto por uma série de modelos GARCH, ancorados a um conjunto de dados de séries de tempo, que empregam diferentes funções de perda, os quais tem o objetivo de capturar diferentes características da dinâmica própria da volatilidade. Combinam-se predições individuais, mediante o uso já seja da média aritmética simples, ou de uma rede neuronal artificial. Este modelo proposto emprega-se para predizer a volatilidade com maior exatidão que cada um dos modelos GARCH considerados.

PALAVRAS CHAVE: Volatilidade, modelos de predição da volatilidade, combinações de prognósticos.

CLASIFICACIÓN JEL: C530, C580,G170

RECIBIDO: febrero de 2012 **APROBADO:** febrero de 2013

CORRESPONDENCIA: Carrera 80 No. 65-223, Bloque M8A, Of. 206, Departamento de Ciencias de la Computación y la Decisión, Facultad de Minas, Universidad Nacional de Colombia - Sede Medellín, Medellín, Colombia.

CITACIÓN: Orozco, J. M. & Velásquez J. D. (2013). A new forecasting combination system for predicting volatility. *Innovar, 23*(50), 5-16.

ABSTRACT: Forecast combination models have been broadly studied and often used to improve forecast accuracy. This article presents a new non-linear composite model to forecast the volatility of asset returns. Our model is composed of a set of GARCH models fitted to a time series dataset using different loss functions, with the aim of capturing different features of volatility dynamics. Individual forecasts are combined by using either the simple arithmetical average method or an artificial neural network. The proposed model is used to forecast the monthly excess returns of S&P500 time series, finding that this new approach is able to forecast volatility with more accuracy than each individual GARCH model considered.

KEYWORDS: Volatility, Forecast volatility models, forecast combinations.

Introduction

Volatility forecasting is an important and difficult task in financial engineering. Forecasts are used in option pricing, risk management, portfolio analysis and hedging strategies, among others (Poon & Granger, 2003; Tang *et al*., 2009; Hull & White, 1987). There are mainly two reasons to explain the difficulty for obtaining accurate forecasts: first, the volatility is an unobserved latent variable that evolves stochastically through time (Amendola & Storti, 2008; Patton, 2011; Bollerslev & Engle, 1993) and the asset squared returns usually are assumed as an unbiased estimator of volatility; however, the squares returns will not always lead to the same estimated values for the volatility as if the latent variable were used. Second, volatility must be approximated from return time series that exhibit complex features as time-varying variance, clusters of similar variance, persistence in the sense that current information remains important for forecasts of all horizons, heavy tailed marginal distributions, asymmetric response to the sign of past returns or leverage effect, and nonlinear and non-stationary behavior (Poon & Granger, 2003; Donaldson & Kamstra, 1997; Bollerslev & Engle, 1993).

There are two groups of approaches used for obtaining volatility forecast when only one model is considered: The first group is related to the evaluation of the forecasting power of many well-established volatility models (Poon & Granger, 2003; Taylor, 2004; Verhoeven *et al*., 2002; Hu & Tsoukalas, 1999; Amendola & Storti, 2008) and to the use of alternative

pragmatic methodologies.As stated in previous paragraphs, the second group is related to the evaluation of the forecasting power of many well-established volatility models (Poon & Granger, 2003; Taylor, 2004; Verhoeven *et al*., 2002; Hu & Tsoukalas, 1999; Amendola & Storti, 2008) and to the use of alternative pragmatic methodologies. For example, Taylor (2004) develops a smooth transition exponential model in which the parameters are time-varying; Brooks and Persand (2003) evaluated volatility forecasting in a financial risk management setting in terms of value-at-risk (VaR); Lopez (2001) considered probability scoring rules that were tailored to a forecast user's decision problem and confirms that the choice of loss function directly affects the forecast evaluation results.

The second group is devoted to the use of traditional statistical rigorous econometric approaches with the aim of understand the in-sample properties and explain the historical behavior of the volatility. Most of the approaches are based on the seminal work of Engle (1982); for example, Bollerslev (1986) postulated the GARCH model as a more parsimonious model than the ARCH model of Engle (1982); other approaches as the GJR-GARCH, the exponential GARCH (EGARCH) of Nelson (1991), the A-GARCH, NA-GARCH and V-GARCH models of Engle et al. (1993), among others, were developed to account for more complex issues, with the aim of representing the asymmetric impact of positive and negative information on the market, distribution of the standardized innovation and persistence of the volatility process. The topic of volatility modeling is reviewed by Andersen et al. (2005a). An extension of previous idea is to suppose the existence of several regimens in the time series and to use a different type of model –or the same type of model with different estimated parameters– for each regimen; the SETAR or STAR models are used when the regimens are determined by observable variables; and Markov-switching models when the regimens are determined by unobserved variables. Examples of this approach are the works of Chan and McAleer (2003) which considered a smooth transition ST-GARCH model for the errors; Lundbergh and Teräsvirta (1998),applied a STAR-STGARCH to characterize nonlinear behavior in the conditional mean and conditional variance; Klaassen (2002) used a regime-switching GARCH model for distinguishing two regimes with different volatility levels; Nazifi and Fatahi (2012) study Markov- switching GARCH models to estimate and forecast the volatility of Tehran stock market. An extend discussion about this topic is provided by Franses and Dijk (2006).

However, the forecasts of several independent models would be considered simultaneously for obtaining a single volatility forecast. The idea of combining several alternative forecasts is justified by many empirical experiences, indicating that the accuracy of the combined forecast is often higher than each alternative forecast (Clemen & Winkler, 1986; Clemen, 1989; Hashem, 1997). In statistics, these methodologies are known as forecast combination techniques (Granger, 1989), while in computational intelligence are called ensemble methods (Haykin, 1996). The conceptual key idea behind all of these approaches is the diversification of the forecasting models; this diversification may be achieved mainly in three ways: first, by using different types of models; second, by using the same type of model with identical internal configuration but different parameter values; and third, by using the same type of model with different internal configuration.

There are several successful experiences about the combination of individual volatility forecasts. Amendola and Storti (2008) propose the use of the generalized method of moments for estimating the weights of a linear combiner. Tsangari (2007) presents a new non-linear, non-parametric, kernel-based method for combining individual forecasts effectively, where the functional form is not assumed to be known. Hu and Tsoukalas (1999) propose an ANN model to combine four model forecasts, in order to provide improved volatility forecasts.

The objective of this paper is to present a new volatility forecast combination methodology based on the following points:

1. The GARCH must be rewritten in the form of the ARMA approach, such that, model parameters would be estimated by minimizing a function of the difference between the squared shocks and the estimated volatility (Tsay, 2010).

2. The behavior, advantages, and limitations of different error functions have extensively studied in time series literature (Hyndman *et al*., 2006). It is well-known that the use of different functions for estimating the parameters in time series and regression models, allows the modeler to capture different behaviors in a time series. For example, the mean squared error is very sensitive to the presence of outliers, unlike the mean absolute deviation. In our approach, we fit several GARCH models by minimizing different error functions with the aim of capture different aspects of the volatility behavior.

