Financial interval time series forecasting and forecast combination: Empirical evidence for US and Brazilian stock markets

Abstract

In finance, interval time series (ITS) represent the evolution of lows and highs prices of an asset throughout time. These price ranges are related to the concept of volatility as they are able to capture the intraday price variability. Hence, their accurate forecasts play an essential role in derivative pricing, trading strategies, risk management and portfolio allocation. This paper evaluates the forecasting performance of intervalar models and their forecasting combinations for financial ITS. Interval-valued forecasting approaches are able to take into account the intrinsic relationship between interval bounds by processing daily lows and highs simultaneously. One-step-ahead individual predictions of the interval Holt’s exponential smoothing (Holt$^1$), the interval multilayer perceptron neural network (iMLP), and the hybrid Holt’s smoothing multi-output support vector regression (Holt$^1$-MSRV) methods, as well as their combinations using simple average, simple average with trimming, variance-based averaging, and ordinary least squares averaging, are evaluated considering as empirical application the S&P 500 and IBOVESPA stock indices. Using traditional accuracy metrics, quality measures designed for ITS, and a simple trading strategy as economic criteria, the results suggest that the combination of intervalar methods gives more accurate predictions than their individual forecasts.

Keywords: Interval time series, forecast combination, interval-valued data, stock market, finance.

1. Introduction

Accurate prediction of asset prices plays a central role in risk management, portfolio selection and derivative pricing (Pettenuzzo et al. 2014). The temporal evolution of equities, stock indices, interest and exchange rates are observed as single-valued financial time series (Arroyo et al. 2011). This measurement is useful in many practical cases, but in situations where several values are observed at each time period (day, hour, minute) it may be neglecting relevant price fluctuations. For instance, if only the opening (or closing) asset price is measured daily, the resulting time series will hide the intraday variability and important information is missed (Degiannakis and Floros 2013; Haniff and Pok 2010).
Besides intraday time series could be forecasted, they reveal characteristics such as irregular temporal spacing, strong diurnal patterns and complex dependence, which result in obstacles for traditional time series models (Webb et al., 2016; Molgedey and Ebeling, 2000). Further, the accurate prediction of the whole sequence of intraday prices for one day ahead is almost impossible in dynamic practical situations. An alternative to alleviate these limitations is when the highest and the lowest values of prices are measured at each time period, what originates interval time series (ITS) (Engle and Russell, 2009).

The concept of ITS, considered in the field of Symbolic Data Analysis (Bock and Diday, 2000), is related to situations in which the data show some variability, as observed in asset prices. Variables of similar nature include, e.g. electricity prices, power load and generation, meteorology, production rates, traffic flows (Weron, 2014). In particular, considering the highs and lows prices of assets, financial ITS modeling and forecasting have received considerable attention in the recent literature with the introduction of several interval time series forecasting methods (Lu et al., 2015; Froelich and Salmeron, 2014; Xiong et al., 2015a; Wang et al., 2013).

Analysis of large data sets, such as high frequency data, based on the ITS framework is a new domain to study statistically detectable patterns. It has attracted a large number of researchers in economics and finance (Rodrigues and Salish, 2015). The daily high and low financial prices can be seen as references values for investors in order to place buy or sell orders, e.g. through candlestick charts, a popular technical indicator (Cheung and Chinn, 2001; Xiong et al., 2017). He and Wan (2009) also stated that the highs and lows are related to prices at which the excess of demand changes its direction. It is worthy to mention that financial ITS take into account variability and/or uncertainty, reducing the amount of random variation relative to that found in classic single-valued data of an asset (e.g. closing prices) (Xiong et al., 2017; Lima and de Carvalho, 2010).

Additionally, high and low prices are related with the concept of volatility. Alizadeh et al. (2002) show that the difference between the highest and lowest log prices over a fixed sample interval, also known as the log range, is a highly efficient volatility measure. Brandt and Diebold (2006) and Shu and Zhang (2006) also stated that the range-based volatility estimator appears robust to microstructure noise such as bid-ask bounce, which overcomes the limitations of traditional volatility.

---

1The literature that considers the high-low range prices as a proxy for volatility dates back to the 1980s with the work of Parkinson (1980).
models based on closing prices that fall to use the information contents inside the reference period of the prices, resulting in inaccurate forecasts (Chou 2005).

Recently, many empirical studies related to financial interval time series have been conducted in the literature, advocating the superior performance of intervalar methods against traditional univariate and multivariate statistical methodologies, since these methods are not able to model the intervalar structure of the data by processing the interval bounds individually (Lima and Carvalho 2008; Hu and He 2007; He and Hu 2009; Arroyo et al. 2011; Xiong et al. 2015b). Besides indicating the high potential in practical applications when assets prices are viewed as intervals and by using intervalar methods, these studies fail to provide consistent results as to which actual approach performs best. In a wide variety of forecasting scenarios, it is virtually impossible to visualize consistent results that indicate which method performs best. This is justified by the well-known no free-lunch theorem (Wolpert and Macready 1997): if an algorithm performs well on a certain class of problems then it necessarily pays for that with degraded performance on the set of all remaining problems. No algorithm performs best on all possible data sets (Lemke and Gabrys 2010). In practice an algorithm may perform best for a specific class of problems and data.

The idea to combine forecasting algorithms appeared as a way to overcome the no free lunch limitation. Actually, combination of forecasts dates the late 1960s (Bates and Granger 1969; Crane and Crotty 1967) when it was shown that significant reduction of forecast errors may be achieved by the weighted combination of forecasts produced by distinct algorithms and methods. Few years later, Bunn (1975) suggested a Bayesian approach to combine forecasts. The author shows how subjective probabilities can be meaningfully assigned over a set of forecasting models and updated when the forecast realizations become known. Since then many studies have indicated the superior performance of forecast combinations over individual forecasts, as e.g. Stock and Watson (2004) and Timmermann (2006).

Hence, this paper suggests an empirical study investigating the forecasting and forecasting combination performances of intervalar approaches for financial interval time series. The interval Holt’s exponential smoothing (Holt\textsuperscript{1}) (Maia and de Carvalho 2011), the interval multilayer perceptron neural network (iMLP) (Roque et al. 2007), and the hybrid Holt’s smoothing multi-output support vector regression (Holt\textsuperscript{1}-MSRV) (Xiong et al. 2017) are considered in this work. These methods are able to process interval-valued data naturally. Holt\textsuperscript{1} and iMPL are linear and nonlinear intervalar approaches, while Holt\textsuperscript{1}-MSRV is a hybrid method in the “linear and nonlinear” modeling framework designed to address both linear and nonlinear patterns hidden in the data, as observed
in financial time series. Experiments are conducted using the main stock indices of the US and Brazilian financial markets, the S&P 500 and IBOVESPA, respectively, for the period from January 2000 to December 2015, focusing on one-step-ahead forecasts. Forecast combination is done using different methods: simple average, simple average with trimming, variance-based averaging and ordinary least squares averaging. Comparisons are conducted using traditional accuracy measures and statistical tests, as well as in terms of error metrics designed for interval-valued data, including additional forecast descriptive statistics such as efficiency and coverage rates. Further, individual and combined forecasts are used as input on a trading strategy to measure the results as an economic criteria.

