

Article

Forecasting Selected Colombian Shares Using a Hybrid ARIMA-SVR Model

Lihki Rubio *  and Keyla Alba 

Department of Mathematics and Statistics, Universidad del Norte, Barranquilla 080001, Colombia;
akeyla@uninorte.edu.co

* Correspondence: lihkir@uninorte.edu.co

Abstract: Forecasting future values of Colombian companies traded on the New York Stock Exchange is a daily challenge for investors, due to these stocks' high volatility. There are several forecasting models for forecasting time series data, such as the autoregressive integrated moving average (ARIMA) model, which has been considered the most-used regression model in time series prediction for the last four decades, although the ARIMA model cannot estimate non-linear regression behavior caused by high volatility in the time series. In addition, the support vector regression (SVR) model is a pioneering machine learning approach for solving nonlinear regression estimation procedures. For this reason, this paper proposes using a hybrid model benefiting from ARIMA and support vector regression (SVR) models to forecast daily and cumulative returns of selected Colombian companies. For testing purposes, close prices of Bancolombia, Ecopetrol, Tecnoglass, and Grupo Aval were used; these are relevant Colombian organizations quoted on the New York Stock Exchange (NYSE).

Keywords: hybrid model; ARIMA; support vector regression (SVR); forecasting; time series analysis; daily returns; cumulative returns

MSC: 91B84



Citation: Rubio, L.; Alba, K.

Forecasting Selected Colombian Shares Using a Hybrid ARIMA-SVR Model. *Mathematics* **2022**, *10*, 2181. <https://doi.org/10.3390/math10132181>

Academic Editor: Tatiana Filatova

Received: 13 April 2022

Accepted: 7 June 2022

Published: 23 June 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

A national financial system is a set of institutions and operations through which resources are transferred from people with the capacity to invest to the people or companies that require those investments to carry out their activities [1]. Hence, it is crucial to be aware of those institutions that have the greatest impact on the securities market to make investments. At present, it is easier to be in frequent contact with financial institutions providing services such as taking deposits, granting loans, or managing investments [2].

There are a wide variety of companies in Colombia in which different investors may be interested; four of them (Bancolombia, Ecopetrol, Tecnoglass, and Grupo Aval) are considered to have the highest impact on the country's economy, and it is of great interest for investors to obtain predictive models to understand their behavior and make appropriate decisions. These stocks may be of high relevance to long-term investors interested in solid companies showing a good historical performance. For example, Tecnoglass (TGLS) had, in 2020, a growth of 26.21% over the previous year, generating, during the pandemic, operating profits of 11.9% above a normal year such as 2019, and maintaining the same volume of net income as in the previous year. Figure 1 shows the cumulative returns of TGLS and the satisfactory recovery of this company's shares after COVID-19. The smallest pandemic impact was observed in the Bancolombia shares, which is one of the strongest companies in Colombia with a presence in more than ten countries and a 27-year long listing on the NYSE.

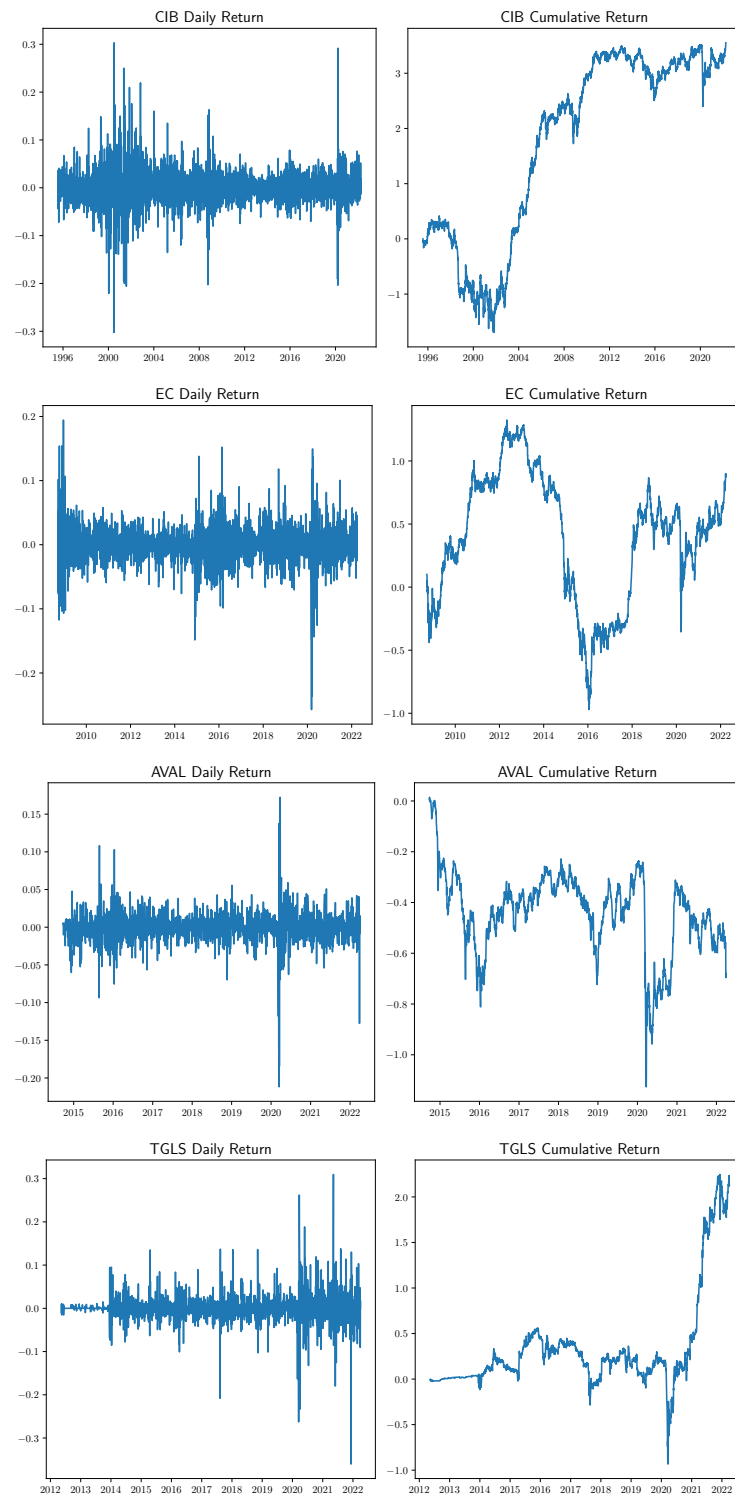


Figure 1. Daily and cumulative returns for Bancolombia, Ecopetrol, Grupo AVAL, and Tecnoglass shares. Source: Yahoo Finance API for Python.

As each of these four companies' shares have been preferred over the years, confidence and trading volume can be attractive when investing in the stock market [3], as can the study of appropriate predictive models. There are different methods or techniques for identifying future movements of a stock, such as fundamental and technical analysis, as well as the use of statistical estimations to produce predictions for making correct decisions [4,5]. The main objective of this research is to make use of a hybrid model to obtain robust predictions of the

daily and cumulative returns of shares in Bancolombia, Ecopetrol, Tecnoglass, and Grupo Aval. Hybrid models have the advantage of capturing linear and non-linear patterns in a time series; for this reason, in this work, this type of analysis is used, and its effectiveness is verified through accuracy and goodness-of-fit metrics to confirm the benefits of using a combination of predictive techniques instead of using each one separately. Predictions obtained in this research can be used to forecast other NYSE-listed stocks since some of them exhibit similar behaviors in the time series studied in this work for testing purposes.

2. Literature Review

The support vector regression (SVR) model is a promising tool for the financial forecasting of stocks with large fluctuations, while (ARIMA) models are remarkable for fitting linearity without sacrificing trend predictions for long-term horizons, which is crucial, for example, for portfolio optimization [6]. Due to its effectiveness for solving non-linear estimation problems, SVR models have been widely used for time series forecasting in different fields, e.g., for wind power and financial forecasting, as well as for predicting industrial machinery production value and accuracy [7–11]. Hence, in this paper, we use the SVR model as a component of the hybrid model, as it has shown to be effective for time series predictions.

The advantage of using ARIMA models is their versatility and the fact that they can be adjusted to predict different temporal phenomena [12]. It is possible to adjust a variety of time series data using either ARs (autoregressive models), MAs (moving averages), or ARMA, which is a combination of AR and MA. Due to its effectiveness as a classical forecasting method, it has been widely used, for example, to predict water quality, workloads in cloud applications, the EBITDA index for financial performances, and short-term customer loads [13–16]. Since ARIMA is a linear model, it cannot capture non-linear patterns in a time series; therefore, different predictive models based on machine learning and deep learning techniques have been used in time series forecasting to predict strong fluctuations.

Artificial neural networks (ANNs) have recently gained considerable interest from researchers to predict time series. The main benefit of using neural networks is their ability to forecast non-linear behaviors in data [17]. When using ANNs, in contrast to traditional linear models, it is not necessary to describe a concrete model format. Instead, the model is built and fitted based on data characteristics. Such a model-based design is appropriate for a wide range of experimental datasets without a set of rules to provide appropriate data-generation methods.

The use of hybrid models or the combination of a pair of techniques to enhance forecast performance is one of the most powerful alternatives to improve prediction errors [18]. Hybrid models are either homogeneous, such as with the use of a variety of neural network settings, or heterogeneous, using both linear and non-linear techniques [19,20]. This can be of great benefit for time series predictions, since time series can show non-linear behavior over time; however, it can become linear according to input considerations. For these types of hybrid models, while the machine learning model handles non-linearity, ARIMA handles the linear, non-stationary part [21,22].

Hybrid prediction strategies have been used for time series forecasting in recent decades, such as the hybrid ARIMA–ANN model implemented in [21], where it was noted for its effectiveness in predicting linearity and nonlinearity in time series. The hybrid ARIMA–BPNN, used to make predictions about the Chinese stock market, was based on the use of the ARIMA model and backpropagation neural networks (BPNNs) to effectively capture linear and nonlinear behaviors hidden in the original dataset and improve the forecast accuracy obtained by each model separately [23]. Another hybrid model based on ARIMA is the ARIMA–LSTM, which retains the advantages of the two models in addition to the efficiency of approximating of the seasonal component provided by the LSTM model. This hybridization provides a robust prediction under certain parameterizations [24].

Recently, a study on gold price forecasting with ARIMA and SVR models showed that the SVR model is better than the ARIMA model, and suggests that SVR should be used in any commodity price forecasting due to its high accuracy. All previous studies of individual and hybrid models have been useful for predicting time series, and many of these comparisons and applications were used to formulate new hybrid models in order to improve prediction results [25–27].