3. When several forecasts for the same event are available, it is possible to obtain a composite forecast using forecast combination techniques. In this work, we consider possible alternatives to obtain a composite forecast: the simple arithmetical average method, the multiple linear regression and an artificial neural network.

The structure of this paper is organized as follows. In Section 2, we present the GARCH model, the parameterization and methodologies for estimating model parameters, and the proposed forecast combination model. In section 3, we give a numerical experiment using a benchmark time series for exemplifying and testing the performance of the forecast volatility models. The final section gives some concluding remarks.

Proposed methodology

The GARCH model

One of the most used models for statistical modeling and forecasting the conditional volatility is the generalized autoregressive conditional heteroskedastic (GARCH) approach of Bollerslev (1986). In a GARCH model, the current variance, σ_t^2, is a function of the past squared shocks, $\{a_{t-i}^2; i = 1, \dots, p\}$, and the past variances, $\{\sigma_{t-j}^2; j = 1, \dots, q\}$:

$$\sigma_t^2 = \kappa + \sum_{i=1}^{p} \alpha_i . a_{t-i}^2 + \sum_{j=1}^{q} \beta_j . \sigma_{t-j}^2 \quad \text{with } a_t = \sigma_t \varepsilon_t \quad \text{and } \varepsilon_t \sim N(0,1) \qquad (1)$$

where the parameters κ, α_i and β_j, are subject to the following restrictions: $\kappa > 0$, $\alpha_i \geq 0$, $\beta_j \geq 0$ and $\sum_{i=1}^{\max(p,q)} (\alpha_i + \beta_j) < 1$.

The classical and well-known procedure for the estimation of the parameters is derived supposing that the normalized residuals ε_t follow a standard normal distribution; see Tsay (2010). Thus, the optimal values of the parameters are calculated by maximizing the natural logarithm of the residuals likelihood function:

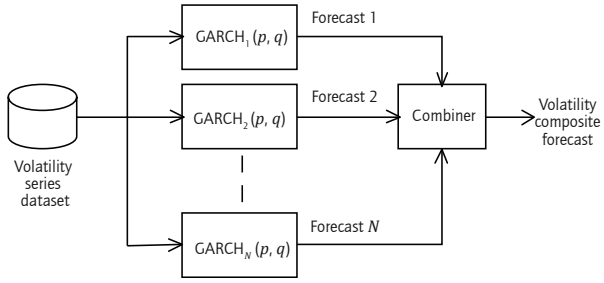
$$\log L = -\frac{N}{2} \log 2\pi - \frac{1}{2} \sum_{t=1}^{N} \left[\log \sigma_t^2 + \frac{a_t^2}{\sigma_t^2} \right] \qquad (2)$$

where N is the time series length.

General model structure

The general structure of the proposed system is presented in the Figure 1. It is composed of N GARCH(p, q) models with the same structure but with different parameter values; all GARCH(p, q) models are fitted using the same sample of data, but using different optimizations with the aim of capturing different features of the volatility

FIGURE 1. Structure of the volatility forecasting system.



Source: Authors' elaboration.

dynamics. Each model is used to produce an individual forecast using the available information; finally, all individual forecasts are combined to obtain a volatility composite forecast.

Alternative parameter estimation procedure for the GARCH model

Tsay (2010) presents an alternative estimation procedure based on the minimization of a function of the difference between the volatility and its forecasting. By defining $u_t = a_t^2 - \sigma_t^2$, the GARCH model presented in Eq. (1) would be written out as:

$$a_t^2 = \kappa + \left[\sum_{i=1}^{\max(p,q)} (\alpha_i + \beta_i) a_{t-i}^2\right] - \left[\sum_{j=1}^{q} \beta_j u_{t-j}\right] + u_t \qquad (3)$$

Thus, the previous equation has the same structure of an ARMA model where the first expression inside parenthesis corresponds to the autoregressive component while the second correspond to the moving average component. As a consequence, the parameters in the model presented in Eq. (3) would be estimated by the following minimization:

$$\min \sum_t u_t^2 \quad \text{with} \quad u_t = a_t^2 - \sigma_t^2 \qquad (4)$$

Note that maximize the Eq. (2) is not the same thing that the expression in Eq. (4) and the optimal parameters in each case are different. As indicated by Tsay (2010), the theoretical implications and the practical merits of this approach are unknown and it is necessary further research.

Estimation of the Parameters of the GARCH model using different error functions

The use of Eq. (2) for estimating the optimal values of the parameters in a GARCH(p, q) is based on the assumption of normality of the standardized residuals ε_t; in this sense, when the log-likelihood function is higher, then the probability distribution of ε_t is nearer to a theoretical standard normal distribution; however, it does not mean that the parameters are optimal in terms of the volatility forecast accuracy.

The estimation procedure proposed by Tsay (2010) is focused in the forecast accuracy by minimizing a function of the forecast error; this is, the parameters of the model defined in Eq. (3) are estimated executing the minimization presented in Eq. (4). However, the parameters of each GARCH(p, q) model would be estimated by minimizing any general loss function.

Below, we present a list of commonly employed loss functions on volatility forecast evaluation (Gooijer & Hyndman, 2006; Patton, 2011; Dunis & Huang, 2002). These functions are not robust to noise (Patton, 2011), such that, the optimal forecast is not the true conditional variance. However, we use these functions for fitting the GARCH(p, q) models of our system because of they are able to capture other statistical features of time series that are ignored when the GARCH model is estimated by maximizing the likelihood function; as a consequence, we expect to improve our forecast accuracy respect the classical approach. The functions are a measure that gives us a basis to compare our volatility forecasts with the realized volatility. Under the assumptions for the conditional distribution of daily returns given in Patton (2011), the MSE loss function generates an optimal forecast equal to the conditional variance, and thus satisfies the necessary condition for robustness. In addition, we include the linex loss function [proposed by Varian (1974) and used by Zellner (1986) and Christoffersen and Diebold (1994)] for modeling the possible existence of asymmetries in the response of volatility to the sign of shocks; in practical terms, this means that the model is able to model in a different way large positive and negative differences between squared shocks and variances; in eq. (8), the parameter c controls the asymmetry level, and $b_i = \frac{1}{N}$ in most simple case.