The contributions of this work can be summarized as follows. First, there is no evidence in the current literature regarding the evaluation of forecast combination in the field of interval time series forecasting, specially for financial ITS forecasting. The second contribution is that not only statistical accuracy but also quality measures designed for interval time series are used to assess the predictability of the individual models and their combinations for interval-valued stock index forecasting. Third, the use of an economic criteria accesses the potential of the results in real-world applications. Finally, the literature still demands works related to ITS forecasting in finance, mostly for emergent economies like Brazil, since the growing availability of high frequency data and its operations (high frequency trading) require more informative predictions, which are not only very important, but also compelling.

This paper is outlined as follows. After this introduction, Section 2 provides a literature review regarding financial interval time series forecasting and forecast combination. Section 3 gives the methodology, including a brief reminder of the interval arithmetics adopted in this work, the intervalar methods used for prediction, the approaches considered for forecast combination and the measures to access models performance. The empirical application on the S&P 500 and IBOVESPA indices is discussed in Section 4. Finally, Section 5 concludes the paper and lists topics for future research.

2. Literature review

The literature addressing ITS in finance generally uses extensions of classic data analytics and forecasting techniques (Fiess and MacDonald, 2002; Lima and Carvalho, 2008). For instance, Fiess and MacDonald (2002) use a vector error correction model (VECM) to forecast low, high and closing values of three foreign exchange rates: Dollar-Yen, Pound-Dollar, and Mark-Dollar. According to
the authors, the data generating processes of the prices time series are different, and the prediction of low and high values improves closing price forecasting. Similarly, Cheung (2007) indicates that high and low values of the Dow Jones Industrial, S&P 500 and NASDAQ indices increase the explanatory capacity of a VECM model.

An interval linear model to forecast the stock market price movements is addressed in Hu and He (2007). The intervals are constructed based on lower and higher equity prices, and the model is calculated from their midpoints (centers) using ordinary last squares. Further, He and Hu (2008) considered the same econometric model but estimated separately for the lower bounds and the upper bounds of the interval time series. In He and Hu (2009), both approaches are compared and the authors conclude that the former yields more accurate results. García-Ascanio and Mate (2010) proposed an empirical model for monthly highs and lows of electricity demand using vector autoregression (VAR).

Arroyo et al. (2011) use ITS to forecast the Dow Jones Industrial index and the Euro-Dollar exchange rate using univariate and multivariate techniques such as exponential smoothing, autoregressive integrated moving average (ARIMA), artificial neural networks, and pattern recognition methods such as the k-nearest neighbors. Forecasts are produced using independent (univariate) and joint (multivariate) predictions of the lower and upper bounds of intervals or, alternatively, their midpoints (center) and half-lengths (radius). Approaches that independently forecasts the lower and upper bounds time series obtained the worst results.

Rodrigues and Salish (2015) suggest univariate threshold models to predict S&P 500 index interval time series by independently forecasting the midpoints and half-lengths of the intervals. Comparisons are made against random walk, vector autoregressive (VAR), and k-nearest neighbor techniques. Results show that the threshold method provides more accurate forecasts than the remaining approaches.

Using intraday data, Yang et al. (2014) compare ARIMA, naive, and interval linear regression models to forecast S&P 500, Dow Jones Industrial and Nasdaq indices. The authors claim that interval-based approaches are superior when compared against single-valued time series models. Recently, Xiong et al. (2014) developed a support vector machine model to simultaneously forecast the minimum and maximum values of the S&P 500, FTSE 100, and Nikkei 225 interval indices. The results, when compared with classic econometric techniques and interval neural networks, show the superior potential of the interval-based model, especially in financial trading.

The current literature advocates the use of ITS framework in economics and finance, since it
provides appropriate mechanisms to analyze large data sets such as e.g. high frequency data, and also supplements the information extracted by the time series of the closing values considering high and low prices in terms of a proxy measure for volatility. However, most of the current approaches, even the ones based on interval-valued data, assumes a linear structure in representing time series process dynamics. Several studies indicate that nonlinear models do provide a richer understanding about the dynamics of variables of interest (Rodrigues and Salish 2015). Henry et al. (2001), Dueker et al. (2007) and Guidolin et al. (2009) are examples that suggest evidences of threshold nonlinearities in exchange rates, bond and stock markets, respectively. Roque et al. (2007), Maia and de Carvalho (2011) and Rodrigues and Salish (2015) also highlight the importance of nonlinearities in the context of ITS. In particular, Rodrigues and Salish (2015) provide empirical results stressing the high capability of forecasting models based on regime dependent ITS forecasts.

Several ITS forecasting methods has been developed in the literature of finance, including traditional statistical techniques as interval exponential smoothing methods such the interval Holt’s exponential smoothing, suggested by (Arroyo et al., 2011). However, these techniques provide good forecasts in cases where ITS under study are linear and stationary.

Due to the intrinsic complexity and volatility of ITS (e.g., interval-valued stock prices), they appear nonlinear and non-stationary. To overcome this limitation, machine learning techniques such as the interval multi-layer perceptrons (iMLP) (Roque et al., 2007) and the multi-output support vector regression (Xiong et al., 2014) have shown a relevant nonlinear modeling capability for ITS in real-world.

In this context, Xiong et al. (2015b) suggest a “linear and nonlinear” modeling framework based on VECM and multi-output support vector regression to forecast agricultural commodity future interval prices. Considering the Chinese future market, they advocate the high accuracy of the suggested method for interval-valued agricultural commodity futures prices forecasting due to its capability of capturing the linear and nonlinear patterns exhibited in future prices.

More recently, Xiong et al. (2017) suggested an hybrid modeling framework combining interval Holt’s exponential smoothing method (Holt) and multi-output support vector regression (MSVR) for ITS forecasting, named (Holt^I-MSRV). Holt and MSVR are committed to capture the linear and nonlinear patterns hidden in ITS, respectively. The results indicated that the proposed Holt^I-MSRV provide better ITS forecasting than individual linear and nonlinear methods.

Arroyo et al. (2011) provide a survey on ITS forecasting methodologies in finance and economics.
The current literature on financial ITS forecasting indicates that more accurate results are achieved when intervalar approaches are considered, since they are able to address the possible interrelations (e.g., cointegration between the daily highs and lows of the stock prices) that are presented amongst the highs and lows series. However, they fail to provide consistent results as to which actual approach performs best. An alternative to overcome this limitation is the use of forecast combination.

A common combination approach is to evaluate several potential forecasting methods on an in-sample dataset and to select the best among them to produce out-of-sample forecasts. The overwhelming interest in time series forecasts combination has led to the development of a large number of techniques. A majority of them use a weighted linear combination of individual forecasts. Statistical averaging such as simple average, trimmed mean, median, etc. are amongst the most basic combination techniques. For instance, Wang and Chang (2010) develop a demand forecasting methodology that combines market and shipment forecasts. They investigated how to assign weights to individual forecasts using three linear schemes (minimum value of the forecast error, adaptive weights, and regression analysis), and two nonlinear methods (fuzzy neural network, and an adaptive network based fuzzy inference system).