On the other hand, some works have been published using a hybrid approach comprising ARIMA and artificial intelligence for financial forecasting [28]. SARIMABP, a hybrid model using SARIMA and a backpropagation neural network model, has been used to forecast seasonal time series data [22]. The hybrid method, called KARIMA, which uses a Kohonen self-organized map and the ARIMA model for short-term forecasting [29]. Using fundamental artificial neural network principles, a hybrid model to overcome the data limitations of ANNs and achieve more precise predictions than ANNs under incomplete data scenarios has been addressed by employing fuzzy regression models [30].

Due to the effectiveness shown by the use of the hybridization of different types of predictive time series models, in this paper, the use of a hybrid ARIMA–SVR model to predict the daily and cumulative returns of a Bancolombia, Ecopetrol, Tecnoglass, and Grupo Aval is proposed, since it has been shown by different authors that forecasting models can complement each other to greatly improve the accuracy of their forecasts, which is extremely important in portfolio optimization, risk management, and decision making. An additional motivation for using the hybrid model is that time series data from the real market are seldom linear or non-linear; frequently, they consist of both linear and non-linear patterns. Therefore, it is problematic for predictors to select the appropriate approach for their particular scenario. Mixing different methodologies can solve the issue of choosing the right model. Therefore, by using a hybrid ARIMA–SVR model, intricate autocorrelation patterns can be predicted with higher precision.

3. Materials and Methods

For testing purposes, each model will be used separately, and then a hybrid model will be built to take advantage of the benefits of each model to handle linearity and nonlinearity present in the original data series. Time series data for Bancolombia, Ecopetrol, Tecnoglass, and Grupo Aval were obtained using the Yahoo Finance API. For the development of the application, a computer with 16 GB of RAM, an 8th Gen. Intel Core i7 Processor, an NVIDIA GeForce GTX 1070 with Max-Q Design, 8 GB GDDR5 was used, running on the x64-bit Microsoft Windows 10 platform. The algorithms for each were implemented using the Python language.

3.1. Data and Sampling

To validate what is proposed in this work, a forecast study of Bancolombia, Ecopetrol, Tecnoglass, and Grupo Aval's daily and cumulative returns is performed using a hybrid model that uses the autoregressive integrated moving average (ARIMA) together with support vector regression (SVR).

Bancolombia has been listed on the Colombian Stock Exchange (BVC) for more than 20 years and is one of the leading institutions in terms of market capitalization. Its presence is focused on the American continent, and it has been consolidated in the Colombian capital market since 1981 [31]. Ecopetrol S.A. Empresa Colombiana de Petróleos, on the other hand, is a semi-public company, specializing in oil and gas production, which is involved in every phase of the hydrocarbon supply chain—from the extraction of raw materials to refining and commercialization. Ecopetrol's shares are currently traded on the Colombian and New York Stock Exchanges [32,33]. Tecnoglass S.A. is a leading glass manufacturer located in Barranquilla, Colombia. It is the second major glass supplier for the United States and is one of the most important companies dedicated to architectural glass processing in Latin America. It was indexed in Nasdaq on 20 December 2013 [34,35]. Grupo Aval S.A. (formerly known as Grupo Acciones y Valores S.A.) (NYSE: AVAL, BVC: GRUPOAVAL) is a Colombian busi-

ness conglomerate engaged in a wide variety of activities, mainly financial. It is Colombia's leading financial group, controlling the four most-important Colombian banks—(Banco de Bogotá, Banco de Occidente, AV Villas, and Banco Popular). In 2014, the company went public at the NYSE after a process which was initiated in 2004, with the issuance of 1874 million preferred shares in the form of ADRs at USD 13.50 per share, representing a value of USD 1265 million [36,37].

The sample size used for predictions corresponds to the last 4-year daily observations, until 1 April 2022, of each share appearing in Yahoo Finance. Daily returns are calculated using the following formula:

$$R_t = \frac{P_t - P_{t-1}}{P_{t-1}},$$

where P_t is the close price of each share at time t . The time series data were divided into training and testing datasets. For the testing set, the last ten days (10%) were considered, and the remaining 90% was used to train each model. The variables that were studied were date and the value of the daily and cumulative returns. Datasets were obtained using the Yahoo Finance API for Python.

Next, the models obtained were optimized by choosing parameters that provided the best adjustment, in terms of Akaike's criterion and hyper-parameterization for the SVR model. Accuracy measures, such as the MAE (mean absolute error), MAPE (mean absolute percentage error), MSE (mean square error), and RMSE (root-mean-square error), were used to compare the model results. Correlation and the calculation of R^2 were used to identify the best fitting for each model; for determining the significance of the differences in predictions, the Diebold–Mariano (DM) test was used. Results obtained by the best forecasting model are presented, and their advantages are discussed in terms of error and goodness of fit.

3.2. ARIMA Model

The ARIMA models, pioneered by Box and Jenkins, are powerful and useful forecasting techniques to predict time series [38]. Essentially, the ARIMA models estimate future values of the time series as a linear combination of historical and residual data. There are three components in such a model: the differentiation order for seasonality, and autoregressive and moving average orders [39].

The backward shift operator \mathcal{B} is used as an operator to remove non-stationarity. A time series, y_t , is called homogeneous non-stationary if its first difference, $w_t = (1 - \mathcal{B})y_t = y_t - y_{t-1}$, or the d th difference, $w_t = (1 - \mathcal{B})^d y_t$, is also a stationary time series. Furthermore, y_t is referred to as an autoregressive integrated moving average (ARIMA) model with orders p, d , and q , noted $\text{ARIMA}(p, d, q)$, if its d th difference produces a stationary $\text{ARMA}(p, q)$ procedure. Hence, an $\text{ARIMA}(p, d, q)$ is often expressed as

$$\phi(\mathcal{B})(1 - \mathcal{B})^d y_t = c + \theta(\mathcal{B})\varepsilon_t \quad (1)$$

where

$$\phi(\mathcal{B}) = 1 - \sum_{i=1}^p \phi_i \mathcal{B}^i \quad \text{and} \quad \theta(\mathcal{B}) = 1 - \sum_{i=1}^q \theta_i \mathcal{B}^i \quad (2)$$

the backward shift operators for $\text{AR}(p)$ and $\text{MA}(q)$ are defined as $\phi(\mathcal{B})y_t = c + \varepsilon_t$ and $y_t = \mu + \theta(\mathcal{B})\varepsilon_t$, with $c = \mu - \phi\mu$, where μ and ε_t are the mean and white noise, respectively, and $E(\varepsilon_t) = 0$.

The orders for the ARIMA models can be calculated using the autocorrelation and partial autocorrelation structures. Coefficients are computed by the maximum likelihood approach pioneered by Box and Jenkins [12]. From evaluative checks such as Akaike's information criterion (AIC) and the Jarque–Bera normality test for residual series, it is identified as the optimal model.

3.3. Support Vector Regression Model

The support vector regression (SVR) model is used for nonlinear regression problems, and it is a novel technique in time series forecasting [40]. SVR uses a linear function hypothesis space in a high-dimensional feature space to be trained according to the principle of structural risk minimization (SRM) [41,42]. SVR inputs are mapped into a high-dimensional feature space by using a non-linear function [43,44]. The decision function for the SVR model can be expressed as

$$y = w \cdot \phi(x) + b, \quad (3)$$

where x is the input, w and b are constant vectors, and $\phi(x)$ is a nonlinear function. The principal objective of SVR algorithms is to find the best parameters w, b in Equation (3), to solve an optimization problem described as follows

$$\min \left\{ \frac{1}{2} w^2 + C \sum_{i=1}^n (\zeta_i + \zeta_i^*) \right\} \quad \text{subject to:} \quad \begin{aligned} y_i - (w \cdot \phi(x_i) + b) &\leq \varepsilon + \zeta_i \\ (w \cdot \phi(x_i) + b) - y_i &\leq \varepsilon + \zeta_i^* \\ \zeta_i, \zeta_i^* &\geq 0, \quad i = 1, \dots, n \end{aligned} \quad (4)$$

where ζ_i and ζ_i^* are slack variables that measure the training errors above and below an ε tube, and C is a positive constant penalty coefficient that determines the degree of penalized loss when a training error occurs. The nonlinear SVR can also be expressed as a dual pattern using the Lagrange multipliers after minimizing Equation (4); then, the dual optimization problem of nonlinear SVR changes to the following

$$\begin{aligned} &\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) \mathcal{K} + \varepsilon \sum_{i=1}^n (\alpha_i - \alpha_i^*) - \sum_{i=1}^n y_i (\alpha_i - \alpha_i^*), \\ &\text{with constraints} \quad \sum_{i=1}^n (\alpha_i - \alpha_i^*) = 0 \\ &\quad 0 \leq \alpha_i, \alpha_i^* \leq C, \quad i = 1, \dots, n, \end{aligned} \quad (5)$$

where ζ_i and ζ_i^* are Lagrange multipliers with a \mathcal{K} kernel function, to be calculated as follows: $\mathcal{K} := \mathcal{K}(i, j) = \phi(x_i)^T \phi(x_j)$. The kernel functions widely used in SVR include linear, radial, polynomial, and sigmoid functions [45,46]. Lastly, the nonlinear (SVR) decision function can be stated as

$$f(x_i) = \sum_{i=1}^n (\alpha_i - \alpha_i^*) \mathcal{K}(x_i, x_k) + b. \quad (6)$$

3.4. Hybrid Model

Obtaining reliable predictions of daily and cumulative return values with a forecasting model is crucial for making correct decisions. Khashei and Bijari (2011) [47] propose a hybrid model similar to Zhang (2003) [21] in which the ARIMA model extracts linear components and predicts them; residuals, being nonlinear data subcomponents, are then fit by the SVR model. Therefore, the hybrid model in this work consists of ARIMA and SVR and predicts linear and nonlinear behaviors with the best accuracy forecasts. The hybrid model \mathcal{H}_t is, then, expressed as

$$\mathcal{H}_t = \mathcal{L}_t + \mathcal{N}_t, \quad (7)$$

where \mathcal{L}_t and \mathcal{N}_t represent the linear and non-linear components for the hybrid technique and are calculated based on the original time series. Let $\hat{\mathcal{L}}_t$ be prediction resulting from the use of the ARIMA model in the original dataset at time t ; then, the residual ε_t , defined as

$$\varepsilon_t = \mathcal{H}_t - \hat{\mathcal{L}}_t, \quad (8)$$

is adjusted using the SVR model and can be expressed as

$$\varepsilon_t = \mathcal{F}_{SVR}(\varepsilon_{t-1}, \varepsilon_{t-2}, \dots, \varepsilon_{t-n}) + \Delta_t, \quad (9)$$

where \mathcal{F}_{SVR} is a nonlinear expression related to the SVR model and Δ_t is the random error. Thus, a hybrid model can be expressed as

$$\hat{Z}_t = \hat{\mathcal{L}}_t + \hat{N}_t, \quad (10)$$

where $\hat{\mathcal{L}}_t$ and \hat{N}_t are the forecast values for linearity and non-linearity using the ARIMA and SVR models, respectively.