- Mean absolute deviation:

$$\text{MAD} = \frac{1}{N}\sum_{t=1}^{N} |a_t^2 - \sigma_t^2| = \frac{1}{N}\sum_{i=n}^{N} |u_t| \qquad (5)$$

- Mean squared error:

$$\text{MSE} = \frac{1}{N}\sum_{t=1}^{N} |a_t^2 - \sigma_t^2|^2 = \frac{1}{N}\sum_{i=n}^{N} |u_t|^2 \qquad (6)$$

- Mean cubic error:

$$\text{MCE} = \frac{1}{N}\sum_{t=1}^{N} |a_t^2 - \sigma_t^2|^3 = \frac{1}{N}\sum_{i=n}^{N} |u_t|^3 \qquad (7)$$

- Linex loss function:

$$\text{LLF} == \sum_{t=1}^{N} b_i[\text{Exp}\{c(a_t^2 - s_t^2)\} - c(a_t^2 - s_t^2) - 1], c \neq 0 \quad (8)$$

Forecasts combiner

By analyzing the most relevant literature about ensemble averaging methods and forecasts combination techniques, we select the following techniques for combining the volatility forecasts, with f_i, i = 1, ...n, of the GARCH(p, q) models:

- Simple arithmetical average: $f_* = \frac{\sum_{i=1}^{n} f_i}{n}$
- Combination by multiple linear regression: $f_* = \beta_0 + \beta_1 f_1 + \cdots + \beta_n f_n$, where β_i are the parameters.
- Nonlinear optimal combination by using an artificial neural network. In this research, we use the model proposed by Hu and Tsoukalas (1999), where the forecast is obtained as:

$$f_* = \varphi + \left(\sum_{i=1}^{n} \eta_i f_i\right) + \sum_{h=1}^{H} \omega_h . g\left(\beta_h + \sum_{i=1}^{n} \alpha_{hi} . f_i\right) \quad (9)$$

where φ, β_h, η_h, ω_h, and α_{hi} are the parameters. $g(\cdot)$ is the sigmoid transfer function:

$$g(u) = \frac{1}{1 + \exp(-u)} \quad (10)$$

Note that, the ANN model described by eq. (9) is the combination of a linear regression model and a multilayer perceptron neural network; this allows to the neural network to capture in a better way the linear and nonlinear dynamics in the data. Finally, the model in eq. (9) has $nH + 2H + n + 1$ parameters to be estimated.

Numerical experiment

Data and preliminary analysis

For our experiment, we use the monthly excess stock returns of the S&P500 index from January of 1926 to December of 1991; see Figure 1. The dataset contains 792 observations and it is commonly used as a benchmark for exemplifying and testing volatility models; see Tsay (2010) and Verhoeven *et al.* (2002).

Tsay (2010) analyses this time series using all available information (the 792 observations). First, an AR(3)-GARCH(1,1) is postulated for representing the dynamics of the data; the parameters of this model are estimated by maximizing Eq. ((2). Following, all non-significant parameters are dropped and the model is reduced to a GARCH(1,1). The final model proposed by Tsay (2010, page. 136) would be written out as:

$$r_t = \mu + a_t, \qquad a_t = \varepsilon_t \sigma_t, \qquad \sigma_t^2 = \kappa + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \quad (11)$$

with μ = 0.0076, κ = 0.000086, α_1 = 0.1216 and β_1 = 0.8511.

Experimental setup

In our numerical experiment, we use the first 708 observations for estimating the parameters of all GARCH(1,1) models and the remaining 84 observations for forecasting. Forecast accuracy, for the fitting and forecasting samples, is measured using the mean absolute deviate (MAD) and the mean squared error (MSE) previously defined in Eq. (5) and Eq. (6) respectively.

The setup of our experiment is described by the following steps:

- The GARCH model in Eq. (11) is fitted using the first 708 observations by maximizing the natural logarithm of likelihood function in Eq. (2).
- We rewrite the GARCH(1, 1) model in Eq. (11) as:

$$a_t^2 = \kappa + (\alpha_1 + \beta_1) a_{t-1}^2 - \beta_1 u_{t-1} + u_t \quad (12)$$

  And then, we estimate a set of optimal parameters {κ, (α_1+β_1), β_1} when each one of the loss functions in Eq. (5), Eq. (6), Eq. (7) and Eq. (8) are minimized. Thus, we obtain four sets of parameters for the model presented in Eq. (12).

- For each GARCH model considered in this experiment, we forecast the volatility one month ahead for both the fitting and the forecasting samples.
- Combined forecasts are calculated using the individual forecasts obtained from the models estimated in steps 1 and 2 as follows:
  - ◊ The simple arithmetical average of the forecasts of each GARCH model.
  - ◊ By using a multiple linear regression.
  - ◊ By using ANN models with H = 1, ..., 4 and n = 5. See Eq. (9). f_1, f_2, f_3, f_4 and f_5 are the forecasts of the models estimated in the steps 1 and 2 by maximizing Eq. (2) and minimizing the MAD, MSE, MCE and LLF(c = 0.038) error functions respectively. Each ANN model is trained fifty times with random initial weights. We prefer the ANN model with the set of parameters minimizing the MSE for the fitting sample.

- The fitting and forecasting MSE and MAD are calculated. The obtained results, multiplying the MSE by 10^4 and the MAD by 10^2, are presented in Tables 2, 3, 4 and 5.

In Table 1, the set of parameters for the estimate models maximizing Eq. (2) and minimizing Eq. (5), Eq. (6), Eq. (7) and Eq. (8) are presented. It is interesting to note that the parameters of models in Table 1, do not satisfy the statistical constraints for guaranty positive values of σ_t^2, but, this is not a problem because of the models are used for forecasting in normal conditions of the market and it is very improbable to obtain negative values for σ_t^2. In addition, very large values for a_t^2 occur in abnormal conditions as crashes and bubbles, and it is impossible to forecast such events with any model. Finally, statistical restrictions are not absolute and necessary for obtaining good forecasting models.

In all cases, we use the MSE as the evaluation criteria for testing the volatility forecast accuracy (Tsangari, 2007; Lopez, 2001), where small values of MSE indicate good forecasting accuracy. Additionally, as it is shown in Patton (2011), the MSE error function is unbiased and it satisfies the necessary condition for a loss function to be robust to noise in the volatility proxy.

For this work, we implement a prototype in Microsoft Excel and R program. The optimal parameters of the experts and artificial neural networks were obtained using the solver complement and evolutionary algorithms.

Forecast accuracy of the individual models

We compare and measure the performance of the models using two criteria: the mean squared error (MSE), and the mean absolute deviation (MAD). In Table 2, we present the MSE (MAD) values for the fitting and forecasting samples when the five individual models are considered. The M_2 model (GARCH fitted by minimizing the MSE) results, present the lower value for MSE in both samples among all five models. But the MAD values for fitting and forecasting samples are not as good. The standard M_1 model (GARCH fitted by maximizing the log L) is overcome by the M_2 model in MSE values, and by the M_3 model (GARCH fitted by minimizing the MAD) in MAD values; this can happen because of the criterion of optimization of these models. The M_5 model (GARCH fitted by minimizing the LLF) results show good MSE values in comparison with the M_3 model but not in MAD values. In this way we look for a model which conforms to our expectation of reaching a balance for the MSE (MAD) values across both fitting and forecasting samples.

Forecast accuracy of the combined forecasts obtained using a simple average

In Table 3, we present the values of the MSE and MAD statistics calculated for the fitting and forecasting samples. In this case, we consider the average of the all combinations obtained by considering two, three, four and five models. By inspecting the Table 3, we found that the M_{26} model (the average of the forecasts from M_2 and M_3 models)

TABLE 1. Parameters of individual GARCH(1, 1) models.

