Revista Colombiana de Estadística - Applied Statistics
January 2021, volume 44, issue 1, pp. 123 a 157
http://dx.doi.org/10.15446/rce.v44n1.85116

Combining Interval Time Series Forecasts. A First Step in a Long Way (Research Agenda)

La combinación de predicciones de series temporales de intervalo (STI)

Carlos G. Maté[1,a]

[1]Department of Industrial Organization, ICAI School of Engineering, Comillas Pontifical University, Madrid, Spain

Abstract

We observe every day a world more complex, uncertain, and riskier than the world of yesterday. Consequently, having accurate forecasts in economics, finance, energy, health, tourism, and so on; is more critical than ever. Moreover, there is an increasing requirement to provide other types of forecasts beyond point ones such as interval forecasts. After more than 50 years of research, there are two consensuses, "combining forecasts reduces the final forecasting error" and "a simple average of several forecasts often outperforms complicated weighting schemes", which was named "forecast combination puzzle (FCP)". The introduction of interval-valued time series (ITS) concepts and several forecasting methods has been proposed in different papers and gives answers to some big data challenges. Hence, one main issue is how to combine several forecasts obtained for one ITS. This paper proposes some combination schemes with a couple or various ITS forecasts. Some of them extend previous crisp combination schemes incorporating as a novelty the use of Theil's U. The FCP under the ITS forecasts framework will be analyzed in the context of different accuracy measures and some guidelines will be provided. An agenda for future research in the field of combining forecasts obtained for ITS will be outlined.

***Key words*:** Efficient market hypothesis; Equal weights; Financial markets; Forecast combination; Optimal weight; Random walk model.

Resumen

Cada día observamos un mundo más complejo, incierto y con mayor riesgo que el mundo de ayer. Luego, tener pronósticos precisos en economía, finanzas, energía, salud, turismo, etc.; es más crítico que nunca. Además, existe un requisito creciente de proporcionar otro tipo de pronósticos más allá de los puntuales, como los pronósticos de intervalos. Después de más

[a]Ph.D. E-mail: cmate@comillas.edu

de 50 años de investigación, hay dos consensos, "combinar pronósticos reduce el error de pronóstico final" y "un promedio simple de varios pronósticos a menudo supera complicados esquemas de ponderación", que se denominó "rompecabezas de combinación de pronósticos (FCP)". La introducción de los conceptos de series de tiempo de intervalo (ITS) y varios métodos de pronóstico se han propuesto y dan respuestas a algunos desafíos de los grandes datos. Entonces, un problema es cómo combinar varios pronósticos obtenidos para una ITS. Este documento propone algunos esquemas combinados con un par o varios pronósticos ITS. Algunos extienden esquemas previos para datos puntuales, incorporando como novedad la U de Theil. El FCP en el marco de pronósticos ITS se analizará con diferentes medidas de exactitud y se proporcionarán algunas pautas. Se describirá una agenda para futuras investigaciones en la combinación de pronósticos obtenidos para ITS.

***Palabras clave*:** Combinación de pronósticos; Hipótesis de mercado eficiente; Mercados financieros; Modelo de caminata aleatoria; Peso óptimo; Ponderaciones iguales.

1. Introduction

There is an immense number of various forecasting methods and models. A subsection of these uses only the past values of an observed variable to predict future outcomes; in this case, both the input and output of a method can be modeled by a time series, i.e. a sequence of observations of the same variable in uniform time periods.

On the one hand, the value of using more than just point forecasts has been shown in different contexts such as finance or tourism. For example, in exchange rates forecasting, Wang & Wu (2012) incorporate out-of-sample interval forecasting and Sarno & Valente (2005) use density forecasts. Recently, Li et al. (2018) have stated that interval forecasts can provide more comprehensive information to improve tourism forecasting accuracy. They give guidelines for producing accurate interval forecasts that benefit policy-making for a wide array of applications in practice.

On the other hand, the individual values of a time series do not necessarily need to be single numbers; they could, for example, be represented by a more complex structure, such as a function (Hyndman & Shang 2009 or Gao et al. 2019), a density function (Tay & Wallis 2000 or lately Bassetti et al. 2020), a histogram (Arroyo & Maté, 2009), a boxplot of some subset of the data, or an interval.

Interval-valued data is a particular case of symbolic data as it is viewed in the field of symbolic data analysis (SDA). SDA states that symbolic variables (lists, intervals, frequency distributions, etc.) are better suited for describing complex real-life situations than single-valued variables; further details can be found in Billard & Diday (2003), Billard & Diday (2006) and Noirhomme-Fraiture & Brito (2011). Hsu & Wu (2008) propose using interval-valued data to establish a model and to predict.

Intervals could be used as a framework to incorporate the magnitude of the measurement error in estimates. For example, recently, Glennon et al. (2018) investigate the extent to which forecast combination methods can be used to improve the property-level and find that relatively simple forecast combination schemes have the potential to reduce the error in property value estimates, possibly by a wide margin.

An interval time series (ITS) assigns to each time period an interval covering the values taken by the observed variable. Each interval has four characteristic attributes, since it can be defined in terms of lower and upper boundaries, center, and radius. In finance, the evolution for every asset (equities, commodities, or exchange rates) in every period of time (for example, day or month) of lows and highs is an ITS. See, as an example, Figure 1 for the case of monthly intervals of lows and highs prices for General Electric in the first part of the year 2017.

The analysis and forecasting of ITS is a young research area, dating back less than 20 years, and still presents a wide array of open issues. For a recent survey of methodologies and techniques for ITS forecasting readers are referred to different papers such as Arroyo & Maté (2006), Han et al. (2008), Maia et al. (2008), García-Ascanio & Maté (2010), Maia & de Carvalho (2011) and Arroyo et al. (2011*a*), among others.

In a different path, forecasts combination is a mature field of research with key papers such as Bates & Granger (1969), Clemen (1989) and Timmermann (2006). Additional important references are Gneiting (2011) or Gneiting & Katzfuss (2014). The connection with multivariate analysis has been highlighted in Maté (2011), among others. In engineering, environmental sciences, energy and other areas, this field is called ensemble forecasts (see, for example, Avci et al. (2018)). With this name there are several applications in big data such as Ordiano et al. (2018) or Galicia et al. (2019) in ensemble learning. The problem of model selection (see, for example, Moral-Benito (2015) or Gibbs (2017)) is related to the forecast combination problem and is revisited in Kourentzes et al. (2019).

After more than 50 years of research, there is a general consensus "combining forecasts reduces the final forecasting error". This consensus has been also established through some forecasting competitions such as the last one, M4 (see Makridakis et al. 2020). According to Atiya (2020), forecast combinations were big winners in the M4 competition.

As part of this consensus, there is also a well-known fact "a simple average of several forecasts often outperforms complicated weighting schemes", which was named "forecast combination puzzle (FCP)" by Stock & Watson (2004). Recently, Thomson et al. (2019) suggest an analytical framework that can be used for enhancing the assessment of forecast performances and guiding decisions as to which forecasters should be pooled to obtain an effectively combined forecast. It is shown that composite forecasts (formed using a simple average) support previous research (e.g., Armstrong 2001, Clemen 1989 or Timmermann 2006, among many others) and confirm the benefits of forecasts combination.