4. Results and Discussion

4.1. Prediction Evaluation Measures

To identify the most appropriate approach to the original time series, errors were determined based on the mean absolute percentage error (MAPE), the mean absolute deviation (MAD), and the mean squared deviation (MSD) measures, each determined using the following formulas:

$$\begin{aligned} \text{MAPE} &= \frac{\sum_{t=1}^{T_f} |(y_t - \hat{y}_{t-1}) / y_t|}{T_f} \times 100 \\ \text{MAD} &= \frac{\sum_{t=1}^{T_f} |y_t - \hat{y}_{t-1}|}{T_f} \\ \text{MSD} &= \frac{\sum_{t=1}^{T_f} (y_t - \hat{y}_{t-1})^2}{T_f}. \end{aligned} \quad (11)$$

For significance of differences in predictions, the Diebold–Mariano (DM) test was used. It is an asymptotic hypothesis test for the mean of the loss series

$$\delta_k^{\mathcal{M}_1, \mathcal{M}_2} = \mathcal{L}(\epsilon_k^{\mathcal{M}_1}) - \mathcal{L}(\epsilon_k^{\mathcal{M}_2}),$$

where $\epsilon_k^{\mathcal{M}} = y_k - \hat{y}_k$ is the forecasting error using the model \mathcal{M} at the time step k , with $\mathcal{L}(\cdot)$ as the loss function, generally taken as $\mathcal{L}(\epsilon_k^{\mathcal{M}}) = |\epsilon_k^{\mathcal{M}}|^p$, $p = 1, 2$. Considering a forecast from model \mathcal{M}_1 and a forecast from model \mathcal{M}_2 , the null hypothesis to be contrasted is H_0 : the loss differential mean for model \mathcal{M}_1 is less than or equal to that of model \mathcal{M}_2 [48].

4.2. Application of Models

Descriptive parameters are calculated to obtain important information related to measures of central tendency, dispersion, and position (see Table 1). According to the results, it can be seen that the most volatile stocks are Bancolombia and Ecopetrol, and they are also the ones with the longest trading history on the NYSE. The distribution of price frequencies for each of these stocks is approximately symmetric, except for Ecopetrol, which is right-skewed. In terms of the total volume of shares traded, Bancolombia and Tecnoglass have increased their value the most. Bancolombia's shares are the most expensive overall, and Grupo AVAL's shares are the cheapest.

Table 1. Descriptive parameters for each share's closing price: Bancolombia, Ecopetrol, Grupo AVAL, and Tecnoglass.

	Bancolombia	Ecopetrol	Grupo AVAL	Tecnoglass
$N_{records}$	6721	3410	1897	2492
μ	29.6831	25.3802	7.5447	10.7590
σ	19.9106	15.7001	1.7814	4.6862
y_{min}	1.0500	5.4000	3.3400	2.2900
Q_1	12.2500	12.6050	6.2700	8.2975
\tilde{x}	30.2700	19.1200	7.7900	9.9400
Q_3	45.1600	38.7575	8.4700	11.5600
y_{max}	70.5000	67.4800	13.7800	33.7000

4.2.1. Application of ARIMA Model

The standard statistical paradigm assumes that a series under stationary conditions and a model that fits the historical data is also the optimal model for forecasting [49]. The cumulative returns of Bancolombia, Ecopetrol, Grupo AVAL, and Tecnoglass show non-stationary signals, with p -values greater than 0.05, as evidenced in Table 2. The non-stationarity null hypothesis cannot be rejected, so each share's cumulative return shows a non-stationary signal; consequently, a right differentiator order should be selected to make it stationary. The autocorrelation function for the cumulative returns, on the right, shows an exponential decay, indicating a trend in the data and a positive correlation for a large number of lags, which requires further differencing. Autocorrelation for daily returns on the left shows the lag moves into the far-negative zone rapidly, indicating the stationarity of the time series (see Figure 2).

Table 2. Dickey–Fuller test for the non stationarity of cumulative returns.

Share	Dickey–Fuller Test
Bancolombia (CIB)	p -value: 0.8396
Ecopetrol (EC)	p -value: 0.9874
Grupo AVAL (AV)	p -value: 0.0678
Tecnoglass (TGLS)	p -value: 0.3776

A first and second differentiator is applied until stationarity is reached; from the first differentiator applied, stationarity is achieved. The ARIMA model is trained to confirm the right order and differentiation and to have the best model orders (p, d, q) . According to the information-theoretic approaches, the Akaike goodness-of-fit (AIC) criterion was used, which provides the best coefficients for model quality, based on the minimization of the AIC coefficient [50]. The following were parameters obtained for the ARIMA model from the training datasets of each share studied here, namely, CIB, EC, AVAL, and TGLS: $(1, 0, 1)$, $(1, 0, 2)$, $(2, 0, 0)$, and $(4, 0, 4)$ for daily returns, and $(1, 1, 0)$, $(2, 1, 3)$, $(0, 1, 0)$, and $(3, 1, 3)$ for cumulative returns, respectively. The first parameter p is the autoregressive order, the second d is the differentiation order, and the moving average order q is the third. Parameters were calculated by choosing the lowest Akaike coefficient after a loop over pre-defined ranges for each of these orders.

Figure 3 shows the model adjustment for the real time series of the daily and cumulative returns. Correlation plots between the original and predicted time series are plotted with the R^2 respective value. Figures 3 and 4 show the goodness of fit for daily and cumulative returns, which can be confirmed with the R^2 and correlation plot. Accuracy metrics, such as MAE, MSE, MAPE, and RMSE, were applied to evaluate the efficiency of the ARIMA model for each share, as can be seen in Tables 3 and 4.

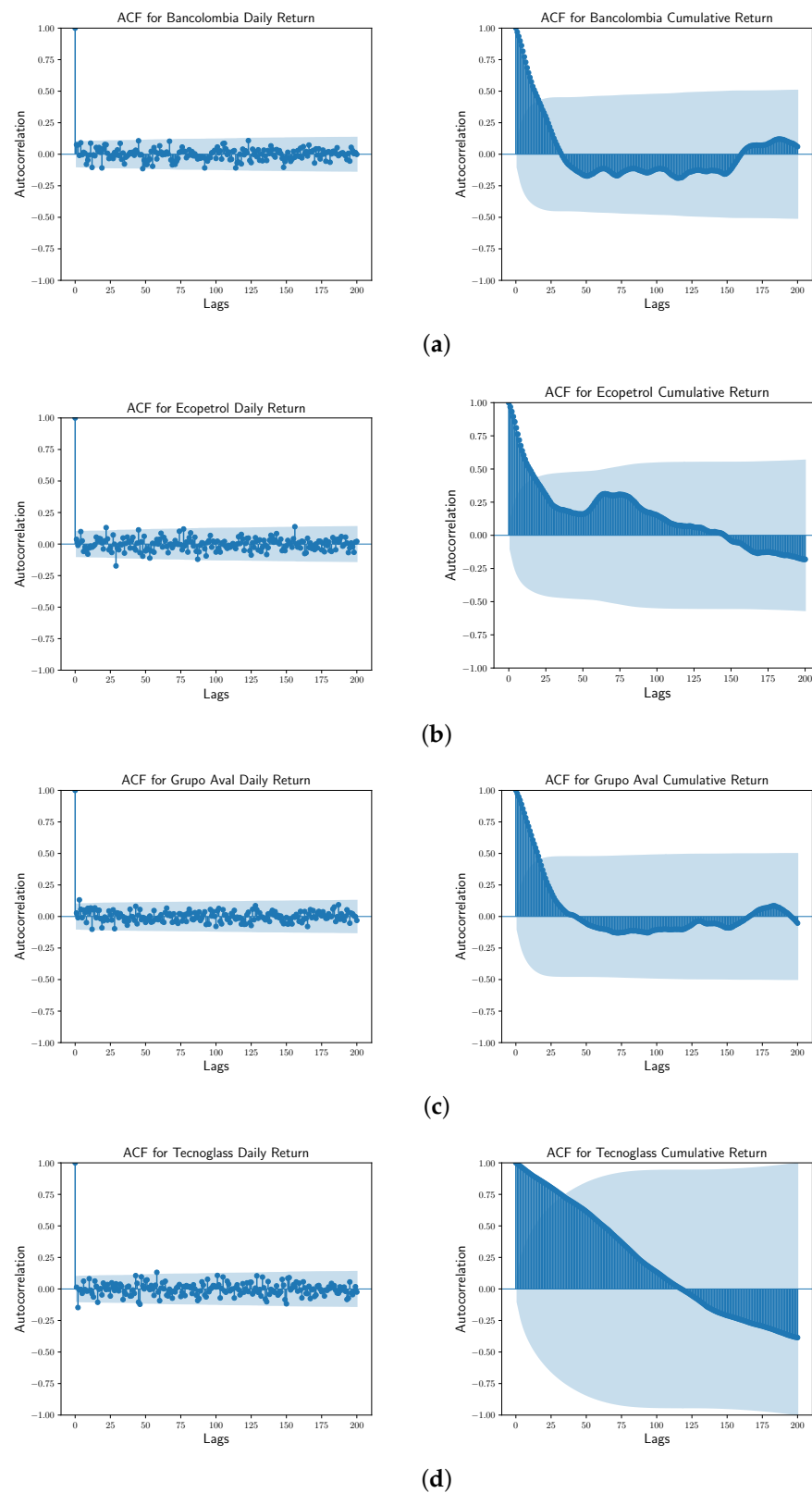


Figure 2. Autocorrelation plots for (a) Bancolombia, (b) Ecopetrol, (c) Grupo Aval, and (d) Tecnoglass time series data—daily returns (left) and cumulative returns (right). The dashed blue line specifies a significance threshold. Lags, consistently outside the pair of dashed blue lines, represent non-stationary trends.