The use of meta-learning methods to improve forecasting performance in combining techniques is suggested by Lemke and Gabrys (2010). Meta-learning aims at identifying a set of features to describe the time series and a pool of individual forecasting methods. The results show the superiority of a ranking-based combination over simple model selection approaches.

The combination of exponential smoothing point and interval forecasts using weights derived from the Akaike information criteria on several time series datasets is examined in Kolassa (2011). The author suggests that simple and weighted combinations do not consistently outperform one another, and simple combinations sometimes perform worse than the single forecasts selected by the information criteria.

In the context of electricity prices, Nowotarski et al. (2014) evaluates the use of forecast averaging to perform a backtesting analysis on day-ahead electricity prices in three major European and US markets. The findings support the additional benefit of combining forecasts of the individual methods to derive more accurate forecasts.

Similarly, Zhao et al. (2014) develop a time-varying-weight combination method based on high-order Markov chains and a time-varying weighted average (HM-TWA) method to forecast monthly electricity consumption in China. The out-of-sample performance evaluation showed that HM-
TWA outperforms the individual models and traditional combining methods. The effectiveness of HM-TWA is further verified through comparisons with existing combination models.

In the field of computational intelligence, Adhikari (2015) suggests linear combination of forecasting using weights determined by a neural network. The neural network successively recognizes weighting patterns of the individual models using their past forecasting records to produce the forecasts. Using eight real-world time series data, the author shows that the approach achieves significantly better forecasting accuracy than each of the individual models and other linear combination schemes. More recently, Sadaei et al. (2016) introduced a hybrid method to combine autoregressive fractional integrated moving average (ARFIMA) with fuzzy time series models to forecast long memory (long-range) time series. The results indicated the superiority of the hybrid method when compared against the ARFIMA and other classic methods.

Summarizing, the literature advocates the higher accuracy of forecast combinations against individual forecasts. Thus, this paper aims to evaluate forecast combination approaches for financial ITS using linear, nonlinear and hybrid intervalar methods.

3. Methodology

This section details the methodology adopted in this paper. The formulation of interval time series and the basic concepts regarding intervalar arithmetic is first described. After, the intervalar methods used to compute the individual forecasts are addressed, as well as the forecast combination approaches. Finally, the performance measures are shown, including traditional accuracy metrics and the quality measures designed for interval-valued data. The economic criteria by means of a simple trading strategy is also detailed.

3.1. Interval-valued time series and interval arithmetics

An interval-valued variable $X$ is defined as a closed bounded set of real numbers in the form:

$$X = [X^L, X^U] \in \mathcal{I},$$

where $\mathcal{I} = \{[X^L, X^U] : X^L, X^U \in \mathbb{R}, X^L \leq X^U\}$ is the set of closed intervals of the real line $\mathbb{R}$, $X^L$ the lower bound, and $X^U$ the upper bound of the interval. Notice that, in this paper, $X^L$ and $X^U$ are the daily low and high prices of an asset, respectively.

The midpoint of an interval $X$, denoted by $x$, is:

$$x = \frac{X^L + X^U}{2}.$$
An interval time series (ITS), \( \{X_t\} \), is a sequence of interval-valued variables observed in successive time steps \( t = 1, 2, \ldots, N \) and expressed as a two-dimensional vector \( [X^L_t, X^U_t]^T \), i.e., \( X_t = [X^L_t, X^U_t]^T \).

Interval arithmetic extends traditional arithmetic to operate on intervals. This paper uses the arithmetic operations introduced by Moore et al. (2009):

\[
X + Y = [X^L + Y^L, X^U + Y^U],
\]
\[
X - Y = [X^L - Y^U, X^U - Y^L],
\]
\[
XY = \left[ \min\{X^L Y^L, X^L Y^U, X^U Y^L, X^U Y^U\}, \max\{X^L Y^L, X^L Y^U, X^U Y^L, X^U Y^U\} \right],
\]
\[
X/Y = X \left( \frac{1}{Y} \right), \text{ with } 1/Y = [1/Y^U, 1/Y^L].
\] (3)

Interval arithmetic subsumes classic arithmetic. This means that if an operation of interval arithmetic takes real numbers as operands, considering them as intervals of length zero, then we obtain the same result as if the operation were performed using traditional arithmetic.

### 3.2. Forecasting methods

This paper considers three intervalar forecasting methods: the interval Holt’s exponential smoothing (Holt\(^I\)) (Maia and de Carvalho, 2011), the interval multilayer perceptron neural network (iMLP) (Roque et al., 2007), and the hybrid Holt’s smoothing multi-output support vector regression (Holt\(^I\)-MSRV) (Xiong et al., 2017).

The Holt\(^I\) model can be described as the following formulation:

\[
\hat{P}^X_t = AX_t + (I - A) \left( \hat{P}^X_{t-1} + \hat{Q}^X_{t-1} \right),
\] (4)
\[
\hat{Q}^X_t = B \left( \hat{P}^X_t - \hat{P}^X_{t-1} \right) + (I - B) \hat{Q}^X_{t-1},
\] (5)

where \( X_t = [X^L_t, X^U_t]^T \), \( \hat{P}^X_t = [\hat{P}^L_t, \hat{P}^U_t]^T \), \( \hat{Q}^X_t = [\hat{Q}^L_t, \hat{Q}^U_t]^T \), \( I \) is an identity matrix of order 2, \( \hat{X}_{t+1} = \hat{P}^X_t + \hat{Q}^X_t \), and

\[
A = \begin{bmatrix}
\alpha_{11} & \alpha_{12} \\
\alpha_{21} & \alpha_{22}
\end{bmatrix},
\] (6)
\[
B = \begin{bmatrix}
\beta_{11} & \beta_{12} \\
\beta_{21} & \beta_{22}
\end{bmatrix},
\] (7)

are the smoothing parameters matrices constrained to the range \([0, 1]\). These parameters are estimated by minimizing the interval sum of squared one-step-ahead forecast errors:

\[
\min_{\alpha_{ij}, \beta_{ij}} R(A, B), \text{ subject to } 0 \leq \alpha_{ij}, \beta_{ij} \leq 1,
\] (8)
where
\[
\mathcal{R}(A, B) = \sum_{t=3}^{N} (X_t - \hat{X}_t)^T (X_t - \hat{X}_t)
\]
\[
= \sum_{t=3}^{N} (X_t^L - \hat{P}_{t-1}^L - \hat{Q}_{t-1}^L) + \sum_{t=3}^{N} (X_t^U - \hat{P}_{t-1}^U - \hat{Q}_{t-1}^U),
\]
\[(9)\]
and \(N\) is the sample size.

The start vectors for \(\hat{P}_X\) and \(\hat{Q}_X\) are taken to be \(X_2\) and \(X_2 - X_1\), respectively. The solution of this problem can be obtained using the limited memory BFGS method for bound constrained optimization (Byrd et al., 1995). For more details regarding the Holt\(^1\) method refer to Maia and de Carvalho (2011). Notice that Holt\(^1\) concerns a linear forecasting approach for interval-valued data.