Lately, Shaub (2020) concludes that an examination of a simple ensemble forecasting method shows that equal arithmetic averaging of base models can serve as a hedge for model selection risk and help to prevent some large forecasting errors. This simple combination approach can be of particular interest to business users and software engineers who need to forecast a large number of time series automatically, without intervention, and without a global model that needs retraining with the addition of new time series or data drawn from a new data generating process.

Concerning large datasets, Song & Liu (2017) propose one framework of tourism forecasting with big data. They identify three important steps: (1) data exploration, which is the data processing that prepares the proper data for the model; (2) use modeling techniques to predict user behavior based on their previous business transactions and preferences; (3) optimize the forecast results and decrease the forecast failure risk by model selection and combination forecasting. ITS forecasting methods and the combination approaches developed in this paper can help in the above steps.

TABLE 1: GE monthly low-high prices dataset 2017: January to July; with two forecasting methods M1 and M2.

Month	LowGE	HighGE	LowM1	HighM1	LowM2	HighM2
January	31.4	31.84	30.1607	30.9881	29.3952	30.2016
February	29.56	29.81	30.929	31.3575	30.144	30.5616
March	29.82	30.35	29.1166	29.3629	28.3776	28.6176
April	29.75	30	29.3727	29.8948	28.6272	29.136
May	28.93	29.17	29.2988	29.55	28.5552	28.8
June	27.5	27.88	28.4961	28.7325	27.7728	28.0032
July	27.06	27.59	27.0875	27.4579	26.4	26.76096

Given that we are able to use several forecasting methods with ITS, one main issue in SDA with interval-valued data is how to combine several forecasts obtained for an ITS. Figure 1 shows the problem graphically and Table 1 provides the corresponding dataset. The ITS shown by a solid line (red), joining the interval-valued data at every month, reports about the evolution of the monthly prices of General Electric (GE) during the seven first months of the year 2017. The couple of ITS shown by a dotted line (purple) and a dashed line (blue) informs about the evolution of the monthly interval price forecasts of GE with the methods M1 (purple) and M2 (blue), respectively, during the same period. The problem is to obtain a linear combination of these two ITS forecasting methods M1 and M2 which is the best approach to the real ITS.

This paper proposes several combination schemes with a couple or various ITS forecasts and analyzes the forecast combination puzzle under the ITS forecasts framework in the context of different accuracy measures. Section 2 provides an overview of the framework of the linear combination of multiple crisp forecasts. Sections 3, 4, and 5 introduce essential ITS concepts such as notation, distance measures, accuracy measures, and so on; when forecasting ITS. Section 6 briefly describes the methods of combining linearly a couple of ITS forecasts in an optimal way. Section 7 explains the methodology showing empirical results. Section 8 discusses the problem of the combination of several ITS forecasts and suggests a research agenda in this field. Finally, Section 9 collects some concluding remarks and outlines several open lines of research.

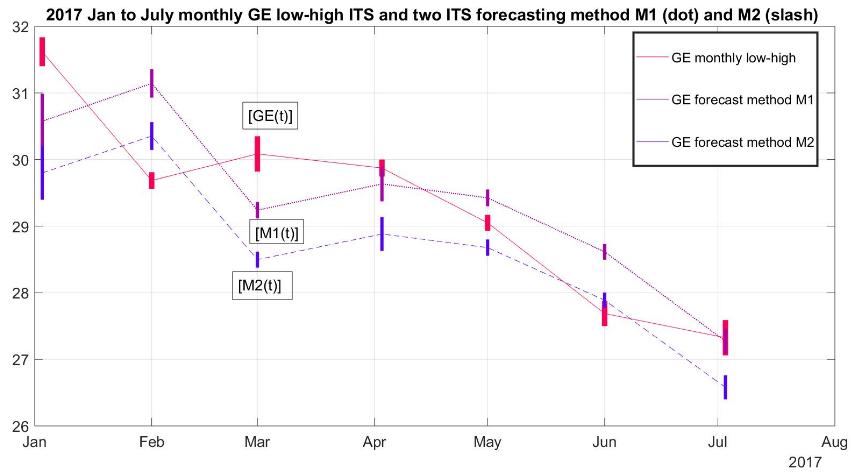

Figure 1: GE monthly ITS prices in 2017 (January to July) with two forecasting methods M1 and M2 for this ITS.

2. Linear Combination of Multiple Crisp Time Series Forecasts

Let $F_{t+h,i}$ the forecast of one time series x_t at time t+h according to the method i, with $i = 1, \dots, k$; for $h = 1, \dots, l$. Let $F_{t+h,C}$ the linear combined forecast with time-varying weights. That is

$$F_{t+h,C} = \sum_{i=1}^{k} \omega_{t,i} * F_{t+h,i} \tag{1}$$

with $t = 1, \dots, T$. Usually, it is assumed that $\sum_{i=1}^{k} \omega_{t,i} = 1$. In addition, sometimes the restriction $\omega_{t,i} \geq 0$ is incorporated. With both restrictions, the problem is named a convex linear combination approach.

The problem is to get an optimal linear combination forecast $\hat{x}_{t+h}$, for $h = 1, \dots, l$, given x_t, with $t = 1, \dots, T$. Additional details of the above setup can be found in Timmermann (2006).

The distinction between using a small number of methods or many predictors has been considered in different papers such as Stock & Watson (2006), which only consider linear forecasts, that is, forecasts that are linear in the predictors, because this has been the focus of almost all research on economic forecasting.

Some economic time series are often integrated (non-stationary), and this fact has implications for the form of a linear combination of the forecasts. See, for example, Hallman & Kamstra (1989).

2.1. The Random Walk Crisp Time Series Model, Naïve Forecasting and Efficient Market Hypothesis (EMH)

A random walk time series is defined as a process where the current value of a variable is composed of the past value plus an error term defined as white noise (a variable with zero mean and variance one which usually is assumed following a normal law). Algebraically a simple random walk is represented as follows:

$$X_t = X_{t-1} + \varepsilon_t \tag{2}$$

A process of this type implies that the best prediction of X for the next period is the current value. That is,

$$\hat{X}_t = X_{t-1} \tag{3}$$

which is known as the naïve forecasting method. That is, the process does not allow to predict the value of the change $(X_{t+1} - X_t)$ or predicts a zero expected change. The change in X is random. It can be shown that the mean of a random walk process is constant but its variance is not. Therefore a random walk process is non-stationary, and its variance increases with t. In practice, the presence of a random walk process makes the forecasting process very simple, since all the future values of X_{t+s} for s > 0, are simply X_t.

Fama (1995) describes briefly and simply the theory of random walks and some of the important issues it raises concerning the work of market analysts. More recently, Rapach & Zhou (2013) survey the literature on stock return forecasting and point out that forecast combination, among other strategies, improve forecasting performance by addressing the substantial model uncertainty and parameter instability surrounding the data-generating process for stock returns.