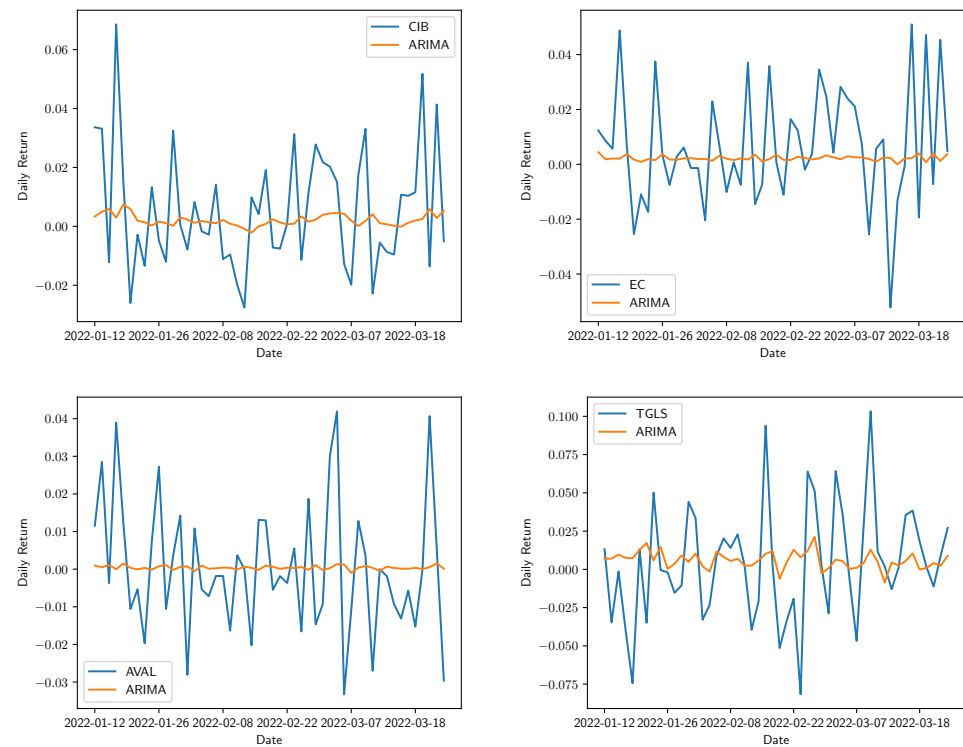


Figure 3. Real vs. ARIMA adjustment for daily returns of Bancolombia, Ecopetrol, Grupo AVAL, and Tecnoglass. The real time series is represented by the share symbol and its prediction with the model name in the legend.

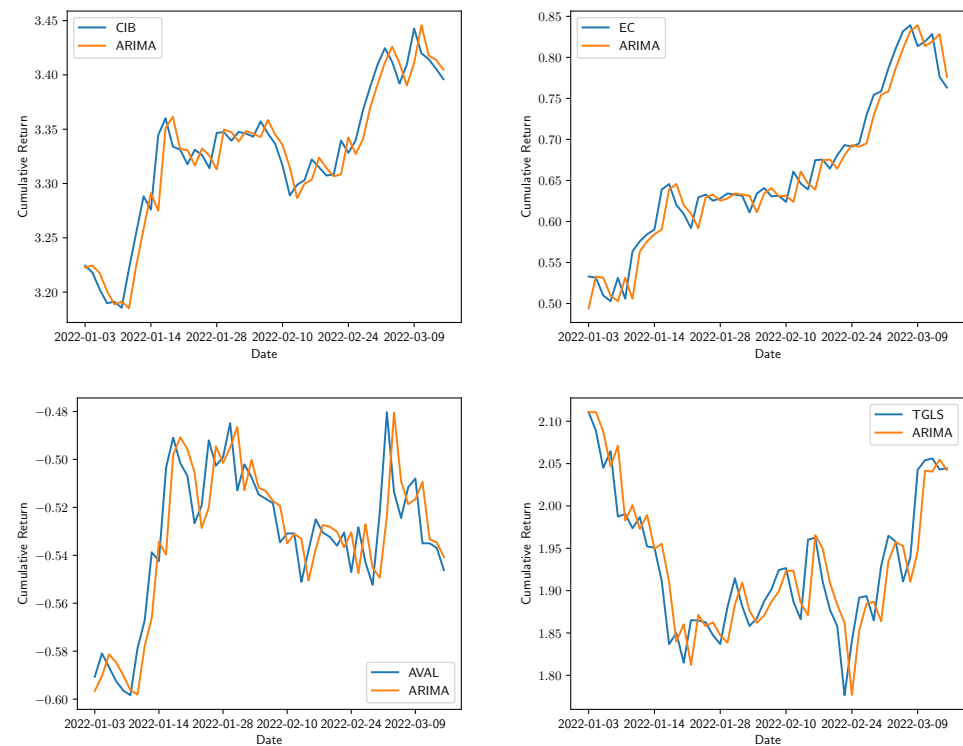


Figure 4. Real vs. ARIMA adjustment for cumulative returns of Bancolombia, Ecopetrol, Grupo AVAL, and Tecnoglass. The real time series is represented by the share symbol and its prediction with the model name in the legend.

Table 3. MAE, MSE, and R^2 for the daily returns forecast using the ARIMA model. The Diebold–Mariano test was used for the prediction pair $(DM, p\text{-value})_{ARIMA, Hybrid}$. MAE and MSE columns are multiplied by 10^{-7} .

Share	MAE	MSE	R^2	$(DM, p\text{-Value})_{ARIMA, Hybrid}$
Bancolombia (CIB)	156,642.5839	4209.16114	0.0630787	$(7.7018, 1.3424 \times 10^{-13})$
Ecopetrol (EC)	164,542.6646	4563.8175	0.0031868	$(8.8210, 5.1582 \times 10^{-17})$
Grupo AVA1 (AV)	118,944.0264	2435.5536	0.0418544	$(8.4931, 5.5444 \times 10^{-13})$
Tecnoglass (TGLS)	330,208.9796	26,142.6679	0.1570711	$(4.0985, 5.1533 \times 10^{-13})$

Table 4. MAE, MSE, and R^2 for the cumulative returns forecast using the ARIMA model. The Diebold–Mariano test was used for the prediction pair $(DM, p\text{-value})_{ARIMA, Hybrid}$. MAE and MSE columns are multiplied by 10^{-7} .

Share	MAE	MSE	R^2	$(DM, p\text{-Value})_{ARIMA, Hybrid}$
Bancolombia (CIB)	156,412.9033	4258.5618	0.9392642	$(7.1732, 4.4188 \times 10^{-12})$
Ecopetrol (EC)	156,730.7965	4104.2421	0.9658714	$(8.4510, 7.9365 \times 10^{-16})$
Grupo AVA1 (AV)	123,512.9019	2546.7269	0.7500343	$(8.0887, 1.0073 \times 10^{-14})$
Tecnoglass (TGLS)	325,112.7213	24,722.3193	0.8636632	$(4.0987, 5.1714 \times 10^{-5})$

Figure 3 shows the ARIMA model’s performance in forecasting the linear behaviors of the daily and cumulative returns of each share, maintaining the trend for a long-term horizon; however, it is necessary to use another forecasting model to capture non-linear patterns. The support vector regression model (SVR) was applied for this goal.

4.2.2. Support Vector Regression Application

The support vector regression model is used to predict original datasets, and then to forecast residuals from the ARIMA model, considering ARIMA as a linear pre-filter. Parameters for the best fitted model were calculated with a grid search methodology based on the training dataset; parameter values with the minimum MSE values were used to determine the final model.

Figures 5 and 6 show the SVR prediction for the training sets using the best kernel and parameters after cross-validation was applied. The model with the RBF kernel provided the best goodness of fit compared to the linear, polynomial, and sigmoid kernels. Accuracy metrics were used to compute forecasting errors (see Tables 5 and 6), correlation, and R^2 to study the forecasting quality of the selected SVR model.

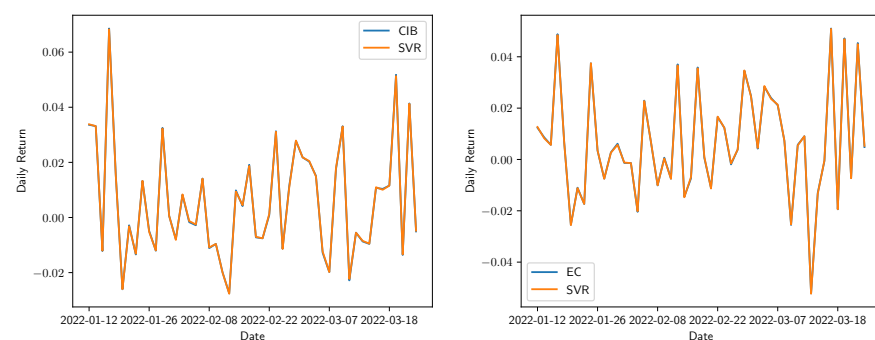


Figure 5. Cont.

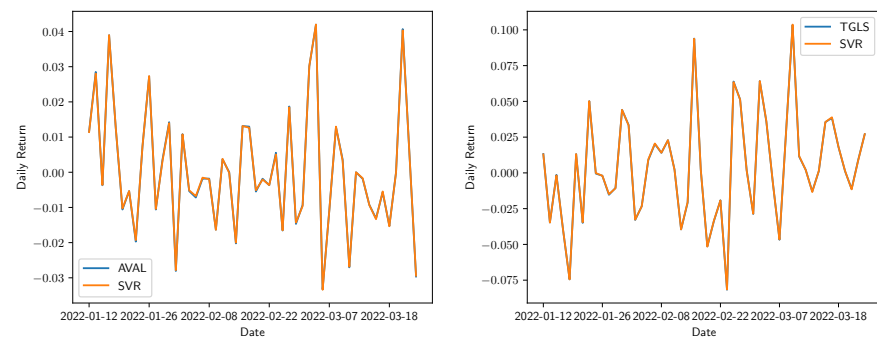


Figure 5. Real vs. SVR adjustment for the daily returns of Bancolombia, Ecopetrol, Grupo AVAL, and Tecnoglass. The real time series is represented by the share symbol and its prediction with the model name in the legend.

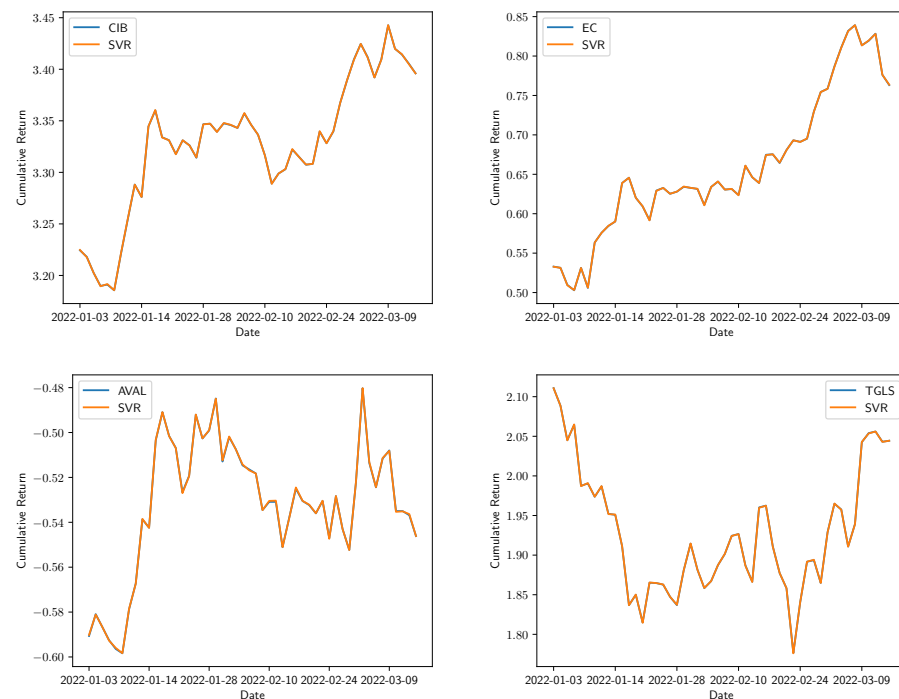


Figure 6. Real vs. SVR adjustment for the cumulative returns of Bancolombia, Ecopetrol, Grupo AVAL, and Tecnoglass. The real time series is represented by the share symbol and its prediction with the model name in the legend.