The interval multilayer perceptron neural network (iMLP), proposed by Roque et al. (2007), is a nonlinear forecasting method designed to process interval-valued input and output data. As an extension of the classic multi-layer perceptron network, iMLP parameters are single-valued numbers and the operations follow the rules of interval arithmetic.

The parameters are estimated in the framework of supervised learning by minimizing an error function of the form:
\[
E = \frac{1}{N} \sum_{t=1}^{N} D(X_t, \hat{X}_t) + \lambda \Phi(\hat{f}),
\]
\[(10)\]
where \(N\) is the number of observations used for learning, \(\lambda \Phi(\hat{f})\) is a regularization term and \(D(X_t, \hat{X}_t)\) is a dissimilarity measure between the actual value \(X_t\) and the forecasting value \(\hat{X}_t\), that is computed using the Euclidian distance for intervals:
\[
D(X_t, \hat{X}_t) = (X_t^L - \hat{X}_t^L)^2 + (X_t^U - \hat{X}_t^U)^2.
\]
\[(11)\]
Roque et al. (2007) suggests the random initialization of network weights and the use of a back-propagation procedure to obtain the weights that minimize the error function as learning mechanism. Details regarding the iMPL model identification are found in Roque et al. (2007).

In order to address both linear and nonlinear patterns hidden in the data, the Holt’s smoothing multi-output support vector regression (Holt\(^1\)-MSRV) (Xiong et al. 2017) combines Holt\(^1\) and multi-output support vector regression methods as nonlinear hybrid “linear and nonlinear” modeling framework. The model assumes that an interval time series \(X_t\) is composed by a linear component and a nonlinear component in the form:
\[
X_t = L_t + N_t,
\]
\[(12)\]
where \( L_t \) and \( N_t \) are the linear and nonlinear components to be estimated, respectively.

Estimation of Holt\(^1\)-MSRV comprises three main steps (Xiong et al., 2017):

**Step 1:** The Holt\(^1\) method is applied to model the linear component \( L_t \). Thus, the residuals of the Holt\(^1\), \( e_t = X_t - \hat{L}_t \), are calculated since they contain information on the nonlinear component of ITS, where \( \hat{L}_t \) is the forecast of Holt\(^1\) at time \( t \).

**Step 2:** The multi-output support vector regression (MSVR) method is used to model the residuals \( e_t \) as the following:

\[
e_t = f(e_{t-1}, \ldots, e_{t-d}) + \epsilon_t,
\]

where \( f(\cdot) \) is a nonlinear function determined by MSVR, \( \epsilon_t \) is a random vector of errors, and \( d \) is the number of lagged error values considered by the model.

The MSVR is a generalization of the standard SVR that is able to solve the problem of regression estimation for multiple variables by means of the flexible multi-output structure. The theoretical foundation behind the MSVR technique can be found in Xiong et al. (2017) and Peréz-Cruz et al. (2002). As in Xiong et al. (2017), a MSRV model with \( 2d \) inputs, \( (e_{t-1}, \ldots, e_{t-d}) = [e_t^L, e_t^U, e_{t-1}^L, e_{t-1}^U, \ldots, e_{t-d}^L, e_{t-d}^U]^T \in \mathbb{R}^{2d} \), and two outputs, \( e_{t+1} = [e_{t+1}^L, e_{t+1}^U]^T \in \mathbb{R}^2 \), is applied, with each output corresponding to the forecast of the interval bounds. The idea behind this step is to capture the nonlinear component \( \hat{N}_t \) of the series by modeling the residuals of Holt\(^1\) with MSRV.

**Step 3:** Finally, the forecasts \( \hat{X}_t \) are computed by:

\[
\hat{X}_t = \hat{L}_t + \hat{e}_t,
\]

where \( \hat{e}_t \) is the estimated residuals using MSRV.

### 3.3. Forecasting combination approaches

Combinations of forecasts are useful because models of a real-world data generation process are likely to be misspecified, and picking only one of the available models is risky, especially if the data are not stationary (Lemke and Gabrys, 2010). The aim of forecast combination is to reduce modeling risk and to improve accuracy by exploiting the different strengths of different models while compensating for their weaknesses. This paper examines forecast combination techniques to evaluate their performance when interval forecasting techniques are combined. Here combinations are specified by the follow averaging procedures: simple average, simple average with trimming, variance-based averaging and ordinary least squares averaging.
Most linear combination is based on linear ensembles. Let \( \mathbf{Y} = [Y_1, Y_2, \ldots, Y_N]^T \) be the actual values of a data set of size \( N \), and \( \hat{\mathbf{Y}}^k = [\hat{Y}_1^k, \hat{Y}_2^k, \ldots, \hat{Y}_N^k]^T \) be the respective forecasts from the \( k \)-th model, with \( k = 1, \ldots, M \), where \( M \) is the number of forecasting methods. Linear combination methods produce forecasts as follows:

\[
\hat{Y}_t = w_1 \hat{Y}_t^1 + w_2 \hat{Y}_t^2 + \ldots + w_M \hat{Y}_t^M, \tag{15}
\]

where \( w_k, k = 1, 2, \ldots, M \), are non-negative and unbiased combination weights, i.e., \( w_k \geq 0 \) \( \forall k \) and \( \sum_{k=1}^M w_k = 1 \), and \( t = 1, \ldots, N \). Notice that, in this paper, combination weights are single-valued, thus the extension of the averaging methods is straightforward when considering interval-valued data, i.e. by using interval arithmetic.

Simple average (S-AVG) is the simplest method: it assigns equal weights to all individual forecasters and as such, it is free from weight estimation errors

\[
w_k = \frac{1}{M}, \quad \forall \ k. \tag{16}
\]

Simple average with trimming (S-AVGT) averages individual forecasts as well, but do not consider 20\% worst performance forecasts. The 20\% rate is between the 10-30\% recommended in Jose and Winkler (2008) and Lemke and Gabrys (2010).

Variance-based averaging (VB-AVG) weights forecasts using past forecasting performance. Let \( e_t^k = (Y_t - \hat{Y}_t^k) \) be the forecast error of the \( k \)-th method at \( t, k = 1, \ldots, M \). VB-AVG weights are found from the normalized unbiased inverse quadratic forecasting errors (i.e., the inverse of mean squared relative error), that is:

\[
w_k = \left( \frac{\sum_{t=1}^N (e_t^k)^2}{\sum_{t=1}^N (e_1^2)^2 + (e_2^2)^2 + \ldots + (e_M^2)^2} \right)^{-1}. \tag{17}
\]

In VB-AVG forecast errors are calculated using in-sample pairs of training and validation datasets. This scheme is based on the assumption that the forecasts from models with large in-sample errors receive less weight, and vice versa [Adhikari 2015].