Roughly speaking and following Fama & Blume (1966), the Efficient Market Hypothesis (EMH) establishes that assets are always priced at their fair values, fully reflecting all the information possessed. As a consequence, speculators cannot predict future returns and systematically beat the market without leaning towards riskier assets. Therefore, under the EMH, stock returns follow random walk models and cannot be predicted to any extent. For an updated analysis of the EMH see, for example, Naseer et al. (2015), among many other references.

2.2. Accuracy Measures in Crisp Time Series Forecasting

In order to compare several forecasting methods or different combination approaches, accuracy measures of $\hat{x}_t$, such as MSE defined by

$$MSE = \frac{\sum_{t=1}^{T} \left(x_t - \widehat{x}_t\right)^2}{T} \tag{4}$$

or RMSE, the square root of the MSE, or other measures are used (see, for example, Hyndman & Koehler 2006). Assuming we have information for $t = 1, \dots, T$ of a time series and a forecasting method M for such series, the MSE at time T for the method M will be noted by $MSE(M, T)$, or $MSE(M)$ if there is no doubt about the period considered.

One quite interesting measure for the above issue is Theil's U. It compares one forecasting model or method to a naïve model by

$$U = \sqrt{\frac{\sum_{t=2}^{T} (x_t - \widehat{x}_t)^2}{\sum_{t=2}^{T} (x_t - x_{t-1})^2}} \tag{5}$$

The U statistic compares the model with the simple random walk.

- When the value of U is equal to 1, the performance of the currently observed model is the same as that of the simple random walk.
- For $U > 1$, the model performs worse than the simple RW.
- For $U < 1$, the model's performance is better than the simple RW.

The connection between EMH, random walk model, and Theil's U deserves future thorough research. One paper relating these three important keywords is Riddington (1993).

2.3. Linear Combination of Two Crisp Time Series Forecasts. The use of Theil's U

In this case, the equation (1) is given by

$$F_{t+h,C} = \omega * F_{t+h,1} + (1 - \omega) * F_{t+h,2} \tag{6}$$

In the seminal paper Bates & Granger (1969), the optimal weight when using two forecasting methods M1 and M2 is given by

$$\omega = \frac{MSE(M2) - ACE(M1, M2)}{MSE(M1) + MSE(M2) - 2 * ACE(M1, M2)} \tag{7}$$

where $ACE(M1, M2)$ stands for the average crossed error using methods M1 and M2 and is given by

$$ACE(M1, M2) = \frac{\sum_{t=1}^{T} (x_t - \widehat{x}_{t,M1})(x_t - \widehat{x}_{t,M2})}{T} \tag{8}$$

In the case of two unbiased methods M1 and M2 to forecast the time series, the MSE of every method is the variance noted by σ_1^2 and σ_2^2. Then, the weight obtained in (7) is given by

$$\omega = \frac{\sigma_2^2 - \rho\sigma_1\sigma_2}{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2} \quad (9)$$

which is the key result in Bates & Granger (1969). Several studies have been accomplished to analyze the nature of the above weights, including the possibility of negative weights. See, for example, Dickinson (1975).

Obviously, weights can be varying with time and to automate the process for the case of a rolling window.

Weights can also be optimized according to the use of Theil‘s U, which is a new approach until our knowledge goes. That is, choosing weights to optimize the corresponding Theil‘s U of the combination of two forecasts is a new approach and the optimal weight is given by

$$\omega = \frac{U_{M2}^2 - ACEU(M1, M2)}{U_{M1}^2 + U_{M2}^2 - 2 * ACEU(M1, M2)} \quad (10)$$

where $ACEU(M1, M2)$ stands for the average crossed error under Theil‘s U using methods M1 and M2 and is given by

$$ACEU(M1, M2) = \frac{\sum_{t=2}^{T} (x_t - \widehat{x}_{t,M1})(x_t - \widehat{x}_{t,M2})}{\sum_{t=2}^{T} (x_t - x_{t-1})^2} \quad (11)$$

One quite important advantage of computing the Theil‘s U of the optimized combined method is that we can discard such an approach when that value is greater or equal than 1. This lacks in the literature of forecasts combination research.

2.4. Weighting Usual Schemes in the Combination of Several Crisp Forecasts. A New Additional Scheme

Several weighting schemes can be found in the literature. Following Genre et al. (2013), we will restrict to the class of linear combinations and focus on those methods which emphasize parsimony, with the goal to minimise the estimation error as much as possible.

- Trimming and other statistical combinations

  These include the median and other trimmed mean measures that remove extreme values from the cross-section of forecasts, assigning a zero weight to some forecasts and equal weights to all others.

- Performance-based combinations

  These approaches assign higher weights to forecasts with relatively good forecasting track records and lower weights to forecasts with poor performances. One of them is the Inverse of Root Mean Squared Error (IRMSE) averaging. Weights are obtained through the following expressions

$$\omega_{t,i} = \frac{RMSE_{t,i}^{-1}}{\sum_{i=1}^{k} RMSE_{t,i}^{-1}} \tag{12}$$

Another alternative, not found in the literature until our knowledge goes, would be the inverse of Theilťs U Squared (IThUS) averaging. Weights are obtained through the following expressions

$$\omega_{t,i} = \frac{(U^2)_{t,i}^{-1}}{\sum_{i=1}^{k} (U^2)_{t,i}^{-1}} \tag{13}$$

The better (worst) the method is, the higher (lower) the corresponding weight for such method is.

- Sequential combination

  When combining more than two forecasts, Winkler & Clemen (1992) propose combining sequentially. Instead of combining all k forecasts at once, we could combine them sequentially in $k-1$ steps, adding one forecast at each step.

3. Basic Concepts When Analyzing ITS

In the following, some important definitions and facts about ITS will be provided.

3.1. Definitions and Notation

Definition 1 (Interval)**.** An interval $[x]$ over the base set $(E, \leq)$ is an ordered pair $[x] = [x^L, x^U]$ where $x^L, x^U \in E$ are the lower and upper bounds of the interval, respectively, such that $x^L \leq x^U$.

An equivalent representation of an interval is given by the centre (midpoint) and radius (half-range) of the interval, namely $\langle x \rangle = \langle x^C, x^R \rangle$, where $x^C = \frac{1}{2}(x^L + x^U)$ and $x^R = \frac{1}{2}(x^U - x^L)$.

The class of nonempty compact intervals of the real line has been represented in several ways.

[a] $\mathcal{K}_c(\mathbb{R})$ (Theory of sets notation)

[b] $\mathrm{I}(\mathbb{R})$ or $\mathbb{IR}$ (Interval analysis notation)

Hence, $\mathcal{K}_c(\mathbb{R}) = \{[a, b] : a, b \in \mathbb{R}, a \leq b\}$ will denote the class of nonempty compact intervals.

3.2. Some Stochastic Issues With Interval Time Series

In this section, we consider the stochastic issues that arise when using interval-valued time series (ITS). Following Arroyo et al. (2010), Tay & Wallis (2000), Kubica & Malinowski (2006) and Sinova & Van Aelst (2015); we first define the important concepts of interval random variable and interval stochastic process, and consequently define the concept of ITS.