Table 5. MAE, MSE, and R^2 for the daily returns forecast using the SVR model. The Diebold–Mariano test was used for the prediction pair $(DM, p\text{-value})_{SVR, Hybrid}$. MAE and MSE columns are multiplied by 10^{-7} .

Share	MAE	MSE	R^2	$(DM, p\text{-Value})_{SVR, Hybrid}$
Bancolombia (CIB)	1896.1414	0.5732	0.9998737	(0.1316, 0.8954)
Ecopetrol (EC)	2046.0967	0.6497	0.9998610	(−1.2238, 0.2218)
Grupo AVAL (AVAL)	1849.0694	0.5449	0.9997991	(1.6671, 0.0964)
Tecnoglass (TGLS)	2054.6102	0.6239	0.9999799	(1.0358, 0.3010)

Table 6 shows the accuracy metrics for the (SVR) model using a radial kernel, with C and ϵ and γ as the cross-validation parameters. Results in Figures 5 and 6 show a considerable fit to the daily returns using the selected model with parameters $C = 0.1$, $\epsilon = 10^{-6}$, $\gamma = 1.0$ for Bancolombia and Grupo Aval, and $C = 1.0$, $\epsilon = 10^{-6}$, $\gamma = 1.0$ for

Ecopetrol and Tecnoglass. The ARIMA and SVR models, separately, cannot handle all patterns in the time series. However, the hybrid model has the potential of minimizing errors considerably, as it is a combination that uses parameters from each model to capture what is significant separately.

Table 6. MAE, MSE, and R^2 for the cumulative returns forecast using the SVR model. The Diebold–Mariano test was used for the prediction pair $(DM, p\text{-value})_{SVR, Hybrid}$. MAE and MSE columns are multiplied by 10^{-7} .

Share	MAE	MSE	R^2	$(DM, p\text{-Value})_{SVR, Hybrid}$
Bancolombia (CIB)	2043.3500	0.6387	0.9999908	(−1.4533, 0.1470)
Ecopetrol (EC)	2184.8263	0.7228	0.9999939	(−1.0231, 0.3069)
Grupo AVAL (AVAL)	1968.4033	0.5853	0.9999378	(0.9748, 0.3304)
Tecnoglass (TGLS)	1916.5194	0.5896	0.9999967	(1.9944, 0.0469)

4.2.3. Application of the Hybrid Model

For the proposed hybrid model, ARIMA was used as a pretreatment for line filtering in the dataset. Error terms or residuals from the ARIMA model were then fed into the SVR model. The parameters with minimum MSE scores were chosen to fit the ARIMA residuals. The SVR model was used to predict non-linear patterns that ARIMA, as a statistic linear model, failed to capture.

The hybrid model estimation $\hat{\mathcal{H}}_t$ can be expressed as

$$\hat{\mathcal{H}}_t = \hat{\mathcal{L}}_t + \hat{\mathcal{N}}_t,$$

where $\hat{\mathcal{L}}_t$ is the linear term modeled by the ARIMA model and $\hat{\mathcal{N}}_t$ is the non-linearity; the residuals from ARIMA were predicted by the SVR model. Let ε_t be the residual at t , derived from ARIMA; then,

$$\varepsilon_t = \mathcal{H}_t - \hat{\mathcal{L}}_t, \quad (12)$$

are modeled using SVR and can be expressed in the following way:

$$\varepsilon_t = \mathcal{F}_{SVR}(\varepsilon_{t-1}, \varepsilon_{t-2}, \dots, \varepsilon_{t-n}) + \Delta_t, \quad (13)$$

where \mathcal{F}_{SVR} is the non-linear function associated with the SVR model with random errors Δ_t . Figures 7 and 8 show ARIMA-SVR adjustment for daily and cumulative returns of Bancolombia, Ecopetrol, Grupo AVAL, and Tecnoglass. Autocorrelation plots were added to the right to identify goodness of fit of the model.

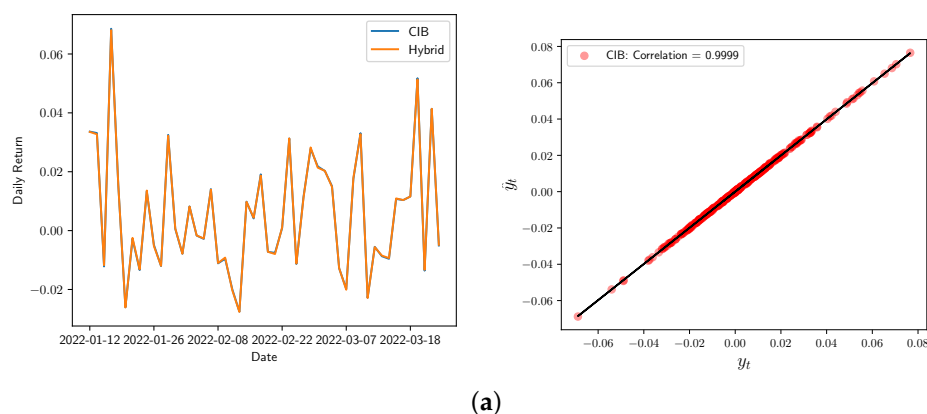


Figure 7. Cont.

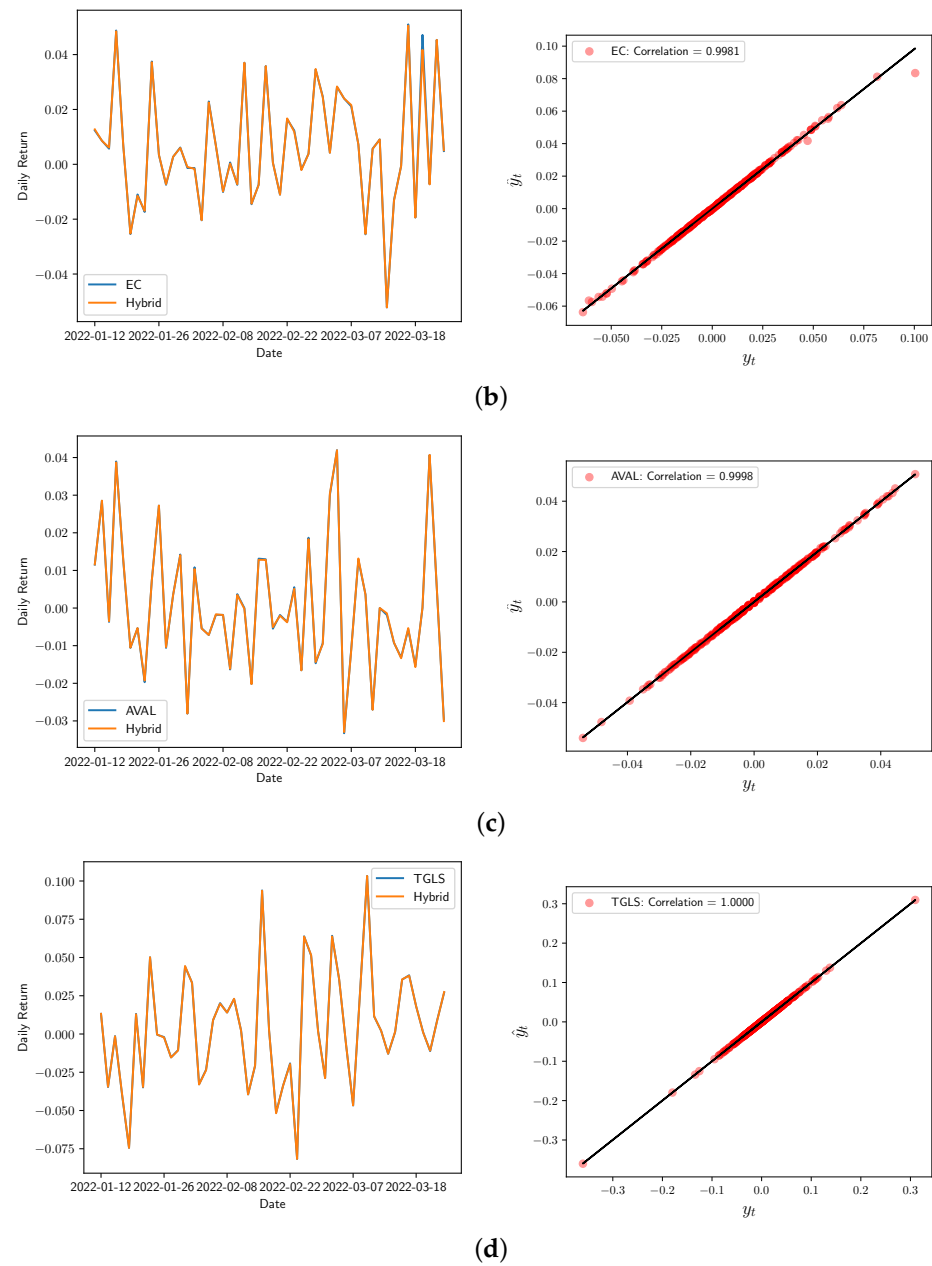


Figure 7. Real vs. Hybrid adjustment for the daily returns (left) and correlation $\text{Corr}(y_t, \hat{y}_t)$ for the real and forecasted share prices (right) of: (a) Bancolombia, (b) Ecopetrol, (c) Grupo AVAL, and (d) Tecnoglass.

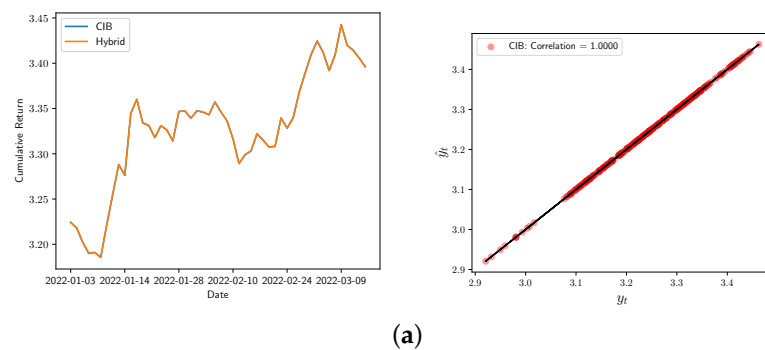


Figure 8. Cont.