The ordinary least squares averaging (OLS-AVG) is easy-to-implement approach with a good performance track-record. The combined forecast is assumed to be of the form:

\[
Y_t = w_1 \hat{Y}_t^1 + w_2 \hat{Y}_t^2 + \ldots + w_M \hat{Y}_t^M + \epsilon_t, \tag{18}
\]

where \( \epsilon_t \) is a white noise.
However, to compute the weights \( w_k, k = 1, \ldots, M \), by solving (18) using traditional least squares algorithm, in this paper we take advantage of the intervals midpoints, i.e.:

\[
\hat{y}_t = \hat{w}_1 \hat{y}_1^t + \hat{w}_2 \hat{y}_2^t + \ldots + \hat{w}_M \hat{y}_M^t,
\]  

(19)

where \( \hat{w}_k \) are the estimated weights, and \( \hat{y}_k^t \) is the midpoint of the \( t \)-th forecasted interval \( \hat{Y}_k^t \) obtained from the \( k \)-th model, \( k = 1, \ldots, M \).

3.4. Performance measurements

In order to access models performance, traditional time series error measures such as root mean squared error (RMSE) and symmetric mean absolute percentage error (SMAPE) are considered. To translate the performance metrics to intervals, let an ITS, \( \{Y_t\}_{t=1}^N \), be a sequence of intervals observed in successive instants in time \( t = 1, \ldots, N \). In the context of financial ITS, \( Y_L^t \) and \( Y_U^t \) correspond to the low and high stock price values at \( t \), respectively.

Traditional error measurements such as RMSE and SMAPE are not appropriate for ITS since the difference between the observed and the forecasted interval does not faithfully represent the concept of deviation besides considering interval arithmetics. One way to overcome this limitation is measuring the error in each one of the ITS attributes (lower and upper bounds). This is done by considering the two attribute time series, \( \{Y_L^t\} \) and \( \{Y_U^t\} \), and then estimating the error for each series by comparing it to the respective forecasted attribute time series, \( \{\hat{Y}_L^t\} \) and \( \{\hat{Y}_U^t\} \) ([Arroyo et al., 2011]). Hence, error metrics are computed for the two interval time series attributes. Therefore, RMSE and SMAPE are calculated for both lows (RMSE\(^L\), SMAPE\(^L\)) and highs (RMSE\(^U\), SMAPE\(^U\)) interval bounds, as follows:

\[
\text{RMSE}^L = \sqrt{\frac{1}{N} \sum_{t=1}^{N} (Y_L^t - \hat{Y}_L^t)^2},
\]

(20)

\[
\text{RMSE}^U = \sqrt{\frac{1}{N} \sum_{t=1}^{N} (Y_U^t - \hat{Y}_U^t)^2},
\]

(21)

\[
\text{SMAPE}^L = \frac{1}{N} \sum_{t=1}^{N} \frac{|Y_L^t - \hat{Y}_L^t|}{(|Y_L^t| + |\hat{Y}_L^t|)/2},
\]

(22)

\[
\text{SMAPE}^U = \frac{1}{N} \sum_{t=1}^{N} \frac{|Y_U^t - \hat{Y}_U^t|}{(|Y_U^t| + |\hat{Y}_U^t|)/2},
\]

(23)

where \( \hat{Y}_L^t \) and \( \hat{Y}_U^t \) are the lower and high \( t \)-th forecasted prices, respectively, \( Y_L^t \) and \( Y_U^t \) the \( t \)-th actual low and high, respectively, and \( N \) the sample size.
Since the data have an interval structure, it implies that both characteristics (lower and upper bounds) that describe intervals have to be taken in consideration jointly, using for example dissimilarity measures based on interval distance between observed interval and its forecast. To quantify the overall accuracy of the fitted and forecasted ITS, the mean distance error (MDE) of intervals is considered:

$$MDE = \frac{\sum_{t=1}^{N} D(Y_t, \hat{Y}_t)}{N},$$

where $Y_t$ and $\hat{Y}_t$ are the observed and forecasted ITS and $D(\cdot)$ is an interval distance. As suggested in Arroyo et al. (2011) and Rodrigues and Salish (2015), the Euclidean distance for intervals is considered:

$$D_E(Y_t, \hat{Y}_t) = \sqrt{(Y_t^L - \hat{Y}_t^L)^2 + (Y_t^U - \hat{Y}_t^U)^2}.$$  

(25)

Another distance measure for interval time series forecasting also adopted in this paper is the normalized symmetric difference (NSD) of intervals, computed as (Rodrigues and Salish, 2015):

$$D_{NSD}(Y_t, \hat{Y}_t) = \frac{\overline{\omega}(Y_t \cup \hat{Y}_t) - \overline{\omega}(Y_t \cap \hat{Y}_t)}{\overline{\omega}(Y_t \cup \hat{Y}_t)},$$

(26)

where $\overline{\omega}(\cdot)$ indicates the width of the interval.

The mean distance error of intervals using both Euclidian ($MDE^E$) and NSD ($MDE^{NSD}$) as distance function are computed. The advantage is that NSD distance is a normalized distance measure and it is not affected by the data magnitude as the Euclidian distance.

The computation of descriptive statistics for ITS are also conducted, as suggested by Rodrigues and Salish (2015). This paper calculates the coverage rate

$$R_C = \frac{1}{N} \sum_{t=1}^{N} \frac{\overline{\omega}(Y_t \cap \hat{Y}_t)}{\overline{\omega}(Y_t)},$$

(27)

and the efficiency rate

$$R_E = \frac{1}{N} \sum_{t=1}^{N} \frac{\overline{\omega}(Y_t \cap \hat{Y}_t)}{\overline{\omega}(\hat{Y}_t)}.$$  

(28)

These rates provide additional information on what part of the observed ITS is covered by its forecasts (coverage) and what part of the forecast covers the observed ITS (efficiency). If the observed intervals are fully enclosed in the predicted intervals then the coverage rate will be 100%, but the efficiency could be less than 100% and reveal the fact that the forecasted ITS is wider than the actual ITS. Hence, these statistics must be considered jointly. Therefore, the indication of a good forecast is observed when the coverage and efficiency rates are reasonably high and the difference between them is small (Rodrigues and Salish 2015).
In addition to the accuracy measurement, significant differences between a pair-wise of forecasting models or combining forecasting approaches are evaluated through the Turkey’s HSD test (Ramsay and Schaefer, 1996) considering a 5% level.

Finally, results are compared using an economic criteria. Considering the predicted highs and lows, a simple trading strategy is performed, as in the work of Xiong et al. (2017). Let $X_t^O$ and $X_t^C$ be the opening and closing prices at $t$, respectively, and $\hat{X}_{t+1}^U$ and $\hat{X}_{t+1}^L$ be the forecasted high and low prices for day $t+1$ after market closes on day $t$. The trading strategy is comprised by four steps as follows (Xiong et al., 2017): i) on a given day $t$, a ‘buy’ signal for the asset is generated if $\hat{X}_{t+1}^U - X_t^O > X_t^O - \hat{X}_{t+1}^L$; ii) if the ‘buy’ signal is observed for $k$ consecutive days beginning with day $t$, buy the asset on day $t+k-1$ using the closing value $X_{t+k-1}^C$; otherwise, hold the capital; iii) on another day $s$ subsequent to buying the asset, a ‘sell’ signal is generated if $\hat{X}_{s+1}^U - X_s^O < X_s^O - \hat{X}_{s+1}^L$; iv) sell the asset on day $s+k-1$ using the closing value $X_{s+k-1}^C$ of that day if a ‘sell’ signal has been observed for $k$ consecutive trading days beginning with day $s$; otherwise, hold the asset.