Let $(\Omega, \mathcal{F}, P)$ be a probability space, where Ω is the set of elementary events, $\mathcal{F}$ is the σ-field of events and $P : \mathcal{F} \to [0, 1]$ the σ-additive probability measure. We define a partition of Ω into sets $A_X(x)$ such that $A_X(x) = \{\omega \in \Omega | X(\omega) = x\}$, where $x \in [x^L, x^U]$.

Definition 2 (Interval random variable)**.** A mapping $[X] : \mathcal{F} \to \mathcal{K}_c(\mathbb{R}) \subset \mathbb{R}$ such that for each $x \in [x^L, x^U]$, there exists a set $A_X(x) \in \mathcal{F}$, is called an interval random variable (iRV).

The minimum and maximum temperatures in a place every day (or every time frame considered), in Celsius degrees, is an example of an iRV.

Definition 3 (Interval-valued stochastic process)**.** An interval-valued stochastic process is a collection of interval random variables that are indexed by time, that is $\{[X_t]\}$ for $t \in T \subset \mathbb{R}$ with each $[X_t]$ following the above definition.

The evolution through the time of the minimum and maximum temperatures is an interval-valued stochastic process.

Definition 4 (Interval-valued time series)**.** An interval-valued time series (ITS) is a realization of an interval-valued stochastic process. It may be equivalently denoted as $\{[x_t]\} = \{[x_t^L, x_t^U]\} = \{\langle x_t^C, x_t^R\rangle\}$ for $t = 1, 2, \ldots, T$.

Definition 5 (Interval-valued random walk)**.** An *ITS* $[Y_t]$ is an interval-valued random walk (iRW) with drift if

$$[Y_t] = [\mu] + [Y_{t-1}] + [\varepsilon_t] \tag{14}$$

As a consequence

$$Y_t^L = \mu^L + Y_{t-1}^L + \varepsilon_t^L \tag{15}$$

$$Y_t^U = \mu^U + Y_{t-1}^U + \varepsilon_t^U \tag{16}$$

That is, under an iRW with drift, the interval-valued variable at time t is the result of one constant interval plus the interval valued variable at time $t-1$ plus an error at time t in the way of an interval. The case of $[\mu] = [0]$ is the iRW or iRW without drift.

The connection between stationarity, random walks, trends, unit roots, time series regression, and so on; is out of the scope of this paper (see, for example, Nelson & Plosser 1982, Phillips & Perron 1988 and Kwiatkowski et al. 1992).

The iRW model for an ITS provides the simplest forecast at every time for an ITS, the so-called naive forecast.

Definition 6 (Interval-valued naive forecast)**.** The interval-valued naive forecast for an *ITS* $[Y_t]$, for $t = 2, 3, \dots, T$; is given by

$$[\hat{Y}_t] = [Y_{t-1}] \tag{17}$$

As a consequence

$$\hat{Y}_t^L = Y_{t-1}^L \tag{18}$$

$$\hat{Y}_t^U = Y_{t-1}^U \tag{19}$$

Sometimes, the interval-valued naive forecast will be named as the iRW forecasting method.

3.3. Choosing a Distance Measure With Interval Valued-Data

In order to evaluate the accuracy of a forecast, a measure of the distance between the forecast and the actual value needs to be defined. For classic time series, the absolute or the squared value of the difference between the observation and the forecast is considered as the distance. What remains is to find a definition of distance that can be applied to intervals. Several distance measures for two intervals of the real line have been defined in the literature. (Kao et al., 2014) list the most frequently used ones. We will restrict to the Euclidean distance

$$d_E([x], [y]) = \frac{1}{\sqrt{2}}\sqrt{(x^L - y^L)^2 + (x^U - y^U)^2} \tag{20}$$

Each distance is expressed both in terms of lower and upper bounds, and in terms of centre and radius. Other interesting distances have been proposed in the literature. See, for example, Irpino & Verde (2008).

3.4. Main Methods and Models to Forecast ITS

Following (Arroyo et al., 2011*a*), among others, ITS can be applied to any setting where data can be registered in an almost continuous way using sensors and given a need to track and record the range of values. That is, observing the phenomenon in two frequencies high and low such as days and months, the monthly ITS (of prices, temperatures and so on) arises. In some contexts such as economic ones, the above framework has one limitation. For example, it cannot be used for monthly inflation forecasting, but can be used for quarterly or yearly inflation forecasting.

There are two main classes of methods and models producing interval forecasts for an ITS. See, among others, Arroyo et al. (2011*a*), Han et al. (2008) or Han et al. (2012).

Purely interval methods and models, which can use interval arithmetic to work with interval inputs and transform them into interval-valued outputs. For example, VAR and VECM (see Arroyo et al. 2011*a* and García-Ascanio & Maté 2010), iMLP (see Muñoz et al. 2007), kNN for ITS (ikNN, see Arroyo & Maté 2009), iHolt ((Maia & de Carvalho, 2011)), and so on. For linear regression approaches see, for example, Lima Neto & De Carvalho (2010) and Blanco-Fernández et al. (2011).

Pseudo-interval methods and models, which simulate interval manipulation by modeling two classical time series (one for the centers and one for the radii, or one for the lower bounds and one for the upper bounds) separately, and then combine the respective results into intervals. For example, ARIMA, ARFIMA, MLP, kNN, Holt-Winters, and so on. For details, see, among others, Arroyo et al. (2011*a*).

Table 2: Several approaches to forecast ITS.

Methods based on	Pseudo int. methods	Purely int. methods
Smoothing	Holt-Winters	iHolt
Past values	ARIMA	VAR and VECM
Past patterns	kNN	ikNN
Linear regression	AR	CRM, M, etc.
Neural networks	MLP	iMLP

Table 2 collects different approaches to forecast ITS taking into account the way in that every approach works with the past information of the ITS.

4. Essential Concepts When Analyzing Several Forecasts for One ITS

In this section we consider a realized ITS $\{[x_t]\}$ and its ITS forecast $\{[\hat{x}_t]\}$, with $t = 1, \dots, T$.

4.1. Mean Distance Error

One way to evaluate the accuracy of one ITS forecasting method is computing the distance at every time and then to average these distances. This gives the Mean Distance Error (MDE) to quantify the accuracy of the forecasting method. When using Euclidean distance and following Arroyo et al. (2011*a*), the square of the MDE for the method M with information until time T, will be given by

$$MDE^2 = \frac{1}{2}\frac{\sum_{t=1}^{T}\left[(x_t^L - \widehat{x}_t^L)^2 + (x_t^U - \widehat{x}_t^U)^2\right]}{T}, \tag{21}$$

That is, this MDE is the root mean squared error (RMSE) for evaluating the accuracy of one ITS forecasting method. Clearly, the square value, MDE^2, is the average of the MSE, given by (4), for lower and upper bounds. That is,

$$MDE^2 = \frac{MSE_L + MSE_U}{2} \tag{22}$$

A general definition of the MDE with other distances like Hausdorff, Ichino-Yaguchi, or de Carvalho, is given in Arroyo & Maté (2006).