Criteria	Obtained model
Fitted maximizing the log L	$\sigma_t^2 = 1.3023 + 0.1400a_{t-1}^2 + 0.8180\sigma_{t-1}^2$
Fitted minimizing the MSE	$a_t^2 = 1.3800 + (-0.8880 + 1.8494)a_{t-1}^2 - 1.8494u_{t-1} + u_t$
Fitted minimizing the MAD	$a_t^2 = 0.4318 + (-0.9528 + 1.8463)a_{t-1}^2 - 1.8463u_{t-1} + u_t$
Fitted minimizing the MCE	$a_t^2 = 3.9455 + (-0.8586 + 1.8953)a_{t-1}^2 - 1.8953u_{t-1} + u_t$
Fitted minimizing the LLF	$\sigma_t^2 = 1.0573 + (-0.8721 + 1.8534)a_{t-1}^2 - 1.8534u_{t-1} + u_t$

Source: Own elaboration.

TABLE 2. Evaluation of the volatility forecasting of individual models

Model	Fitting MSE (MAD)	Forecasting MSE (MAD)
GARCH (1, 1) models		
M_1: Fitted maximizing the log L	12294.0 (38.1)	3718.6 (28.7)
M_2: Fitted minimizing the MSE	* 12262.3 (38.6)	* 3678.7 (29.2)
M_3: Fitted minimizing the MAD	13426.9 (31.3)	3768.9 (23.0)
M_4: Fitted minimizing the MCE	15768.4 (73.2)	5905.4 (60.6)
M_5: Fitted minimizing the LLF	12326.1 (40.1)	3740.0 (30.1)

*Lower values of MSE in the column.

Source: Own elaboration.

TABLE 3. Evaluation of the volatility forecasting combination by simple arithmetical average

Model	Fitting MSE (MAD)	Forecasting MSE (MAD)
M_6: Average of M_1, M_2, M_3, M_4, M_5	12314.0 (41.8)	3738.6 (32.1)
M_7: Average of M_1, M_2, M_3, M_4	12321.5 (42.3)	3744.1 (32.6)
M_8: Average of M_1, M_2, M_3, M_5	12304.5 (36.1)	3657.9 (27.2)
M_9: Average of M_1, M_2, M_4, M_5	12540.6 (46.0)	3902.4 (35.8)
M_{10}: Average of M_1, M_3, M_4, M_5	12341.2 (42.7)	3762.2 (32.8)
M_{11}: Average of M_2, M_3, M_4, M_5	12343.7 (42.9)	3759.0 (33.0)
M_{12}: Average of M_1, M_2, M_3	12381.9 (35.0)	3647.3 (26.2)
M_{13}: Average of M_1, M_2, M_4	12662.3 (48.2)	3997.3 (37.9)
M_{14}: Average of M_1, M_2, M_5	12278.2 (38.9)	3710.0 (29.3)
M_{15}: Average of M_1, M_3, M_4	12364.3 (43.7)	3783.2 (33.9)
M_{16}: Average of M_1, M_3, M_5	12336.8 (35.4)	3659.5 (26.5)
M_{17}: Average of M_1, M_4, M_5	12753.7 (48.7)	4030.8 (38.3)
M_{18}: Average of M_2, M_3, M_4	12373.7 (43.9)	3781.4 (34.1)
M_{19}: Average of M_2, M_3, M_5	12342.1 (35.5)	*3645.4 (26.7)
M_{20}: Average of M_2, M_3, M_5	12744.6 (49.0)	4030.8 (38.5)
M_{21}: Average of M_3, M_4, M_5	12373.7 (43.9)	3781.4 (34.1)
M_{22}: Average of M_1, M_2	12271.3 (38.3)	3696.7 (28.9)
M_{23}: Average of M_1, M_3	12530.3 (33.6)	3662.1 (24.9)
M_{24}: Average of M_1, M_4	13154.9 (53.7)	4307.1 (42.9)
M_{25}: Average of M_1, M_5	12297.1 (39.1)	3727.0 (29.4)
M_{26}: Average of M_2, M_3	12553.1 (33.6)	*3639.3 (25.2)
M_{27}: Average of M_2, M_4	13145.3 (54.2)	4315.9 (43.3)
M_{28}: Average of M_2, M_5	12278.3 (39.3)	3708.1 (29.6)
M_{29}: Average of M_3, M_4	12506.5 (47.0)	3891.4 (36.9)
M_{30}: Average of M_3, M_5	12441.9 (34.2)	3651.8 (25.6)
M_{31}: Average of M_4, M_5	13341.1 (55.0)	4376.3 (43.8)

*Lower values of MSE in the column.

Source: Own elaboration.

results present the lower MSE for the forecasting sample in comparison with the competing models and individual models, but in fitting sample it does not hold for the MSE value; the M_{26} model has the lower MAD values. Also, we observe that M_{19} model (the average of the forecasts from M_2, M_3 and M_5 models) has a MSE value near to the MSE of M_{26} model for the forecasting sample and it has low MSE value for the training sample.

A common practice in forecast combining literature is considering the average of the forecast values of the all available models; in our case, this corresponds to consider the M_6 model in Table 3. Reported results show that this model only is better than the M_3, M_4 and M_5 models.

Forecast accuracy of the combined forecasts obtained using a multiple linear regression

In this case we consider the most of the possible linear combinations of individual forecasts. The estimation of the weights was doing by minimizing MSE and the obtained results are in Table 4. The better linear regression model, with respect to MSE value in the forecast sample, is the model that combines forecasts from models M_1 and M_3, the obtained MSE value is 3676.8, which is lower than MSE value in the combination by ANN with three nodes and the same forecasts; also we observe that simple arithmetical average gets a low MSE value of 3662.1 in forecasting sample with the same forecasts. We compare the accuracy between M_{48} model and individual forecasts from models M_1 and M_3, the linear multiple regression of these forecasts (model) is better because it has less MSE value than the individuals.

Accuracy of individual forecasts combined using an artificial neural network

In this case, we consider most of the possible combinations of the individual forecasts using the artificial neural network described by equations (9) and (10). For each set of possible inputs, ANN models with $H = 1, ..., 4$ were considered.