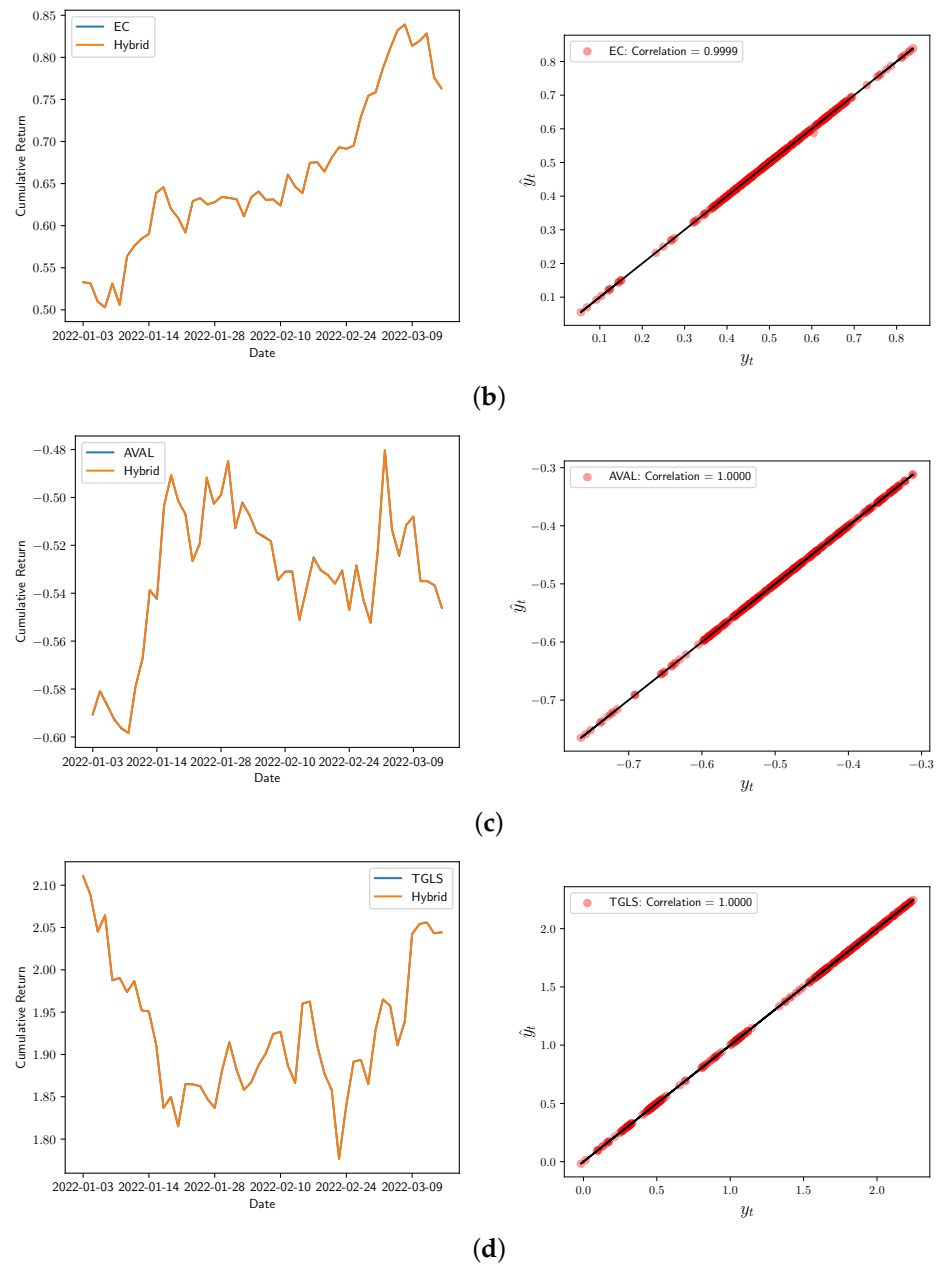


Figure 8. Real vs. Hybrid adjustment for the cumulative returns (left) and correlation $\text{Corr}(y_t, \hat{y}_t)$ for the real and forecasted share prices (right) of: (a) Bancolombia, (b) Ecopetrol, (c) Grupo AVAL, and (d) Tecnoglass.

Tables 7 and 8 summarize the forecasting results for each model used. These findings show that the hybrid model improves the component models (ARIMA and SVR) on four precision measures in most cases (except for Bancolombia). For Bancolombia's daily and cumulative returns, the hybrid model presents similar results to ARIMA in terms of errors, but always provides the best goodness of fit, confirmed with the correlation plots and the R^2 results which appear in the last columns of Tables 7 and 8.

Table 7. Accuracy metrics and R^2 for daily returns an all models. The figures in the MAE, MSE, and RMSE columns are to be multiplied by 10^{-7} .

Model	Share	MAE	MSE	RMSE	R^2
ARIMA	Bancolombia (CIB)	156,642.5839	4209.16114	205,162.4007	0.0630787
	Ecopetrol (EC)	164,542.6646	4563.8175	213,630.9311	0.0031868
	Grupo AVAL (AV)	118,944.0264	2435.5536	156,062.6022	0.0418544
	Tecnoglass (TGLS)	330,208.9796	26,142.6679	511,299.0114	0.1570711
SVR	Bancolombia (CIB)	1896.1414	0.5732	2394.1233	0.9998737
	Ecopetrol (EC)	2046.0967	0.6497	2549.0300	0.9998610
	Grupo AVAL (AVAL)	1849.0694	0.5449	2334.3474	0.9997991
	Tecnoglass (TGLS)	2054.6102	0.6239	2497.8044	0.9997999
Hybrid	Bancolombia (CIB)	2043.0542	0.6654	2579.6194	0.9998584
	Ecopetrol (EC)	1892.1912	0.5584	2363.1461	0.9998945
	Grupo AVAL (AVAL)	2022.1969	0.6135	2476.8812	0.9997654
	Tecnoglass (TGLS)	1909.8914	0.5519	2349.3156	0.9999832

Table 8. Accuracy metrics and R^2 for cumulative returns an all models. The figures in the MAE, MSE, and RMSE columns are to be multiplied by 10^{-7} .

Model	Share	MAE	MSE	RMSE	R^2
ARIMA	Bancolombia (CIB)	156,412.9033	4258.5618	206,362.8323	0.9392642
	Ecopetrol (EC)	156,730.7965	4104.2421	202,589.2924	0.9658714
	Grupo AVAL (AV)	123,512.9019	2546.7269	159,584.6773	0.7500343
	Tecnoglass (TGLS)	325,112.7213	24,722.3193	497,215.4389	0.8636632
SVR	Bancolombia (CIB)	2043.3500	0.6387	2527.3378	0.9999908
	Ecopetrol (EC)	2184.8263	0.7228	2688.4308	0.9999939
	Grupo AVAL (AV)	1968.4033	0.5853	2419.3327	0.9999378
	Tecnoglass (TGLS)	1916.5194	0.5896	2428.2147	0.9999967
Hybrid	Bancolombia (CIB)	2081.0660	0.6400	2529.7931	0.9999908
	Ecopetrol (EC)	1860.2882	0.5132	2265.3619	0.9999961
	Grupo AVAL (AVAL)	1849.0696	0.5449	2334.3475	0.9999426
	Tecnoglass (TGLS)	1695.8875	0.4675	2162.2613	0.9999974

4.3. Discussion

Results presented in this section correspond to three different predictions of daily and cumulative returns using the ARIMA, SVR, and ARIMA–SVR models. Return predictions are important for investors, as they allow for investors to plan goals based on tangible results and to understand whether or not it is worthwhile to invest in certain stocks.

Tables 7 and 8 summarize the forecasting results for each model used. These findings show that the hybrid model improves the component models (ARIMA and SVR). Using four precision measures, in most cases (except for Bancolombia's and Grupo Aval's daily returns), the hybrid model outperformed the predictions from the ARIMA and SVR models. For Bancolombia's cumulative returns, the hybrid model presented results similar to SVR in terms of errors and goodness of fit, confirmed with the correlation plot and R^2 results. The Deibold–Mariano test (see Table 9) was used to determine whether two predictions are significantly different, considering as null hypothesis, forecast errors from two models cause the same loss. As can be seen, Table 9 shows that ARIMA and the Hybrid model are significantly different, as well as the SVR and the Hybrid model to forecast Tecnoglass's cumulative returns.

Figures 9 and 10 show the best fit for each model; the error and accuracy measures are recorded in Tables 7 and 8 to confirm the findings. Tables 7 and 8 show that the hybrid model improved, in the majority of cases, the ARIMA and SVR models for each company, except for Bancolombia's cumulative returns. On the other hand, the correlation plot and R^2 calculation show there is a high proportion of the total variance explained by the hybrid model with R^2 closer to 1, compared to the other two models and all the companies.

As can be seen from the results obtained, the hybrid model is outstanding, as it provides the most accurate predictions with the best goodness of fit compared to those

obtained by the two component models individually. Prediction results for the daily and cumulative returns of the four most-important Colombian stocks listed on the NYSE confirm the significant advantages of using this type of model to capture different types of patterns that can be found in a time series. These types of models can be used for different stocks listed on different exchanges with similar behaviors to obtain the best accuracy and goodness of fit for predictions.

Table 9. The Diebold–Mariano test was used for each prediction pair $(DM, p\text{-value})_{SVR, Hybrid}$, $(DM, p\text{-value})_{ARIMA, Hybrid}$ corresponding to the daily or cumulative returns, respectively.

	Share	$(DM, p\text{-Value})_{ARIMA, Hybrid}$	$(DM, p\text{-Value})_{SVR, Hybrid}$
Daily	Bancolombia (CIB)	$(7.7018, 1.3424 \times 10^{-13})$	$(0.1316, 0.8954)$
	Ecopetrol (EC)	$(8.8210, 5.1582 \times 10^{-17})$	$(-1.2238, 0.2218)$
	Grupo AVAL (AVAL)	$(8.4931, 5.5444 \times 10^{-13})$	$(1.6671, 0.0964)$
	Tecnoglass (TGLS)	$(4.0985, 5.1533 \times 10^{-13})$	$(1.0358, 0.3010)$
Cumulative	Bancolombia (CIB)	$(7.1732, 4.4188 \times 10^{-12})$	$(-1.4533, 0.1470)$
	Ecopetrol (EC)	$(8.4510, 7.9365 \times 10^{-16})$	$(-1.0232, 0.3069)$
	Grupo AVAL (AVAL)	$(8.0887, 1.0073 \times 10^{-14})$	$(0.9748, 0.3304)$
	Tecnoglass (TGLS)	$(4.0987, 5.1714 \times 10^{-5})$	$(1.9944, 0.0469)$