4.2. Interval Average Relative variance(iARV)

The interval average relative variance (iARV) is expressed in our notation as follows:

$$iARV = \frac{\sum_{t=1}^{T}(x_t^L - \widehat{x}_t^L)^2 + \sum_{t=1}^{T}(x_t^U - \widehat{x}_t^U)^2}{\sum_{t=1}^{T}(x_t^L - \overline{x}^L)^2 + \sum_{t=1}^{T}(x_t^U - \overline{x}^U)^2}, \tag{23}$$

where T denotes the number of fitted intervals, $[\widehat{x}_t] = [\widehat{x}_t^L, \widehat{x}_t^U]$ is the tth fitted interval, $[\overline{x}] = [\overline{x}^L, \overline{x}^U]$ is the sample average interval, and $\overline{x}^L$ and $\overline{x}^U$ are the average of the lower and upper bounds, respectively. The $iARV$ statistic compares the predictions of the model with the predictions given by the average interval of the series. Lower values of $iARV$ therefore mean better forecasts, converging to zero for a perfect forecasting model.

4.3. Interval Theil U Statistcs

For the comparison of different forecasting models or methods to a naïve model, it is customary to use statistics such as the Theil's U statistic. For interval-valued data has been proposed by Maia & de Carvalho (2011). It is given by

$$iU = \sqrt{\frac{\sum_{t=2}^{T}(x_t^L - \widehat{x}_t^L)^2 + \sum_{t=2}^{T}(x_t^U - \widehat{x}_t^U)^2}{\sum_{t=2}^{T}(x_t^L - x_{t-1}^L)^2 + \sum_{t=2}^{T}(x_t^U - x_{t-1}^U)^2}} \tag{24}$$

The iU statistic compares the model with the interval random walk.

- When the value of iU is equal to 1, the performance of the currently observed model is the same as that of the iRW.
- For $iU > 1$, the model performs worse than the iRW.
- For $iU < 1$, the model's performance is better than the iRW.

MDE, iARV, and iUTheil statistics are measures of the size of forecast deviations from the actual values; the lower these values, the better the forecasts.

4.4. Results of the Above Measures for the GE Example

Table 3 reports about different accuracy measures for the forecasting methods M1, M2 and iRW; applied to the GE monthly low-high dataset 2017: January to July.

The best method according iARV, iU, and MDE is M1. However, the second-best depends on the accuracy measure considered. Following iU and iARV is M2 and the worst method is iRW. But taking into account MDE, the second-best is the iRW and the worst method is M2.

Table 3: Accuracy measures of methods M1, M2 and iRW with the GE monthly low-high dataset 2017: January to July.

Method	iARV	iU	MDE
M1	0.051706	0.77067	0.7389013
M2	0.061917	0.84334	1.154411
iRW	0.087056	1	1.005031

5. Coverage and Efficiency Rates When Forecasting ITS

In addition to the error measures discussed above, there are some new measures or rates for evaluating the accuracy of the predicted intervals proposed in Rodrigues & Salish (2015). They are based on very simple concepts, but still, give us valuable information about the accuracy of the methods. The main idea of these two measures is in comparing the intersection of the two ranges (actual and forecast) with each of the individual amplitudes (actual and forecast). Recently, Ramos-Guajardo et al. (2020) apply these concepts to hypothesis tests for analyzing the degree of overlap between the expected value of random intervals.

Coverage rate (CR): Let $w\left([x]_t \cap [\hat{x}]_t\right)$ be the length of the existing intersection between the actual interval and the interval predicted with the method M, and $w\left([x_t]\right)$ be the amplitude of the actual interval. The coverage rate at t is then defined as the ratio between these two lengths and is expressed as follows:

$$R_{c,M}(t) = \frac{w\left([x]_t \cap [\hat{x}]_t\right)}{w\left([x]_t\right)}. \tag{25}$$

The closer the value gets to 100%, the closer the intersection of the two intervals is to the actual interval, which means the predicted interval covers a greater part of the actual interval. However, this measure does not express the possibility of the predicted interval being wider than the actual interval. For example, if [10, 12] is the interval to forecast, the interval [9,13] covers the 100% of that interval but it is worse (less efficient) than [10.3, 12.5] to forecast this interval.

Efficiency rate (ER): Let $w\left([x]_t \cap [\hat{x}]_t\right)$ be the length of the existing intersection between the actual interval and the interval predicted with the method M, and $w\left([\hat{x}]_t\right)$ be the amplitude of the predicted interval. The efficiency rate at t is then defined as the ratio between the two lengths and is expressed as follows:

$$R_{e,M}(t) = \frac{w\left([x]_t \cap [\hat{x}]_t\right)}{w\left([\hat{x}]_t\right)}. \tag{26}$$

The closer the coefficient gets to 100%, the greater portion of the predicted interval intersects with the actual interval. Analogously to the previous case, this measure penalizes the situations when the predicted interval is unnecessarily long compared to the actual interval, but does not quantify situations when the actual interval is longer than the predicted one. For example, if [9, 13] is the interval to forecast, the interval [9.5, 12.5] is more efficient in terms of coverage than [10,12] to forecast that interval.

For a forecast to be reasonably accurate, it is necessary that both measures must be as high as possible. From the equations determining the coverage and efficiency rates, it can be seen that for the predicted interval to be as similar to the actual interval as possible, it needs to both cover the actual interval appropriately and not exceed it by a significant length.

Given a realized and a forecasted ITS with the method M, the Mean Coverage Rate (CR) of the method M to quantify the accuracy of the forecast from the point of view of the coverage of the current ITS will be given by

$$CR_M = \frac{1}{T}\sum_{t=1}^{T} R_{c,M}(t). \tag{27}$$

Similarly, the Mean Efficiency Rate (ER) of the method M to quantify the accuracy of the forecast from the point of view of the efficiency will be given by

$$ER_M = \frac{1}{T}\sum_{t=1}^{T} R_{e,M}(t). \tag{28}$$

We look for methods with the highest CR and ER. See, for example, Buansing et al. (2020).

A global measure of coverage and efficiency will be the Mean Coverage-Efficiency rate (CER) as an average of both measures. That is,

$$CER_M = \frac{1}{2}\big(CR_M + ER_M\big). \tag{29}$$

The validity of this measure is higher (lower) as the two measures are more similar (different).

6. Linear Combination of Interval Forecasts for One ITS

Let $[F_{t,i}]$ the interval forecast of one ITS $[x_t]$ at time t, according to the method i, with $i = 1, \ldots, k$. Let $[F_{t,C}]$ the linear combined forecast with time-varying weights. That is

$$[F_{t,C}] = \sum_{i=1}^{k} \omega_{t,i} * [F_{t,i}] \tag{30}$$

The problem is to get a linear combination forecast $\{[\hat{x}_t]\}$ for a realized ITS $\{[x_t]\}$, with $t = 1, \ldots, T$. The first case to be analyzed, in a parallel approach with Bates & Granger (1969), will be that of two ITS forecasting methods.