TABLE 4. Evaluation of the volatility forecasting combination by multiple linear regression

Model	Fitting MSE (MAD)	Forecasting MSE (MAD)
M_{32}: Linear combination of M_1, M_2, M_3, M_4, M_5	12233.9 (38.7)	3691.2 (29.7)
M_{33}: Linear combination of M_1, M_2, M_3, M_4	12233.3 (38.7)	3690.1 (29.7)
M_{34}: Linear combination of M_1, M_2, M_3, M_5	12238.1 (38.6)	3694.0 (29.7)
M_{35}: Linear combination of M_1, M_2, M_4, M_5	12233.9 (38.7)	3691.0 (29.7)
M_{36}: Linear combination of M_1, M_3, M_4, M_5	12235.8 (38.6)	3691.7 (29.7)
M_{37}: Linear combination of M_2, M_3, M_4, M_5	12232.4 (38.7)	3688.1 (29.8)
M_{38}: Linear combination of M_1, M_2, M_3	12258.8 (38.5)	3677.9 (29.2)
M_{18}: Linear combination of M_1, M_2, M_4	12235.9 (38.7)	3688.4 (29.$_{6)}$
M39: Linear combination $_o$f M_1, $M_{2,\,M5}$	12261.2 (38.7)	3678.0 (29.3)
M_{40}: Linear combination of M_1, M_3, M_4	12239.1 (38.6)	3686.8 (29.$_{5)}$
M41: Linear combination $_o$f M_1, *M3,* M_5	12241.1 (38.7*)*	$_3$692.7 (29.7)
M42: Linear $_c$om$b_i$na$t_i$on of M1, M4, M5	12241.2 (38.6)	3684.9 (29.$_{4)}$
M43: Linear combination $_o$f M_2, $M_{3,\,M4}$	12248.5 (38.9)	3686.2 (29.7)
M_{44}: Linear combination of M_2, M_3, M_5	12262.3 (38.6)	3678.6 (29.2)
M_{45}: Linear combination of M_2, M_4, M_5	12256.5 (38.6)	3681.5 (29.2)
M_{46}: Linear combination of M_3, M_4, M_5	12247.1 (38.6)	3681.5 (29.3)
M_{47}: Linear combination of M_1, M_2	12262.1 (38.6)	3677.0 (29.2)
M_{48}: Linear combination of M_1, M_3	12261.5 (38.6)	*3676.8 (29.2)
M_{49}: Linear combination of M_1, M_4	12270.4 (38.6)	3679.9 (29.2)
M_{50}: Linear combination of M_2, M_3	12262.3 (38.6)	3679.0 (29.2)
M_{51}: Linear combination of M_2, M_4	12261.1 (38.6)	3683.1 (29.2)
M_{52}: Linear combination of M_3, M_4	12262.0 (38.5)	3682.8 (29.2)

*Lower values of MSE in the column.

Source: Own elaboration.

In Table 5, we present the values of the MSE and MAD statistics for the fitting and forecasting samples. Several conclusions arise from this table:

- All the considered models have fitting MSE values lower than each individual model in Table 2. Thus, we conclude that each individual forecaster is able to capture valuable nonlinear information about the volatility dynamics.
- For almost all possible combinations of inputs, in terms of forecast MSE values the optimal configuration of the artificial neural network has two neurons in the hidden layer (H = 2).
- The nonlinear combination of individual models is better than each individual model; this is in concordance with the reported experiences in the literature where the composite forecast is more accurate than each individual forecast.
- The M_{86} model is the preferred among all possible combinations by ANN of volatility forecasts because it has a balance for the MSE (MAD) values across the fitting and forecasting samples in comparison with the individual models. In this case, the composite forecast is obtained by the non-linear combination of the forecasts calculated with the M_2, M_3 and M_4 models (which parameters are calculated by minimizing different loss functions). In Figure 1, we plot the squared shocks for the sample and the forecasts from the M_{86} model; in Figures 2, 3 and 4, we compare the forecasts of the M_{86} model with the forecasts of the M_4, M_3 and M_4 models; in these figures are evident that each model captures a different behavior of the squared shocks: the M_2 model, that is estimated by minimizing the MSE, (see Figure 2) presents a similar behavior with the M_{86} model, but the model M_{86} is able to approximate better the regions with high values of the squared shocks. The model M_3, estimated by minimizing the MAD, is able to capture better the lower values of the squared shocks. Finally, the model M_4 is better for approximating the higher values of the squared shocks. Consequently, the artificial neural network is able to integrate the behavior of each considered individual model and to produce a more accurate forecast.
- The M_{54}, M_{58} and M_{64} models results, also present a good and balanced MSE (MAD) values in fitting and forecasting samples.

TABLE 5. Evaluation of the volatility forecasting combination by ANNs with different number of neurons in the hidden layer