Notice that the predicting horizon is $h = 1$ in this paper. The observed consecutive trading days $k$ has to be set in advance and do not change as the steps of the trading strategy are conducted. Here, we take $k = 5$ as an example.

4. Empirical results

This section details the experimental analysis regarding the evaluation of intervalar models in financial ITS forecasting and their forecasting combinations. The datasets of ITS are from the main stock market indices of US and Brazilian markets, the S&P 500 and IBOVESPA, respectively, for the period from January 2000 to December 2015. The respective intervals are constructed by taking the lows and highs indices values as interval lower and upper bounds, respectively.

The data were divided into in-sample and out-of-sample sets. The in-sample set is comprised by the data for the period from January 2000 to December 2011, used for models training and also to determine the weights for the combining methods. The remaining four years of data concerning the out-of-sample set are used to access forecasting performance. The results are one-step-ahead forecasts.

The intervalar methods were implemented in the MATLAB computing environment. Smoothing parameters of the linear Holt\(^1\) method were estimated by minimizing the interval sum of squared

\[^{3}\text{Data are from Economatica.}\]
one-step-ahead forecast errors, as shown in [9], using constrained BFGS optimization approach. The iMLP identification was performed as in Roque et al. (2007), using the backpropagation algorithm for a cost function with no regularization term in a one hidden layer network with tangent hyperbolic transfer functions with the following structure: five hidden units and three lagged interval values \((d = 2)\) as input for S&P 500 index; and three hidden units and four lagged interval values \((d = 3)\) as input for IBOVESPA index. Finally, in the hybrid Holt-MSRV method, the MSRV with two outputs used radial basis function as kernel function with parameters selected by a grid search algorithm, as in Xiong et al. (2017).

Table 1 shows the descriptive statistics for S&P 500 and IBOVESPA ITS lower and upper (lows and highs) bounds. As expected, time series of intervals lower and upper bounds are very similar, mainly in terms of mean and standard deviations. These series have heavy left-side (right-side) tails as indicated by the negative (positive) skewness coefficients, and also lower (higher) values of kurtosis for IBOVESPA index (S&P 500). The Jarque-Bera (Bera and Jarque, 1981) statistics indicate that the both series are non-normal with a 99% confidence level. Figure 1 shows S&P 500 and IBOVESPA ITS over the whole sample period on a log scale. To improve visibility the daily lows of S&P 500 and IBOVESPA indices presented in Figure 1 are the actual daily low price series minus 0.1.

Table 1: Descriptive statistics of lows and highs values of S&P 500 and IBOVESPA indices for the period from January 2000 to December 2015.

<table>
<thead>
<tr>
<th>Statistic</th>
<th>S&amp;P 500</th>
<th>IBOVESPA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Lows</td>
<td>Highs</td>
</tr>
<tr>
<td>Mean</td>
<td>1321.79</td>
<td>1339.31</td>
</tr>
<tr>
<td>Maximum</td>
<td>2126.06</td>
<td>2134.72</td>
</tr>
<tr>
<td>Minimum</td>
<td>666.79</td>
<td>695.27</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>322.28</td>
<td>322.13</td>
</tr>
<tr>
<td>Skewness</td>
<td>0.8173</td>
<td>0.8351</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>3.1704</td>
<td>3.1736</td>
</tr>
<tr>
<td>Jarque-Bera</td>
<td>452.00</td>
<td>471.80</td>
</tr>
<tr>
<td>p-value</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
</tbody>
</table>
Table 2 shows the results from the traditional accuracy measures, RMSE and SMAPE, for S&P 500 and IBOVESPA indices lows (RMSE$_L$, SMAPE$_L$) and highs values (RMSE$_U$, SMAPE$_U$), from the intervalar approaches (Holt$_I$, iMLP and Holt$_I$-MSRV) and also concerning the forecasting combination techniques: simple average (S-AVG), simple average with trimming (S-AVGT), variance-based averaging (VB-AVG) and ordinary least squares averaging (OLS-AVG). Results are one-step-ahead forecasts for the out-of-sample data set, i.e. for the period from January 2012 to December 2015. Better results are highlighted in bold.

Considering both RMSE and SMAPE metrics, the combining forecasting methods performed better for lower and upper bounds of S&P 500 and IBOVESPA ITS, except for the ordinary least squares averaging (Table 2). OLS-AVG performed worst against the remaining combining techniques and showed also inferior results than Holt$_I$-MRSV. That fact may be due to the use of standard least squares algorithm to compute the combining weights by taking advantage of intervals midpoints. Thus, in such case the information related to intervals bounds interdependences is neglected.

Regarding the individual methods, the hybrid “linear and nonlinear” Holt$_I$-MRSV showed the lowest error metric values for both economies evaluated, followed by the nonlinear iMPL and the linear Holt$_I$ models (Table 2). Amongst the combining techniques, better accuracy for both S&P 500 and IBOVESPA indices was achieved by the simple average with trimming (S-AVGT) and the variance-based averaging (VB-AVG), which give a higher weight to the approaches related to lower

Figure 1: Interval-valued S&P 500 and IBOVESPA indices for the period from January 2000 to December 2015.
forecasting errors.

Table 2: Performance comparison of individual forecasts and forecast combinations in terms of RMSE and SMAPE metrics for lows and highs values of S&P 500 and IBOVESPA indices for the period from January 2012 to December 2015.