6.1. Linear Combination of Two Interval Forecasts for One ITS

We will show how to obtain the optimal combination weights, in the case of constant weights, for a combined forecast with two methods M1 and M2 to forecast one ITS. That is, the problem is to get an optimal linear combination forecast for $[x_t]$, noted by $[\widehat{x}_{t,C}]$, from two forecasting methods for such ITS, $[\widehat{x}_{t,M1}]$ and $[\widehat{x}_{t,M2}]$. Hence,

$$[\widehat{x}_{t,C}] = \omega[\widehat{x}_{t,M1}] + (1-\omega)[\widehat{x}_{t,M2}] \tag{31}$$

The interval obtained with the combined forecasting method C will have the following lower and upper bounds

$$\widehat{x}^L_{t,C} = \omega\widehat{x}^L_{t,M1} + (1-\omega)\widehat{x}^L_{t,M2} \tag{32}$$

$$\widehat{x}^U_{t,C} = \omega\widehat{x}^U_{t,M1} + (1-\omega)\widehat{x}^U_{t,M2} \tag{33}$$

assuming that $\widehat{x}^L_{t,C} \leq \widehat{x}^U_{t,C}$. It can be observed that if $x^L_{t,M1} \leq \widehat{x}^U_{t,M1}$ and $x^L_{t,M2} \leq \widehat{x}^U_{t,M2}$ and only when $\omega \geq 0$, that condition is satisfied.

We search the ω value giving the minimum MDE, iARV, or iU.

6.1.1. Under the Mean Distance Error

We consider the mean distance error (MDE) as the accuracy measure. That is, a method M1 has the following MDE square

$$(MDE_{M1})^2 = \frac{1}{2}\frac{\sum_{t=1}^{T}\left[(x^L_t - \widehat{x}^L_{t,M1})^2 + (x^U_t - \widehat{x}^U_{t,M1})^2\right]}{T} \tag{34}$$

In the same way, a method M2 has the following MDE square

$$(MDE_{M2})^2 = \frac{1}{2}\frac{\sum_{t=1}^{T}\left[(x^L_t - \widehat{x}^L_{t,M2})^2 + (x^U_t - \widehat{x}^U_{t,M2})^2\right]}{T} \tag{35}$$

The linear combined forecast, method C, will have the following MDE square, $(MDE_C)^2$

$$\frac{1}{2}\frac{\sum_{t=1}^{T}\left[(x^L_t - (\omega\widehat{x}^L_{t,M1} + (1-\omega)\widehat{x}^L_{t,M2}))^2 + (x^U_t - (\omega\widehat{x}^U_{t,M2} + (1-\omega)\widehat{x}^U_{t,M2}))^2\right]}{T} \tag{36}$$

After some computations, using the standard procedure to minimize a function, the optimal weight is

$$\omega = \frac{MDE^2_{M2} - ACMDE_{M1,M2}}{MDE^2_{M1} + MDE^2_{M2} - 2 * ACMDE_{M1,M2}} \tag{37}$$

where $ACMDE_{M1,M2}$ stands for the average crossed error of the mean distance using methods M1 and M2 and is given by

$$ACMDE_{M1,M2} = \frac{\sum_{t=1}^{T}\ [(x^L_t - \widehat{x}^L_{t,M1})(x^L_t - \widehat{x}^L_{t,M2}) + (x^U_t - \widehat{x}^U_{t,M1})(x^U_t - \widehat{x}^U_{t,M2})]}{2T} \tag{38}$$

Analysis of the optimal weights using the MDE.

- Both methods are perfect ($MDE_{M1} = MDE_{M2} = 0$). Then we have the equal weights (EW) case ($\omega = 0.5$ and $1 - \omega = 0.5$).
- Both methods are similar ($MDE_{M1} \simeq MDE_{M2}$). Then we also have the equal weights (EW) case ($\omega = 0.5$ and $1 - \omega = 0.5$).
- As one method (M2) is some better ($MDE_{M1} \simeq 2*MDE_{M2}$) or much better ($MDE_{M1} \simeq 4 * MDE_{M2}$) than the other one (M1), the weight assigned to the best method is greater or much greater than the other weight.
- In the case of crisp time series (both limits of the interval are the same at every time) and two crisp forecasting methods are combined, the optimal weights are the same than the ones given in Bates & Granger (1969), the seminal paper about forecasts combination.

6.1.2. Under the iARV

The following relationship between iARV and MDE

$$iARV = \frac{2T * (MDE)^2}{\sum_{t=1}^{T} (x_t^L - \overline{x}^L)^2 + \sum_{t=1}^{T} (x_t^U - \overline{x}^U)^2}. \tag{39}$$

justifies that the above developments remain with this measure. The optimal weight is given by (37).

6.1.3. Under the Theil's iU

Now, we consider the Theil's iU as the accuracy measure. In similar reasoning to the above subsection, the optimal weight is given by

$$\omega = \frac{iU_{M2}^2 - ACEiU_{M1,M2}}{iU_{M2}^2 + iU_{M1}^2 - 2 * ACEiU_{M1,M2}} \tag{40}$$

where $ACEiU_{M1,M2}$ stands for the average crossed error in iU computation using methods M1 and M2 and is given by

$$ACEiU_{M1,M2} = \frac{\sum_{t=2}^{T} [(x_t^L - \widehat{x}_{t,M1}^L)(x_t^L - \widehat{x}_{t,M2}^L) + (x_t^U - \widehat{x}_{t,M1}^U)(x_t^U - \widehat{x}_{t,M2}^U)]}{\sum_{t=2}^{T} [(x_t^L - x_{t-1}^L)^2 + (x_t^U - x_{t-1}^U)^2]} \tag{41}$$

In the case of an ITS which is a CTS, due to at every time the upper limit is equal to the lower limit, (40) is (10) and (41) is (11).

Analysis of the optimal weights using the iU statistic.

- Both methods are perfect (iUM1 = iUM2 = 0). Then we have the equal weights (EW) case ($\omega = 0.5$ and $1 - \omega = 0.5$).

- Both methods are similar ($iUM1 \simeq iUM2$). Then we also have the equal weights (EW) case ($\omega = 0.5$ and $1 - \omega = 0.5$).

- As one method (M2) is some better ($iUM1 \simeq 2 * iUM2$) or much better ($iUM1 \simeq 4 * iUM2$) than the other one (M1), the weight assigned to the best method is greater or much greater than the other weight.

6.1.4. When to Combine or Not Two ITS Forecasts for One ITS

Reasons given in Hendry & Clements (2004) to pool crisp forecasts remain in an ITS context. These are the main reasons.

- If two models or methods provide partial, but incompletely overlapping, explanations, then some combination of the two might do better than either alone. In particular, if two forecasts are biased (one upwards, one downwards), it is easy to see why combining could be an improvement over either. That is, if two methods or models forecast reasonably well one ITS; the best decision-making is to combine. The combined method will get better performance than every single method both using MDE or Theil's iU criteria. However, sometimes optimal weights obtained with the Theil's iU criterion will give a better method in terms of higher CR and ER values and lower iU measurements.

- In non-stationary interval time series in centers or centers and radii, most forecasts will fail in the same direction when forecasting over a period within which a break unexpectedly occurs. The combination is unlikely to provide a substantial improvement over the best individual forecasts in such a setting. However, what will occur when forecasting after a deterministic shift depends on the extent of model mis-specification, data correlations, the size of breaks, and so on; so combination may help.