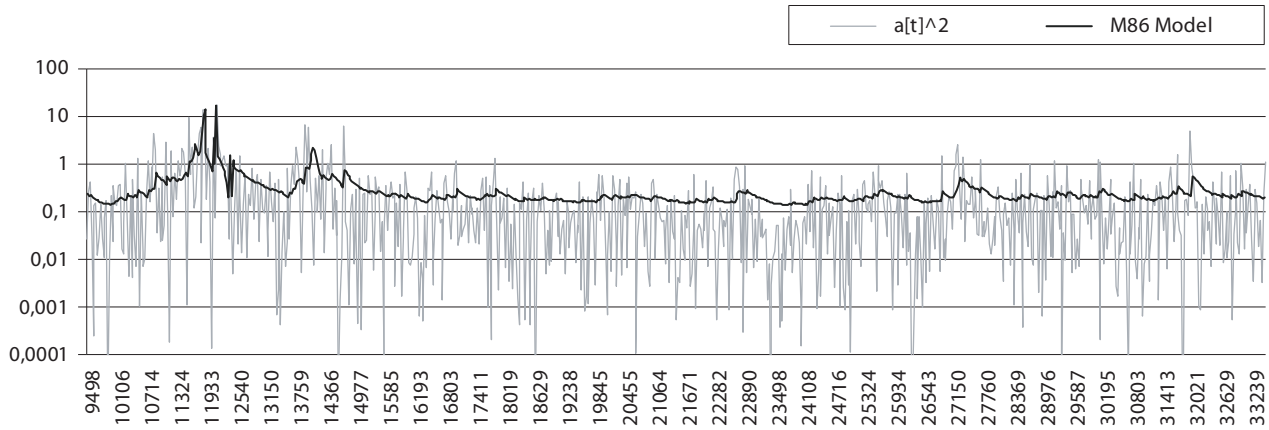
Model	Fitting MSE (MAD)	Forecasting MSE (MAD)
Combining models M_1, M_2, M_3, M_4 and M_5 using ANN		
M_{53}: ANN (H = 1)	10978.0 (38.2)	3724.2 (26.0)
M_{54}: ANN (H = 2)	4521.2 (30.3)	*3667.4 (29.2)
M_{55}: ANN (H = 3)	4383.2 (29.9)	3985.2 (30.1)
M_{56}: ANN (H = 4)	4210.2 (29.1)	4010.7 (32.0)
Combining models M_1, M_2, M_3 and M_4 using ANN		
M_{57}: ANN (H = 1)	11434.8 (38.8)	4159.3 (32.3)
M_{58}: ANN (H = 2)	4521.2 (30.3)	*3665.3 (29.1)
M_{59}: ANN (H = 3)	4351.0 (29.9)	3857.2 (30.2)
M_{60}: ANN (H = 4)	4291.6 (29.2)	3988.4 (29.5)
Combining models M_1, M_2, M_3, and M_5 using ANN		
M_{61}: ANN (H = 1)	10372.0 (39.5)	4378.5 (34.4)
M_{62}: ANN (H = 2)	9507.4 (38.4)	3629.5 (26.0)
M_{63}: ANN (H = 3)	8057.3 (33.9)	3784.0 (31.3)
M_{64}: ANN (H = 4)	5321.1 (34.5)	*3620.8 (28.7)
Combining models M_2, M_3, M_4 and M_5 using ANN		
M_{65}: ANN (H = 1)	11408.8 (38.8)	3941.8 (29.9)
M_{66}: ANN (H = 2)	6529.8 (32.5)	3723.5 (28.1)
M_{67}: ANN (H = 3)	6526.7 (32.4)	*3713.1 (27.9)
M_{68}: ANN (H = 4)	6238.4 (32.5)	3795.0 (27.2)
Combining models M_1, M_2 and M_3 using ANN		
M_{69}: ANN (H = 1)	11694.5 (39.0)	4191.2 (33.1)
M_{70}: ANN (H = 2)	6629.7 (32.6)	*3731.6 (29.3)
M_{71}: ANN (H = 3)	6541.8 (33.5)	3860.3 (30.4)
M_{72}: ANN (H = 4)	6368.2 (33.1)	3911.6 (31.4)
Combining models M_1, M_2, and M_4 using ANN		
M_{73}: ANN (H = 1)	10970.7 (38.7)	3707.2 (25.5)
M_{74}: ANN (H = 2)	10436.2 (40.7)	*3661.6 (33.4)
M_{75}: ANN (H = 3)	10232.8 (39.4)	3723.3 (33.4)
M_{76}: ANN (H = 4)	8009.7 (33.4)	3724.7 (30.7)
Combining models M_1, M_3, and M_4 using ANN		
M_{77}: ANN (H = 1)	11562.2 (39.5)	4180.0 (32.3)
M_{78}: ANN (H = 2)	10438.4 (39.1)	*3603.2 (25.3)
M_{79}: ANN (H = 3)	10329.3 (38.8)	3619.0 (24.6)
M_{80}: ANN (H = 4)	9704.5 (36.6)	3880.7 (27.4)
Combining models M_1, M_3 and M_5 using ANN		
M_{81}: ANN (H = 1)	11660.3 (39.2)	4261.1 (34.1)
M_{82}: ANN (H = 2)	7024.6 (35.5)	3952.9 (32.1)
M_{83}: ANN (H = 3)	6414.3 (32.7)	*3873.4 (30.7)
M_{84}: ANN (H = 4)	6260.8 (33.0)	3915.3 (31.8)
Combining models M_2, M_3 and M_4 using ANN		
M_{85}: ANN (H = 1)	11939.1 (38.3)	3786.2 (29.8)
M_{86}: ANN (H = 2)	4954.1 (30.6)	*3580.6 (27.2)
M_{87}: ANN (H = 3)	4489.7 (30.2)	3878.6 (30.8)
M_{88}: ANN (H = 4)	2619.8 (37.7)	4147.0 (30.7)
Combining models M_2, M_3 and M_5 using ANN		
M_{89}: ANN (H = 1)	11244.5 (38.3)	3895.7 (26.4)
M_{90}: ANN (H = 2)	10391.0 (37.9)	3796.4 (29.1)
M_{91}: ANN (H = 3)	9249.5 (37.7)	*3742.9 (28.9)
M_{92}: ANN (H = 4)	9221.9 (37.8)	3834.3 (30.9)
Combining models M_1 and M_2 using ANN		
M_{93}: ANN (H = 1)	11195.7 (38.7)	3755.5 (25.6)
M_{94}: ANN (H = 2)	6688.4 (32.9)	*3734.4 (29.4)
M_{95}: ANN (H = 3)	6486.6 (32.4)	3803.4 (30.0)
M_{96}: ANN (H = 4)	6483.9 (32.9)	3800.8 (30.6)
Combining models M_1 and M_3 using ANN		
M_{97}: ANN (H = 1)	11278.6 (38.0)	3936.8 (28.7)
M_{98}: ANN (H = 2)	6573.8 (32.8)	3768.5 (29.6)
M_{99}: ANN (H = 3)	6569.0 (32.7)	*3767.7 (29.6)
M_{100}: ANN (H = 4)	6249.9 (32.5)	3733.4 (30.3)
Combining models M_1 and M_4 using ANN		
M_{101}: ANN (H = 1)	11287.6 (38.5)	3888.6 (27.4)
M_{102}: ANN (H = 2)	11203.6 (37.6)	*3684.1 (27.2)
M_{103}: ANN (H = 3)	10283.1 (37.0)	3906.4 (28.3)
M_{104}: ANN (H = 4)	10281.2 (36.7)	3875.0 (28.7)
Combining models M_2 and M_3 using ANN		
M_{105}: ANN (H = 1)	11776.8 (38.2)	3963.8 (31.2)
M_{106}: ANN (H = 2)	6731.2 (34.2)	*3746.0 (30.4)
M_{107}: ANN (H = 3)	6502.6 (32.4)	3763.0 (29.5)
M_{108}: ANN (H = 4)	6481.4 (31.6)	3808.3 (29.2)
Combining models M_2 and M_4 using ANN		
M_{109}: ANN (H = 1)	11646.0 (38.7)	4209.4 (33.2)
M_{110}: ANN (H = 2)	6677.1 (32.7)	*3740.8 (29.4)
M_{111}: ANN (H = 3)	6379.9 (33.3)	3885.4 (31.6)
M_{112}: ANN (H = 4)	6348.8 (32.7)	3836.4 (30.2)
Combining models M_3 and M_4 using ANN		
M_{113}: ANN (H = 1)	11331.5 (38.7)	3871.8 (27.2)
M_{114}: ANN (H = 2)	6462.1 (32.5)	3787.4 (29.8)
M_{115}: ANN (H = 3)	6459.0 (32.4)	3784.5 (29.8)
M_{116}: ANN (H = 4)	6455.6 (32.3)	*3777.6 (29.3)

Continúa...

*Lower values of MSE in the column.

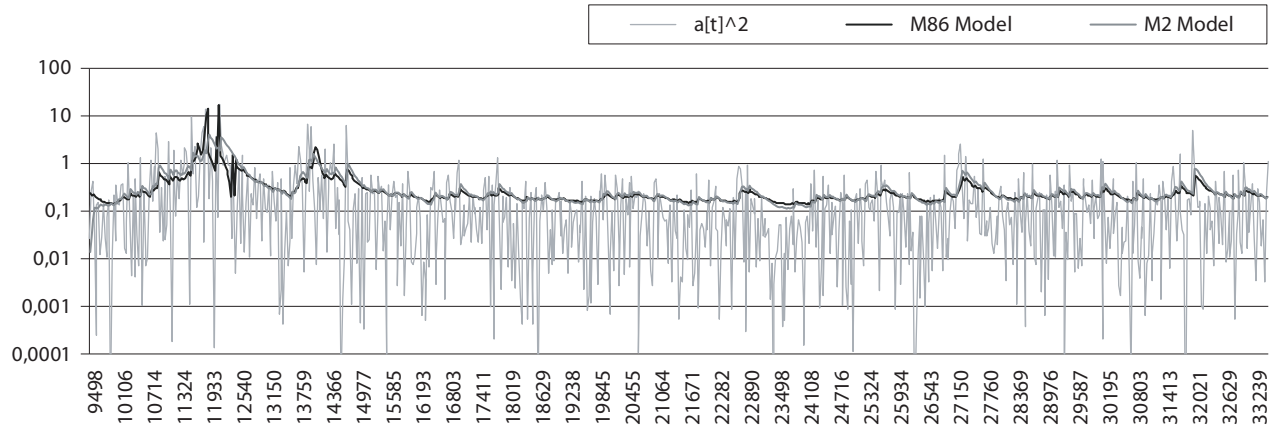
Source: Own elaboration.