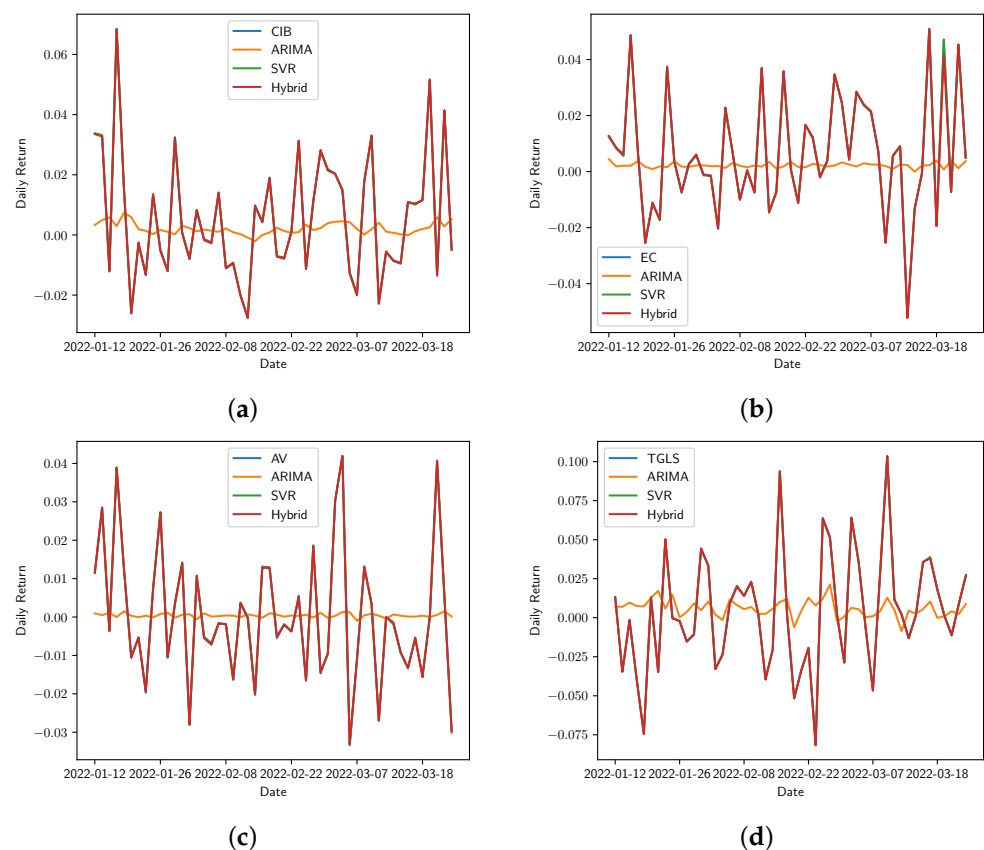


Figure 9. Real vs. the ARIMA, SVR, and Hybrid models' adjustment for daily returns. (a) Bancolombia, (b) Ecopetrol, (c) Grupo AVAL, and (d) Tecnoglass. The real-time series is represented by the stock symbol and its forecasts by the name of the corresponding model.

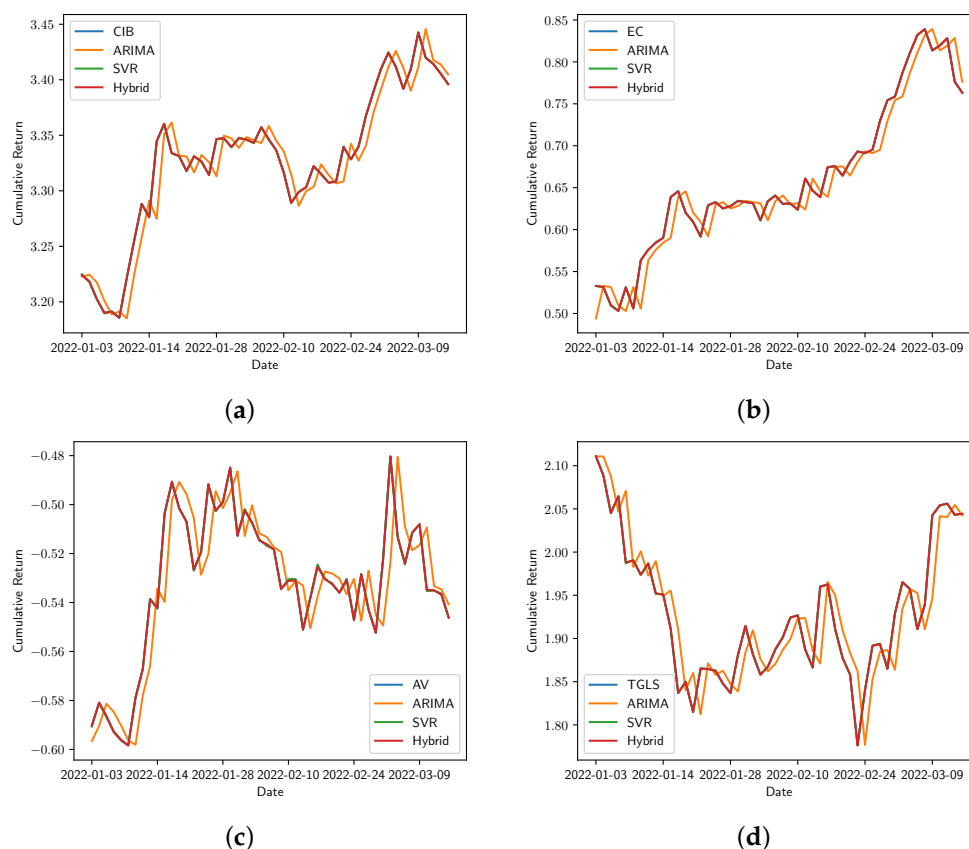


Figure 10. Real vs. the ARIMA, SVR, and Hybrid models' adjustment for cumulative returns. (a) Bancolombia, (b) Ecopetrol, (c) Grupo AVAL, and (d) Tecnoglass. The real-time series is represented by the stock symbol and its forecasts by the name of the corresponding model.

5. Conclusions

This work compares three forecasting models, ARIMA, SVR, and the hybrid ARIMA-SVR model, to forecast the daily and cumulative returns of shares in Bancolombia, Ecopetrol, Tecnoglass, and Grupo Aval. There are no works in the literature that use hybrid models such as the one presented in this paper to predict Colombian shares listed on the New York Stock Exchange, which is of great importance for aiding national and international investors interested in these companies in making correct decisions [51].

A few works can be found where forecasting models were used to predict Colombian shares; some of them use a self-organizing neural fuzzy system (SONFS), for example, to forecast Ecopetrol shares, or an adaptive-network-based fuzzy inference system (ANFIS) for the Colombian power market [52,53], but none of them use a hybrid technique using two different models, as is proposed in this work, to outperform forecasting results; a homogeneous model is always trained for its predictions. Different stocks can show behaviors similar to the set of time series data studied here, as well as other time series used for other forecasting purposes; therefore, the effectiveness of using hybrid techniques can be extended to other types of problems where prediction accuracy is a major concern. Moreover, based on the lack of development for this type of model for Colombian share predictions, this work uses a hybrid model to leverage the effectiveness of the ARIMA and SVR models to predict linear and non-linear patterns, and presents this type of novel model as a suitable alternative to improve accuracy for time series forecasting. The shares studied here have generated a great interest among those who invest in the stock market due to the solidity of the companies representing them.

Time series for these companies were not stationary; therefore, integration was applied until they became stationary. $ARIMA(p, d, q)$, SVR with the RBF kernel and ϵ , C , and γ cross-validation parameters, and the hybrid ARIMA-SVR model were selected to forecast

the daily and cumulative returns of each company, and the datasets were modeled by adjusting the parameters until the best results, in terms of errors and goodness of fit, were obtained. Accuracy errors show the hybridization of each model, namely, ARIMA and SVR, reduces forecast errors, giving greater confidence to investors to use such a predictive model for decision making.

Considering the findings in Tables 7 and 8, it can be inferred that the hybrid technique generated the smallest errors in most cases, except for Bancolombia, providing the best approximation to the real dataset. For Bancolombia, similar predictions between ARIMA and the hybrid model were found, but the hybrid model always had the best goodness-of-fit value. The hybrid ARIMA–SVR model provided several promising options to be used in forecasting analysis; in fact, there are several time series applications in the literature where the model used was a hybrid one [54,55].

The main goal of predicting the daily and cumulative returns of the CIB, EC, AVAL, and TGLS shares quoted on the NYSE and Nasdaq was achieved; however, a limitation was found in trying to forecast Bancolombia's and Grupo Aval's daily returns. As future work, different hybridization techniques to make the predictions more efficient and outperform forecasting results, as well as the use of deep learning models such as LSTM to estimate non-linearity in the time series data, can be studied. Hybrid GARCH models also can be used to forecast the volatility of stocks with strong fluctuations, such as future contracts in the cryptocurrencies market. Another pre-filter based on Fourier transformation, for example, also can be applied to improve results from one of each component of the hybrid model.

Author Contributions: All the authors (L.R. and K.A.) have participated impartially in all the aspects of this work: conceptualization, methodology, investigation, formal analysis, software, validation, visualization, writing—original draft preparation, writing—review and editing. All authors have read and agreed to the published version of the manuscript.

Funding: L.R. is supported by Universidad del Norte. Research Agenda 2021-009.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Cuartas, F.D.J.F. Estructura y evolución del sistema financiero colombiano de la banca comercial a la banca de inversión. *Modum Rev. Divulg. Multidiscip. Cienc. Tecnol. Innov.* **2017**, *1*, 13–34.
2. Cáceres-Chian, V.A.E. *Predicción de Precios de Acciones de Bolsa de Valores Utilizando Support Vector Regression*; Universidad de Lima: Lima, Peru, 2018. [CrossRef]
3. Varón Suárez, O.M.; Cuadros Parra, J.S. Análisis de las Acciones Emitidas por Grupo Bancolombia en la Bolsa de Valores De Colombia, de Cara a una Crisis Económica y Sanitaria. Available online: <http://hdl.handle.net/20.500.12495/5445> (accessed on 11 April 2022).
4. Jung, G.; Choi, S.Y. Forecasting Foreign Exchange Volatility Using Deep Learning Autoencoder-LSTM Techniques. *Complexity* **2021**, *2021*, 6647534. [CrossRef]
5. Tripathi, M.; Kumar, S.; Inani, S.K. Exchange Rate Forecasting Using Ensemble Modeling for Better Policy Implications. *J. Time Ser. Econom.* **2021**, *13*, 43–71. [CrossRef]
6. Huang, W.; Nakamori, Y.; Wang, S.Y. Forecasting stock market movement direction with support vector machine. *Comput. Oper. Res.* **2005**, *32*, 2513–2522. [CrossRef]
7. Cao, L. Support vector machines experts for time series forecasting. *Neurocomputing* **2003**, *51*, 321–339. [CrossRef]
8. Cao, L.; Tay, F.E. Financial forecasting using support vector machines. *Neural Comput. Appl.* **2001**, *10*, 184–192. [CrossRef]
9. Kim, K.j. Financial time series forecasting using support vector machines. *Neurocomputing* **2003**, *55*, 307–319. [CrossRef]
10. Mohandes, M.A.; Halawani, T.O.; Rehman, S.; Hussain, A.A. Support vector machines for wind speed prediction. *Renew. Energy* **2004**, *29*, 939–947. [CrossRef]
11. Pai, P.F.; Lin, C.S. Using support vector machines to forecast the production values of the machinery industry in Taiwan. *Int. J. Adv. Manuf. Technol.* **2005**, *27*, 205–210. [CrossRef]