- If one method owns quite good accuracy measures forecasting very well one ITS and the other method not, then the best decision-making is not to combine.

This decision also depends on the following issues, in decreasing order of more to less importance.

1. The two methods under consideration.

2. The dataset under study.

3. The accuracy measure for which the optimal weights will be obtained.

4. The accuracy measures considered to evaluate the performance of the optimal combination against every single method.

6.2. Linear Combination of Several Interval Forecasts for One ITS

We will show how to obtain combination weights, in the case of varying weights, for a combined forecast with k methods, M1, M2,... and Mk to forecast one ITS $[x_t]$.

Let $[\widehat{x}_{t,Mi}]$ the interval forecast of one ITS $[x_t]$, at time t according to the method i, with $i = 1, \ldots, k$. Let $[\widehat{x}_{t,Ck}]$ the linear combined forecast with time-varying weights. That is

$$[\widehat{x}_{t,Ck}] = \sum_{i=1}^{k} \omega_{t,i} * [\widehat{x}_{t,Mi}]. \tag{42}$$

The resulting interval time series as one combined forecast will have the following lower and upper bounds

$$\widehat{x}^L_{t,Ck} = \sum_{i=1}^{k} \omega_{t,i} * \widehat{x}^L_{t,Mi} \tag{43}$$

and

$$\widehat{x}^U_{t,Ck} = \sum_{i=1}^{k} \omega_{t,i} * \widehat{x}^U_{t,Mi} \tag{44}$$

assuming that $\widehat{x}^L_{t,Ck} \leq \widehat{x}^U_{t,Ck}$. It can be observed that this condition will be satisfied if $\widehat{x}^L_{t,Mi} \leq \widehat{x}^U_{t,Mi}$, for $i = 1, \ldots, k$; and $\omega_{t,i} \geq 0$.

6.2.1. Trimming and Other Statistical Combinations

- Equal weights (EW) combination.

  In this case, all methods are averaged with the same weight. That is,

  $$\omega_{t,i} = \frac{1}{k}, \tag{45}$$

  for $i = 1, \ldots, k$ and at every time t. This procedure does not require the validation of the scheme considering a training period and a testing period.

- Trimmed mean weights (TrMeanW) combination. The concept of trimmed mean is not defined for interval-valued data because there is not a complete order in the space $\mathcal{K}_c(\mathbb{R}) = \{[a, b] : a, b \in \mathbb{R}, a \leq b\}$. One alternative is ranking methods according to one accuracy measure (for example, iU) and to discard the worst method(s).

- Median combination. The median of a set of intervals has several definitions. See, for example, Sinova et al. (2010).

6.2.2. Performance-Based Combinations

These approaches assign higher weights to forecasts with relatively good forecasting track records and lower weights to forecasts with poor performances. Following the ideas given in the crisp case, when using ITS forecasts, the following combinations schemes can be proposed.

- Inverse of Mean Distance Error (IMDE)

  Weights are obtained according to

$$\omega_{t,i} = \frac{MDE_{t,i}^{-1}}{\sum\limits_{i=1}^{k} MDE_{t,i}^{-1}} \tag{46}$$

  where $MDE_{t,i}$ is computed by (21).

- Inverse of Theilt́s U Squared (IThUS) averaging

  Weights are obtained through the following expressions

$$\omega_{t,i} = \frac{(U^2)_{t,i}^{-1}}{\sum\limits_{i=1}^{k} (U^2)_{t,i}^{-1}}. \tag{47}$$

  The better (worst) the method is, the higher (lower) the corresponding weight for such method is.

6.2.3. Sequential Combination

Following Winkler & Clemen (1992), a sequential combination can be accomplished. Instead of combining all k forecasting methods at once, we could combine them sequentially in $k-1$ steps, beginning with the two worst and adding one forecast at each step we always face a problem of combining two interval forecasts for one ITS and to apply the results in Section 6.1. We suggest the iU sequential algorithm where the weighting scheme is that obtained in Section 6.1.3. See the detailed example in the next section.

7. Examples

7.1. Combining two Monthly General Electric (GE) Low-High Price Forecasts in 2017 (Semesters 1 and 2)

Two forecasting methods M1 and M2 (iRW) are considered to forecast the ITS of monthly low-high prices of GE. Method M1 considers as interval forecast for the next month the moving average of the last five months diminished in two dollars.

That is,

$$[\widehat{x}_{t,M1}] = \frac{\sum_{j=1}^{5}[x_{t-j}]}{5} - 2. \tag{48}$$

Obviously, the simple iRW model considers

$$[\widehat{x}_{t,iRW}] = [x_{t-1}]. \tag{49}$$

Figure 2 shows the problem graphically. The centers' line of the ITS, displayed by one solid line (red), reports about the evolution of the centers of the monthly low-high prices of General Electric (GE) during the year 2017. The couple of ITS whose centers are shown by one dot line (purple) and one dash line (blue) informs about the evolution of the centers of the monthly interval price forecasts of GE with the methods M1 and M2(iRW), respectively, during the same period.

In 2017-Semester 1, methods M1 and iRW will be considered to decide about to combine them or not. The OW forecast combination under iU is obtained as

$$[\widehat{x}_{t,OWiU}] = 0.6798 * [\widehat{x}_{t,M1}] + 0.3202 * [\widehat{x}_{t,iRW}]. \tag{50}$$

The OW forecast combination under MDE is obtained as

$$[\widehat{x}_{t,OWMDE}] = 0.7002 * [\widehat{x}_{t,M1}] + 0.2998 * [\widehat{x}_{t,iRW}]. \tag{51}$$

Accuracy measures for each method and every combination are shown in Table 4. All evaluation measures are much better in both combination approaches, clearly outperforming every single method. The iU (MDE) optimal weight approach is (is not) the best in terms of iU, but not is (is) the best according to iARV, MDE, CR, and ER values.

EW combination outperforms both single methods in iARV, MDE, and iU accuracy measures; but it is worst than the best single method (iRW) both in CR and ER values. In this case, EW is worse than the three optimal combinations developed in this paper.

Table 4: Accuracy measures in the case of forecasting 2017-S1 for the monthly low-high prices of GE with two forecasting methods, the iU and MDE OW combinations, and EW combination.

Forecasting method	iARV	MDE	iU	CR	ER
M1 (MA5-2)	2.3440	1.4899	1.5602	0.2898	0.3775
M2 (iRW)	0.7288	0.8308	1	0.6768	0.7121
M3 (OW-iU comb.)	0.3679	0.5902	0.7688	0.7306	0.7462
M4 (OW-MDE comb.)	0.3662	0.5889	0.7699	0.7352	0.7512
M5 (EW comb.)	0.5278	0.7070	0.8485	0.6589	0.6681

Figure 2: GE monthly ITS prices in 2017 (Semesters 1 and 2) with two forecasting methods M1 and M2(iRW).