FIGURE 2. Conditional variance for monthly excess returns of S&P500 and forecast from the model M_{86} (combination of models M_2, M_3 and M_4 using an ANN with $H = 2$ neurons in the hidden layer).



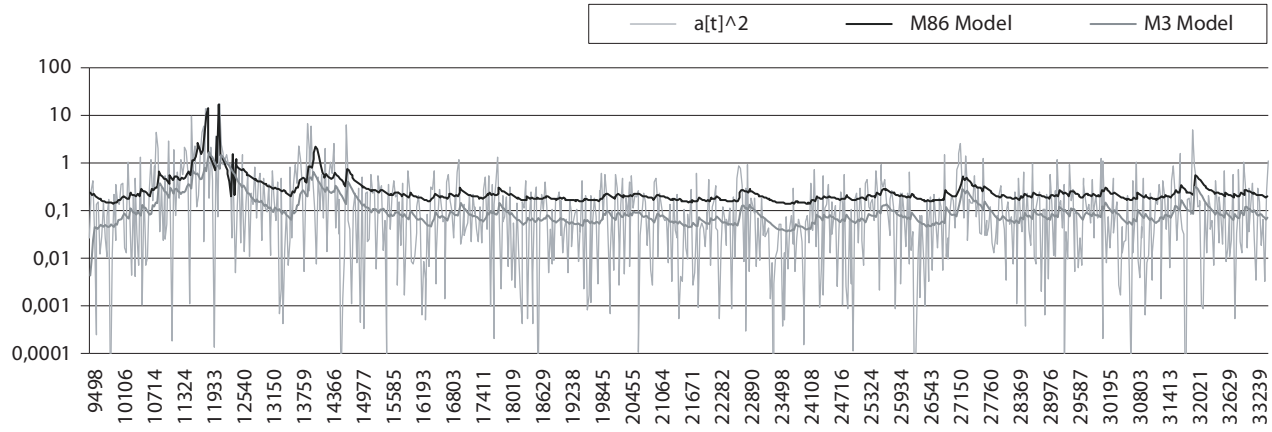
Source: Own elaboration.

FIGURE 3. Comparison of the forecasts calculated using the M_2 and M_{86} models.



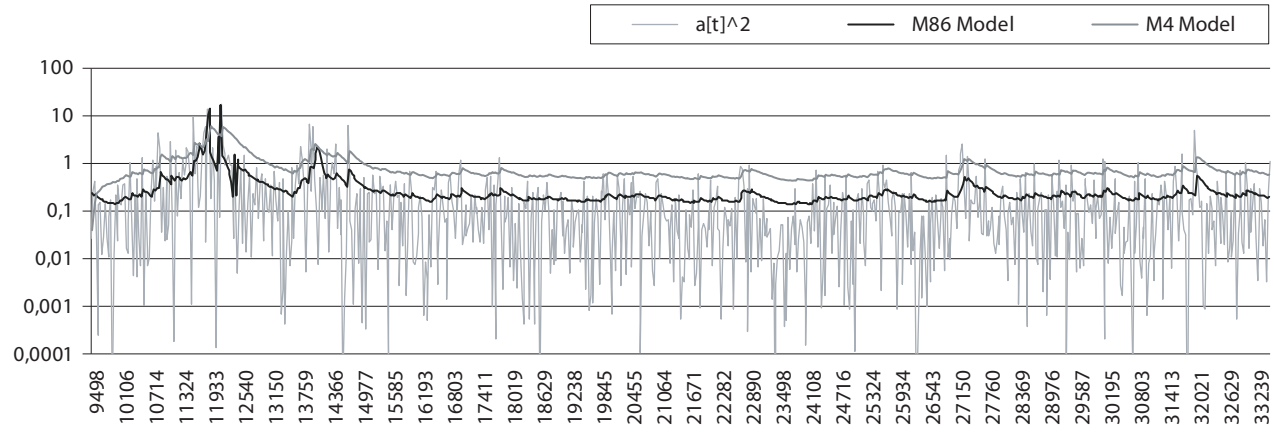
Source: Own elaboration.

FIGURE 4. Comparison of the forecasts calculated using the M_3 and M_{86} models.



Source: Own elaboration.

FIGURE 5. Comparison of the forecasts calculated using the M_4 and M_{86} models.



Source: Own elaboration.

Performance of the preferred model

As we indicated in last section, the M_{86} model is the preferred among all possible combinations by ANN of volatility forecasts, in this section we implement a statistical test to show the difference in accuracy between the preferred model and the standard GARCH model (M_1). We also present an evaluation of the preferred model and the standard GARCH model using Value at Risk evaluation.

Statistical test of predictive accuracy

A demonstration of the effectiveness of the M_{86} model over the M_1 model is shown in a statistical test. A test of the null hypotheses of no difference in accuracy between M_1 and the preferred model is attempted. We implement the test due to Diebold and Mariano (1995), which uses the Morgan-Granger-Newbold statistic for testing of these hypotheses. The test consists in consider the punctual forecasting errors e_{1t} and e_{2t} given by two forecasting models, respectively. The used statistic has a Student's t-distribution with $T - 1$ degrees of freedom. The t statistic is defined by:

$$t = \frac{\rho_{xz}}{\sqrt{\frac{1-\rho_{xz}^2}{T-1}}}$$

Where, $\rho_{xz} = \frac{\sum_t x_t * z_t}{\sqrt{\sum_t x_t^2 * \sum_t z_t^2}}$; $x_t = e_{1t} + e_{2t}$; $z_t = e_{1t} - e_{2t}$, to T, and T is the number of observations.

The Morgan-Granger-Newbold test results show that the null hypothesis of equal predictive accuracy is rejected between M_1 and M_{86} models, at 93% level in training sample with a ρ–value of 7,06694×10^{-54} and in the forecast sample with a ρ–value of 0,0663.

Value at Risk evaluation

In addition, we have compared the GARCH model fitted by maximizing the log L and the preferred model with respect to the value at risk performance (VaR) for the data. The VaR is defined as the maximum potential loss of a portfolio that occurs under normal market conditions with a predefined probability. Assuming that the returns are normally distributed, the 95%-VaR is computed as 95% quantile of the returns distribution:

$$\widehat{VaR}_{t+1} = -1.00 * 1.65 * \hat{\sigma}_{t+1}$$

where $\hat{\sigma}_{t+1}$ is the forecast of the standard deviation given all information until time t.

To evaluate the performance of the 95%-VaR estimators we are using the failure rate criterion. The failure rate (F) is the number of times for which the actual loss is larger than the estimated VaR. This criterion is defined as

$$F = \sum_t D_t$$

where the D_t dummy variable is one if the actual loss is larger than the estimated VaR and zero in other case.