12. Box, G.; Jenkins, G.; Reinsel, G.; Ljung, G. *Time Series Analysis: Forecasting and Control*; Wiley Series in Probability and Statistics; Wiley: Hoboken, NJ, USA, 2015.
13. Faruk, D.Ö. A hybrid neural network and ARIMA model for water quality time series prediction. *Eng. Appl. Artif. Intell.* **2010**, *23*, 586–594. [\[CrossRef\]](#)
14. Calheiros, R.N.; Masoumi, E.; Ranjan, R.; Buyya, R. Workload prediction using ARIMA model and its impact on cloud applications' QoS. *IEEE Trans. Cloud Comput.* **2014**, *3*, 449–458. [\[CrossRef\]](#)
15. Rubio, L.; Gutiérrez-Rodríguez, A.J.; Forero, M.G. EBITDA Index Prediction Using Exponential Smoothing and ARIMA Model. *Mathematics* **2021**, *9*, 2538. [\[CrossRef\]](#)
16. Cho, M.; Hwang, J.; Chen, C. Customer short term load forecasting by using ARIMA transfer function model. In Proceedings of the 1995 International Conference on Energy Management and Power Delivery EMPD'95, Singapore, 21–23 November 1995; Volume 1, pp. 317–322.
17. Zhang, G.; Patuwo, B.E.; Hu, M.Y. Forecasting with artificial neural networks:: The state of the art. *Int. J. Forecast.* **1998**, *14*, 35–62. [\[CrossRef\]](#)
18. Al-Alawi, S.M.; Abdul-Wahab, S.A.; Bakheit, C.S. Combining principal component regression and artificial neural networks for more accurate predictions of ground-level ozone. *Environ. Model. Softw.* **2008**, *23*, 396–403. [\[CrossRef\]](#)
19. Sharkey, A.J. Types of multinet system. In Proceedings of the International Workshop on Multiple Classifier Systems, Cagliari, Italy, 24–26 June 2002; pp. 108–117.
20. Taskaya-Temizel, T.; Casey, M.C. A comparative study of autoregressive neural network hybrids. *Neural Netw.* **2005**, *18*, 781–789. [\[CrossRef\]](#)
21. Zhang, G.P. Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing* **2003**, *50*, 159–175. [\[CrossRef\]](#)
22. Tseng, F.M.; Yu, H.C.; Tzeng, G.H. Combining neural network model with seasonal time series ARIMA model. *Technol. Forecast. Soc. Chang.* **2002**, *69*, 71–87. [\[CrossRef\]](#)
23. Xiong, L.; Lu, Y. Hybrid ARIMA-BPNN model for time series prediction of the Chinese stock market. In Proceedings of the 2017 3rd International Conference on Information Management (ICIM), Paris, France, 12–14 September 2017; pp. 93–97.
24. Fathi, O. Time Series Forecasting Using a hybrid ARIMA and LSTM Model. *Velv. Consult.* **2019**, 1–7. Available online: https://www.velvetconsulting.com/wp-content/uploads/2019/03/Seasonality_modeling_using_ARIMA_LSTM_Hybrid_Model.pdf (accessed on 11 April 2022).
25. Wang, J.J.; Wang, J.Z.; Zhang, Z.G.; Guo, S.P. Stock index forecasting based on a hybrid model. *Omega* **2012**, *40*, 758–766. [\[CrossRef\]](#)
26. Islam, M.S.; Hossain, E. Foreign exchange currency rate prediction using a GRU-LSTM Hybrid Network. *Soft Comput. Lett.* **2020**, *3*, 100009. [\[CrossRef\]](#)
27. Musa, Y.; Joshua, S. Analysis of ARIMA-artificial neural network hybrid model in forecasting of stock market returns. *Asian J. Probab. Stat.* **2020**, *6*, 42–53. [\[CrossRef\]](#)
28. Khashei, M.; Bijari, M.; Ardali, G.A.R. Improvement of auto-regressive integrated moving average models using fuzzy logic and artificial neural networks (ANNs). *Neurocomputing* **2009**, *72*, 956–967. [\[CrossRef\]](#)
29. Van Der Voort, M.; Dougherty, M.; Watson, S. Combining Kohonen maps with ARIMA time series models to forecast traffic flow. *Transp. Res. Part Emerg. Technol.* **1996**, *4*, 307–318. [\[CrossRef\]](#)
30. Khashei, M.; Hejazi, S.R.; Bijari, M. A new hybrid artificial neural networks and fuzzy regression model for time series forecasting. *Fuzzy Sets Syst.* **2008**, *159*, 769–786. [\[CrossRef\]](#)
31. Palacio Álvarez, D.A. Viabilidad de Inversión en Acciones Preferenciales del Grupo Bancolombia Como Emisor de la Bolsa de Valores de Colombia. Available online: <http://hdl.handle.net/20.500.12495/5470> (accessed on 11 April 2022).
32. Top 100 Producers and Their Cumulative Greenhouse Gas Emissions from 1988–2015. *The Guardian* 2020. Available online: <https://www.theguardian.com/sustainable-business/2017/jul/10/100-fossil-fuel-companies-investors-responsible-71-global-emissions-cdp-study-climate-change> (accessed on 11 April 2022).
33. Ecopetrol, Entre las Mejores del Mundo. *Dinero* 2020. Available online: <https://www.semana.com/empresas/articulo/ecopetrol-entre-mejores-del-mundo/162635/> (accessed on 11 April 2022).
34. Acción de Tecnoglass Toca un Nuevo Precio récord en Nasdaq. *El Herald* 2021. Available online: <https://www.elheraldo.co/economia/accion-de-tecnoglass-toca-un-nuevo-precio-record-en-nasdaq-855991> (accessed on 11 April 2022).
35. Tecnoglass, Epicentro de la Industria del Vidrio en las Américas. *El Herald* 2021. Available online: <https://www.elheraldo.co/economia/tecnoglass-epicentro-de-la-industria-del-vidrio-en-las-americas-801438> (accessed on 11 April 2022).
36. Estreno de Aval en Bolsa de Nueva York Superó Expectativas. *Portafolio* 2014. Available online: <https://www.portafolio.co/negocios/empresas/estreno-aval-bolsa-nueva-york-supero-expectativas-64406> (accessed on 11 April 2022).
37. Grupo Aval Compró Multibank Financial Group por Cerca de US\$728 Millones. *La Republica* 2019. Available online: <https://www.larepublica.co/finanzas/grupo-aval-compro-multibank-financial-group-por-cerca-de-us728-millones-2927667> (accessed on 11 April 2022).
38. Asteriou, D.; Hall, S.G. ARIMA models and the Box–Jenkins methodology. *Appl. Econom.* **2011**, *2*, 265–286.
39. Montgomery, D.C.; Jennings, C.L.; Kulahci, M. *Introduction to Time Series Analysis and Forecasting*; John Wiley & Sons: Hoboken, NJ, USA, 2015.

40. Liu, J.; Seraoui, R.; Vitelli, V.; Zio, E. Nuclear power plant components condition monitoring by probabilistic support vector machine. *Ann. Nucl. Energy* **2013**, *56*, 23–33. [[CrossRef](#)]
41. Yu, P.S.; Chen, S.T.; Chang, I.F. Support vector regression for real-time flood stage forecasting. *J. Hydrol.* **2006**, *328*, 704–716. [[CrossRef](#)]
42. Dibike, Y.B.; Velickov, S.; Solomatine, D.; Abbott, M.B. Model induction with support vector machines: introduction and applications. *J. Comput. Civ. Eng.* **2001**, *15*, 208–216. [[CrossRef](#)]
43. Wu, C.; Chau, K.W.; Li, Y.S. River stage prediction based on a distributed support vector regression. *J. Hydrol.* **2008**, *358*, 96–111. [[CrossRef](#)]
44. Wu, M.C.; Lin, G.F.; Lin, H.Y. Improving the forecasts of extreme streamflow by support vector regression with the data extracted by self-organizing map. *Hydrol. Process.* **2014**, *28*, 386–397. [[CrossRef](#)]
45. Chang, C.C.; Lin, C.J. LIBSVM: A library for support vector machines. *Acm Trans. Intell. Syst. Technol.* **2011**, *2*, 1–27. [[CrossRef](#)]
46. Suganyadevi, M.; Babulal, C. Support vector regression model for the prediction of loadability margin of a power system. *Appl. Soft Comput.* **2014**, *24*, 304–315. [[CrossRef](#)]
47. Khashei, M.; Bijari, M. A novel hybridization of artificial neural networks and ARIMA models for time series forecasting. *Appl. Soft Comput.* **2011**, *11*, 2664–2675. [[CrossRef](#)]
48. Diebold, F.X.; Mariano, R.S. Comparing predictive accuracy. *J. Bus. Econ. Stat.* **2002**, *20*, 134–144. [[CrossRef](#)]
49. Fildes, R.; Makridakis, S. The impact of empirical accuracy studies on time series analysis and forecasting. *Int. Stat. Rev. Int. Stat.* **1995**, *63*, 289–308. [[CrossRef](#)]
50. Shibata, R. Selection of the order of an autoregressive model by Akaike's information criterion. *Biometrika* **1976**, *63*, 117–126. [[CrossRef](#)]
51. Los Inversionistas Extranjeros Fueron los que Más Transaron Acciones Durante la OPA. *La Republica* 2022. Available online: <https://www.larepublica.co/especiales/opa-por-nutresa/los-inversionistas-extranjeros-fueron-los-que-mas-transaron-acciones-durante-la-opa-3300083> (accessed on 11 April 2022).
52. Gonzalez, J.A.A.; Rey, C.M.O.; Garcia, J.C.F. A Self-Organizing Neural Fuzzy System to forecast the price of Ecopetrol shares. In Proceedings of the 2012 IEEE Conference on Computational Intelligence for Financial Engineering & Economics (CIFER), New York, NY, USA, 29–30 March 2012; pp. 1–6.
53. Moreno, J. Hydraulic plant generation forecasting in Colombian power market using ANFIS. *Energy Econ.* **2009**, *31*, 450–455. [[CrossRef](#)]
54. Lin, H.; Sun, Q.; Chen, S.Q. Reducing Exchange Rate Risks in International Trade: A Hybrid Forecasting Approach of CEEMDAN and Multilayer LSTM. *Sustainability* **2020**, *12*, 2451. [[CrossRef](#)]
55. Sreeram, L.; Sayed, S.A. Short-term Forecasting Ability of Hybrid Models for BRIC Currencies. *Glob. Bus. Rev.* **2020**, 0972150920954615. [[CrossRef](#)]