<table>
<thead>
<tr>
<th>Metric</th>
<th>Method</th>
<th>Holt\textsuperscript{L}</th>
<th>iMLP</th>
<th>Holt\textsuperscript{L}-MSRV</th>
<th>S-AVG</th>
<th>S-AVGT</th>
<th>VB-AVG</th>
<th>OLS-AVG</th>
</tr>
</thead>
<tbody>
<tr>
<td>Panel A: Interval-valued S&amp;P 500 index</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RMSE\textsuperscript{L}</td>
<td>Holt\textsuperscript{L}</td>
<td>67.70</td>
<td>61.19</td>
<td>56.77</td>
<td>50.17</td>
<td><strong>49.18</strong></td>
<td>50.11</td>
<td>57.65</td>
</tr>
<tr>
<td>RMSE\textsuperscript{U}</td>
<td>67.85</td>
<td>62.13</td>
<td>57.80</td>
<td>52.12</td>
<td>51.11</td>
<td><strong>49.91</strong></td>
<td>63.45</td>
<td></td>
</tr>
<tr>
<td>SMAPE\textsuperscript{L}</td>
<td>0.03036</td>
<td>0.02917</td>
<td>0.02610</td>
<td>0.02476</td>
<td>0.02311</td>
<td><strong>0.02273</strong></td>
<td>0.02781</td>
<td></td>
</tr>
<tr>
<td>SMAPE\textsuperscript{U}</td>
<td>0.03045</td>
<td>0.02836</td>
<td>0.02677</td>
<td>0.02503</td>
<td><strong>0.02489</strong></td>
<td>0.02491</td>
<td>0.03006</td>
<td></td>
</tr>
<tr>
<td>Panel B: Interval-valued IBOVESPA index</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RMSE\textsuperscript{L}</td>
<td>418.33</td>
<td>342.91</td>
<td>313.65</td>
<td>300.09</td>
<td>298.54</td>
<td><strong>279.13</strong></td>
<td>341.26</td>
<td></td>
</tr>
<tr>
<td>RMSE\textsuperscript{U}</td>
<td>484.20</td>
<td>367.12</td>
<td>342.19</td>
<td>327.66</td>
<td>317.00</td>
<td><strong>319.76</strong></td>
<td>343.77</td>
<td></td>
</tr>
<tr>
<td>SMAPE\textsuperscript{L}</td>
<td>0.00771</td>
<td>0.00487</td>
<td>0.00449</td>
<td>0.00410</td>
<td>0.00387</td>
<td><strong>0.00317</strong></td>
<td>0.00529</td>
<td></td>
</tr>
<tr>
<td>SMAPE\textsuperscript{U}</td>
<td>0.00804</td>
<td>0.00548</td>
<td>0.00492</td>
<td>0.00467</td>
<td>0.00451</td>
<td><strong>0.00449</strong></td>
<td>0.00532</td>
<td></td>
</tr>
</tbody>
</table>

In order to access the interval structure of the time series, S&P 500 and IBOVESPA ITS forecasts and forecast combinations were compared using accuracy measures designed for interval-valued data, as well as in terms of intervals descriptive statistics. The mean distance error (MDE) of intervals using both the Euclidean distance (MDE\textsuperscript{E}) and the normalized symmetric difference (NSD) distance of intervals (MDE\textsuperscript{NSD}) were considered. Further, coverage (R\textsuperscript{C}) and efficiency (R\textsuperscript{E}) rates concerns the intervals descriptive statistics. The results according to these measures are shown in Table 3. Better results are in bold.

Again, in terms of MDE\textsuperscript{E} and MDE\textsuperscript{NSD} metrics, the combining approaches outperform all the individual forecast method for both stock indices, except for OLS-AVG, as it ignores the possible mutual dependency between the daily highs and lows of ITS by using the intervals midpoints to compute the combining weights (Table 3). As far as the comparison between the intervalar methods (Holt\textsuperscript{L}, iMLP and Holt\textsuperscript{L}-MSRV) is concerned, the hybrid Holt\textsuperscript{L}-MSRV provides higher accuracy, due to its ability of capturing the linear and nonlinear patterns hiding in ITS simultaneously. When considering the comparison among the combining methodologies, the results indicate again the
superiority of S-AVGT and VB-AVG for both US and Brazilian stock markets.

Concerning intervals descriptive statistics, coverage (R\textsuperscript{C}) and efficiency (R\textsuperscript{E}) rates, they are also reported in Table 3. They provide additional information about the adequacy of the forecasts. In this case, these statistics reveal the percentage of the actual ITS is covered by its forecast (coverage), and what part of the forecast covers the realized ITS (efficiency). To evaluate the models, these rates must be considered jointly and the higher their values the better is the forecast. Further, the closeness of the results of these two statistics can be taken as an indicator of the quality of the forecasts. For both coverage and efficiency rates, combined techniques (except for OLS-AVG) showed the highest values, indicating more accurate predictions (Table 3). Variance-based averaging provides in general better results for both S&P 500 and IBOVESPA indices. Notice that these two statistics values are very close, which corroborate the adequacy of the models. Again, the individual interval forecasting methods performed worst, but better than OLS-AVG approach.

Table 3: Performance comparison of individual forecasts and forecast combinations in terms of interval quality measures for S&P 500 and IBOVESPA ITS for the period from January 2012 to December 2015.

<table>
<thead>
<tr>
<th>Metric</th>
<th>Method</th>
<th>Holt\textsuperscript{1}</th>
<th>iMPL</th>
<th>Holt\textsuperscript{1}</th>
<th>MSRV</th>
<th>S-AVG</th>
<th>S-AVGT</th>
<th>VB-AVG</th>
<th>OLS-AVG</th>
</tr>
</thead>
<tbody>
<tr>
<td>MDE\textsuperscript{E}</td>
<td>88.72</td>
<td>85.56</td>
<td>82.18</td>
<td>80.13</td>
<td><strong>78.19</strong></td>
<td>78.53</td>
<td>82.36</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MDE\textsuperscript{NSD} &amp; 0.4201 &amp; 0.4188 &amp; 0.3918 &amp; 0.3765 &amp; 0.3700 &amp; <strong>0.3687</strong> &amp; 0.4177</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R\textsuperscript{C}   &amp; 0.7819 &amp; 0.7991 &amp; 0.8355 &amp; 0.8510 &amp; 0.8572 &amp; <strong>0.8721</strong> &amp; 0.8699</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R\textsuperscript{E}   &amp; 0.7701 &amp; 0.7743 &amp; 0.7860 &amp; 0.7911 &amp; 0.8208 &amp; <strong>0.8419</strong> &amp; 0.7952</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Panel A: Interval-valued S\&P 500 index

Panel B: Interval-valued IBOVESPA index

To identify the existence of statistically significant differences among the interval methods and their forecast combinations, the Turkey’s HDS test is performed to compare all pairwise differences at the 0.05 level. The results of Turkey’s test are shown in Table 4. For each ITS, we rank the methods and forecast combination approaches from 1 (the best) to 7 (the worst). The VB-AVG,
S-AVGT and S-AVG combining methods perform statistically better than all of other competitors for both S&P 500 and IBOVESPA ITS forecasting, under confidence level of 95%. Regarding the individual forecasts, Holt^I-MSRV outperforms the iMPL, Holt^I and OLS- AVG methods at 95% statistical significance. Additionally, for the IBOVESPA ITS, the results from Table[4] also indicates the statistical superior performance of iMPL over the Holt^I linear method at 0.05 level.

Table 4: Forecasting models and forecast combination approaches ranking from Turkey’s HSD test for S&P 500 and IBOVESPA ITS for the period from January 2012 to December 2015.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Panel A: Interval-valued S&amp;P 500 index</th>
<th>Panel B: Interval-valued IBOVESPA index</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>VB-AVG</td>
<td>VB-AVG</td>
</tr>
<tr>
<td>2</td>
<td>&lt; S-AVGT</td>
<td>&lt; S-AVGT</td>
</tr>
<tr>
<td>3</td>
<td>&lt; S-AVG</td>
<td>&lt; S-AVG</td>
</tr>
<tr>
<td>4</td>
<td>&lt; Holt^I-MSRV</td>
<td>&lt; Holt^I-MSRV</td>
</tr>
<tr>
<td>5</td>
<td>&lt; OLS-AVG</td>
<td>&lt; OLS-AVG</td>
</tr>
<tr>
<td>6</td>
<td>&lt; iMLP</td>
<td>&lt; iMLP</td>
</tr>
<tr>
<td>7</td>
<td>&lt; Holt^I</td>
<td>&lt; Holt^I</td>
</tr>
</tbody>
</table>

(*) indicates the mean difference between the two competing methods is significant at the 5% level.