We have used the VaR to evaluate the performance, the failure rate of the M_{86} model is 4.8% and it is low than the failure rate of the M_1 model (6.2%). We concluded that the model obtained from the combination with an ANN can improve the predictive accuracy. The results are shown in Table 6.

TABLE 6. Number of times for which the actual loss is larger than estimated VaR.

Forecast model	M_1	M_{86}
Failure rate	49(6.2%)	38(4.8%)

Source: Own elaboration.

Conclusions

This work considers the forecasting combination as an alternative to reduce the forecast error of the volatility of the monthly excess returns of S&P500. Several GARCH(1,1) models were estimate, by considering different error measures, and combined by the use of simple arithmetical average and ANNs to produce a new model that integrate all information from individual models. The results have been showed this forecasting combination approach has a good ability for predict with major accuracy the volatility in comparison with the traditional forecast models.

As future work, several research addresses emerge. First, consider several datasets with different frequencies and proceeding from different markets. Second, analyze other alternative models for specifying the experts in the model. And third, test and justify the use of other types of loss functions for optimizing the parameters of the models used as experts.

References

Amendola, A., Storti, G. (2008). A GMM procedure for combining volatility forecasts. *Computational Statistics & Data Analysis*, *52*, 3047–3060.

Andersen, T., Bollerslev, T., Christoffersen, P., Diebold, F. (2005a). Volatility forecasting. *PIER Working,* Paper No. 05-011.

Andersen, T., Bollerslev, T., Christoffersen, P., Diebold, F. (2006). Volatility and correlation forecasting. In: Elliott, G., Granger, C.W.J., Timmermann, A. (Eds.), Handbook of Economic Forecasting. Amsterdam: North Holland Press.

Bollerslev T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics.* 31, 307–327.

Bollerslev, T., Engle, R.F. (1993). Common persistence in conditional variances, *Econometrica*, *61*, 167–186.

Brooks, C., Persand, G. (2003). Volatility forecasting for risk management. *Journal of Forecasting*, *22*(1), 1-22

Chan, F., & McAleer, M. (2003). Estimating smooth transition autoregressive models with GARCH errors in the presence of extreme observations and outliers. *Applied Financial Economics*, *13*(8), 581–592.

Christoffersen, P. F., & Diebold, F. X. (1994). *Optimal Prediction Under Asymmetric Loss* (Working Paper No. 167). National Bureau of Economic Research.

Clemen, R. (1989). Combining forecasts: a review and annotated bibliography. *International Journal of Forecasting*, *5*, 559-583.

Clemen, R., Winkler, R. (1986). Combining economic forecasts. *Journal of Business and Economic Statistics*, *4*, 369-391.

Diebold, F. X., & Mariano, R. S. (1995). Comparing Predictive Accuracy. *Journal of Business & Economic Statistics*, *13*(3), 253–263.

Donaldson, R.G., Kamstra, M. (1997). An artificial neural network-GARCH model for international stock return volatitlity. *Journal of Empirical Finance*, *4*, 17-46.

Duan, J. (1997). Augmented GARCH(P, Q) process and its diffusion limit. *Journal of Econometrics*, *79* (1), 97–127.

Dunis, C. L., Huang, X. (2002). Forecasting and trading currency volatility: An application of recurrent neural regression and model combination. *Journal of Forecasting*, *13*, 317–354.

Engle, R.F. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflations. *Econometrica, 50*, 987–1007.

Engle, R.F., Ng, V. (1993). Measuring and testing the impact of news on volatility. *Journal of Finance*, *48*, 1747–1778.

Franses, P. H., Dijk, D. van (2006). *Non-linear time series models in empirical finance*. Cambridge, United Kingdom: Cambridge University Press.

Glosten, L. R., Jagannathan, R., & Runkle, D. E. (1993). On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks. *Journal of Finance*, *48*(5), 1779-1801.

Gooijer, J.G., Hyndman, R.J. (2006). 25 years of time series forecasting. *International journal of forecasting, 22*, 443-473.

Granger, C. W. J. (1989). Combining forecasts—twenty years later. *Journal of Forecasting*, *8*, 167– 173.

Hashem, S. (1997). Optimal linear combinations of neural networks. *Neural Networks, 10*, 599-614.

Haykin, S. (1996). *Neural Networks: A Comprehensive Study.* Ontario Canada: Prentice-Hall,.

Hu, M. Y., C. Tsoukalas. (1999). Combining Conditional Volatility Forecasts using Neural Networks: An Application to the EMS Exchange Rates. *Journal of International Financial Markets, Institutions and Money, 9*, 407-422.

Hull, J., White, A. (1987). The pricing of options on assets with stochastic volatilities. *Journal of Finance*, *42*, 281–300.

Hyndman, R.J., Koehler, A.B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, *22*, 679-688.

Klaassen, F. (2002). Improving GARCH volatility forecasts with regime-switching GARCH. *Empirical Economics*, *27*(2), 363–394.

Lopez, J. A. (2001). Evaluating the predictive accuracy of volatility models. *Journal of Forecasting*, *20*, 87-109.

Lundbergh, S., & Teräsvirta, T. (1998). Modelling economic high-frequency time series with STAR-STGARCH models (*Working Paper Series in Economics and Finance No. 291*).

Nazifi, M., & Fatahi, S. (2012). Regime Switching GARCH Models and GARCH Models, in Stock Market of the Developing Countries' (*SSRN Scholarly Paper No. ID 1987428*). Rochester, NY: Social Science Research Network

Nelson D.B. (1991). Conditional heteroskedasticity in asset returns: a new approach, *Econometrica, 59*, 347–370.

Patton, A.J. (2011). Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics*, *160*, 246-256.

Poon, S.H., Granger, Clive W.J. (2003). Forecasting volatility in financial markets: A review. *Journal of economic literature, 61*, 478-539.

Tang, L.B., Tang, L.X., Sheng, H.Y. (2009). Forecasting volatility based on wavelet support vector machine. *Expert Systems with Applications*, *36*, 2901–2909.

Taylor, J.W. (2004). Volatility forecasting with smooth transition exponential smoothing. *International Journal of Forecasting, 20*, 273–286.

Tsangari, H. (2007). An alternative methodology for combining different forecasting models. *Journal of Applied Statistics*, *34*(4), 403-421.

Tsay, R. (2010). *Analysis of financial time series*. New Jersey: Hoboken, Wiley.

Varian, H. R. (1974). A Bayesian approach to real estate assessment. In *Studies in Bayesian Econometrics and Statistics in Honor of Leonard J. Savage* (pp. 195–208). North-Holland, Amsterdam.

Verhoeven, P., Pilgram, B., McAleer, M & Mees, A. (2002). Non-linear modelling and forecasting of S&P 500 volatility. *Mathematics and Computers in Simulation, 59*, 233-241.

Zellner, A. (1986). Bayesian Estimation and Prediction Using Asymmetric Loss Functions. *Journal of the American Statistical Association*, *81*(394), 446–451.