Finally, based on the intervals forecasts and intervals forecast combinations, the results of a trading strategy were achieved to access the predictability of the approaches for an investor accordingly as the buy/sell actions detailed in subsection [3.4]. A one-time 0.1% deduction was considered in order to mimic the transaction cost. Also, it is supposed that the investors can enter the market at any time during the evaluation period. Table[5] presents the average return and percentage of trades resulting in positive returns of the individual intervalar methods and combinations from a trading strategy concerning ITS forecasts. All returns are expressed in annualized terms.

Generally speaking, both the average return and percentage of trades resulting in positive returns suggest that the examined methods perform quite well. It should be noted that the smallest average annualized return is 37.19% and 26.31% for S&P 500 and IBOVESPA, respectively, when the individuals forecasts obtained from the Holt^I method are considered. The largest average annualized return for S&P 500 is 45.17% when the simple average with trimming approach is used for forecasting combination. On the other hand, concerning the IBOVESPA index, a average return of 36.55% is the highest achieved, obtained when the variance-based combining technique is performed. It is worthy to note that the percentage of profitable trades is always larger than 50% (Table[5]).

As far as the comparison among the individual forecasting methods and their combinations in
terms of the average annualized returns and percentage of trades with a positive annualized return is concerned, for both economies the forecast combinations are superior to individuals forecasts, except when the OLS-AVG approach is concerned (Table 5). Further, S-AVGT and VB-AVG are the most profitable forecast combining methods when trading strategies are performed for S&P 500 and IBOVESPA ITS indices, respectively. Again, Holt$^1$-MSRV outperforms Holt$^1$ and iMLP for both stock market indices evaluated.

Table 5: Performance comparison of individual forecasts and forecast combinations in terms of economic criteria for S&P 500 and IBOVESPA ITS for the period from January 2012 to December 2015.

<table>
<thead>
<tr>
<th></th>
<th>Holt$^1$</th>
<th>iMPL</th>
<th>Holt$^1$-MSRV</th>
<th>S-AVG</th>
<th>S-AVGT</th>
<th>VB-AVG</th>
<th>OLS-AVG</th>
</tr>
</thead>
<tbody>
<tr>
<td>Panel A: Interval-valued S&amp;P 500 index</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average</td>
<td>37.19%</td>
<td>39.56%</td>
<td>41.87%</td>
<td>43.76%</td>
<td>45.17%</td>
<td>44.19%</td>
<td>38.96%</td>
</tr>
<tr>
<td>Positive</td>
<td>54.65%</td>
<td>58.11%</td>
<td>60.45%</td>
<td>62.32%</td>
<td>63.21%</td>
<td>62.90%</td>
<td>57.16%</td>
</tr>
<tr>
<td>Panel B: Interval-valued IBOVESPA index</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average</td>
<td>26.31%</td>
<td>28.46%</td>
<td>30.19%</td>
<td>33.16%</td>
<td>36.55%</td>
<td>36.55%</td>
<td>24.87%</td>
</tr>
<tr>
<td>Positive</td>
<td>43.77%</td>
<td>45.10%</td>
<td>48.94%</td>
<td>50.98%</td>
<td>53.18%</td>
<td>53.70%</td>
<td>44.03%</td>
</tr>
</tbody>
</table>

Summing up, the empirical results in this paper provide the following conclusions: i) forecast combinations provide superior performance to individual forecasts also for interval-valued financial data concerning the methods and markets considered in the work; ii) combination approaches that take into account the forecasting errors from the individual forecasters achieved better accuracy; iii) the hybrid Holt$^1$-MSRV method performs strikingly better than the linear and nonlinear intervalar approaches, Holt$^1$ and iMPL respectively, indicating that the hybrid “linear and nonlinear” modeling framework can effectively improve prediction performance in the case of ITS forecasting, as in line with the work of Xiong et al. (2017); iv) considering the economic criteria, the results indicate that combination techniques can be used as a promising solution for ITS forecasting by investors in developed and emergent stock markets.
5. Conclusion

Interval time series (ITS) are time series where each period in time is described by an interval-valued variable. In finance, ITS can be described as the evolution of the lows and highs prices of an asset throughout time. Highs and lows prices of an asset are commonly used in technical indicators, as in the construction of candlestick charts, for example. Further, these prices are also related to the concept of volatility by means of range-volatility estimators. Thus, their accurate forecasts for financial markets participants are also very important in the role of derivative pricing, asset allocation and risk management. The literature has been shown that the use of intervalar methods for financial ITS forecasting improves accuracy as they are able to take into account the intrinsic relationship of interval bounds by processing highs and lows prices simultaneously, against traditional time series approaches that consider interval bounds individually. However, the selection of an appropriate forecasting method in many real-world problems is challenging. Currently, a large amount of algorithms are available, and their combination became of major interest because it is unlikely that any single algorithm will surpass all others in any forecasting scenario.

This paper evaluated the forecasts and forecasting combinations of intervalar models for financial ITS forecasting. The interval Holt’s exponential smoothing (Holt\textsuperscript{I}), the interval multilayer perceptron neural network (iMLP), and the hybrid Holt’s smoothing multi-output support vector regression (Holt\textsuperscript{I}-MSRV) methods are considered, as well as their combinations using simple average, simple average with trimming, variance-based averaging and ordinary least squares averaging approaches. The empirical application concerned one-step-ahead forecasts and forecast combinations for ITS of the main stock indices of US and Brazil, S&P 500 and IBOVESPA, respectively, for the period from January 2000 to December 2015. Financial ITS are constructed by means of lows and highs stock indices as interval representatives. In addition to the use of accuracy measures and statistical tests to compute models performance, this paper evaluated the results using quality measures designed for the interval time series framework, as well as in terms of an economic criteria: a simple trading strategy based on the interval forecasts.

The results showed that forecast combinations of intervalar methods provide higher accuracy than the individual forecasts for both S&P 500 and IBOVESPA indices. The hybrid method Holt\textsuperscript{I}-MSRV performed better in comparison to the alternative approaches, since it is constructed under the “linear and nonlinear” modeling framework, which is able to capture both linear and nonlinear patterns of financial ITS. Regarding the forecast combination methods, they can be considered equally accurate in statistical terms, except for the ordinary least squares averaging. A slightly
better performance of simple average with trimming and variance-based averaging methodologies was verified. In general, considering both statistical and economic criteria, the forecast combinations can be used as a promising solution for ITS forecasting. Future work shall include the analysis of medium- and long-term forecasting horizons in forecast combination as well as the application in risk management by using range-based volatility estimators.

References