Concerning 2017-Semester 2, we evaluate if methods M1 and M2 (iRW) should be combined or not. In this case, the MDE-iARV approach and the OW forecast combination under iU do not apply due both give negative weights. Hence, the EW combination can be obtained as

$$[\widehat{x}_{t,EW}] = 0.5 * [\widehat{x}_{t,M1}] + 0.5 * [\widehat{x}_{t,iRW}] \tag{52}$$

TABLE 5: Accuracy measures in the case of forecasting 2017-S2 for the monthly low-high prices of GE with two forecasting methods and the EW combination.

Forecasting method	iARV	MDE	iU	CR	ER
M1	0.6786	1.4899	1.3204	0.2702	0.3432
M2 (iRW)	0.4310	0.8308	1	0.4423	0.3976
M3 (EW comb.)	0.5180	2.3104	1.1348	0.2722	0.3283

Accuracy measures for each method and the EW combination are shown in Table 5. All evaluation measures are much worst in the combination approach than in the iRW method. Therefore, the decision-making in 2017-S2 is not combining, discarding the method M1. Figure 2 is telling us that both methods are overforecasting the ITS but iRW dominates M1 which is the typical case where the combination approach will give worst results than the best single method.

7.2. Combining Five ITS Forecasts for the Monthly SP500 in 2017 and 2018

This last example compares the low-high monthly evolution of the SP500 financial index of the US stocks market during 2017 and 2018, against several weighting schemes of five forecasting methods.

One method is the iRW model and the other four methods are based on different approaches of linear regression methods for interval-valued data such as the center method (CM), the minMax method (mMM), the center and radius method (CRM) and the M model. See Lima Neto & De Carvalho (2010) for CM, mMM and CRM; and Blanco-Fernández et al. (2011) for the M model. These methods and models are fitted using the SP500 information in the period 2004-2016 [1].

7.2.1. Weighting scheme 1: averaging based on equal weights (EW)

Equal weights (EW) forecasting method combines the five methods with the same weights, that is, 0.2 in this case. Hence

$$[\widehat{SP500}_t] = 0.2 * ([\widehat{x}_{t,iRW}] + [\widehat{x}_{t,mMM}] + [\widehat{x}_{t,CM}] + [\widehat{x}_{t,CRM}] + [\widehat{x}_{t,M}]) \tag{53}$$

Figure 3 shows the ITS of the low-high monthly evolution of the SP500 during 2017 and the ITS of the average of the five forecasting methods, named EW combination.

[1] models and dataset information can be obtained upon request

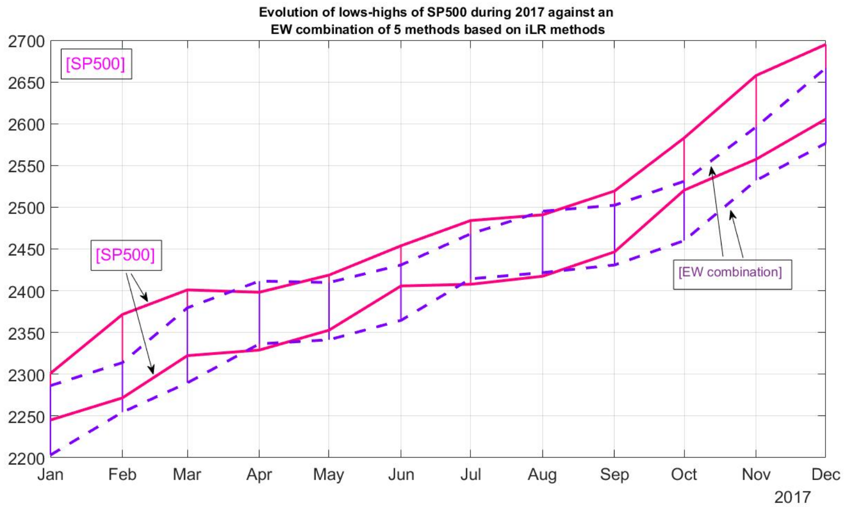

FIGURE 3: SP500 monthly ITS in 2017 with simple average of five forecasting methods.

7.2.2. Weighting Scheme 2: Averaging Based on Weights Under iU

This second weighting scheme is based on the iU approach developed in Section 6.2.2 with the name IThUS. Table 6 shows weights according to this approach (and also the MDE approach).

TABLE 6: Several weighting schemes in the case of forecasting 2017 and 2018 for the monthly low-high of the SP500 with five methods.

Forecasting method	Name	Weight MDE	Weight iU	EW	iU-Seq. Weig.
M1	iRW	0.1513	0.1139	0.2	0.0625
M2	mMM	0.1958	0.1914	0.2	0.062525
M3	CM	0.1974	0.1944	0.2	0.125025
M4	CRM	0.2018	0.2034	0.2	0.25005
M5	M	0.2537	0.2969	0.2	0.4999

7.2.3. Sequential Weighting Based on the iU Sequential Algorithm

In this combination we follow the approach given in Winkler & Clemen (1992). First we take the two worst methods, iRW and mMM and we proceed to get the optimal weight according to the iU approach. We obtain the following combination in the Step1 of the algorithm.

$$[\widehat{SP500}_{t,Step1}] = 0.5001 * [\widehat{x}_{t,mMM}] + 0.4999 * [\widehat{x}_{t,iRW}]. \quad (54)$$

In the Step2 we combine the above method with the CM method and we obtain the following combination.

$$[\widehat{SP500}_{t,Step2}] = 0.5000 * [\widehat{SP500}_{t,Step1}] + 0.5000 * [\widehat{x}_{t,CM}]. \quad (55)$$

That is,

$$[\widehat{SP500}_{t,Step2}] = 0.24995 * [\widehat{x}_{t,iRW}] + 0.25005 * [\widehat{x}_{t,mM}] + 0.5000 * [\widehat{x}_{t,CM}] \quad (56)$$

In the Step3 we combine the above forecasting method of Step2 with the CRM method and we obtain the following combination.

$$[\widehat{SP500}_{t,Step3}] = 0.5000 * [\widehat{SP500}_{t,Step2}] + 0.5000 * [\widehat{x}_{t,CRM}]. \quad (57)$$

Hence,

$$\begin{aligned}[\widehat{SP500}_{t,Step3}] = 0.12497 * [\widehat{x}_{t,iRW}] + 0.12502 * [\widehat{x}_{t,mM}] \\ + 0.25 * [\widehat{x}_{t,CM}] + 0.5 * [\widehat{x}_{t,CRM}]\end{aligned} \quad (58)$$

In the Step4 we combine the above forecasting method of Step3 with the M model and we obtain the following combination.

$$[\widehat{SP500}_{t,Step4}] = 0.5001 * [\widehat{SP500}_{t,Step3}] + 0.4999 * [\widehat{x}_{t,M}] \quad (59)$$

Therefore $[\widehat{SP500}_{t,Step4}]$ will be obtained as

$$\begin{aligned}0.0625 * [\widehat{x}_{t,iRW}] + 0.06252 * [\widehat{x}_{t,mMM}] + 0.12502 * [\widehat{x}_{t,CM}] \\ + 0.25005 * [\widehat{x}_{t,CRM}] + 0.4999 * [\widehat{x}_{t,M}]\end{aligned} \quad (60)$$

7.2.4. Analysis of the Three Weighting Schemes

Table 7 shows accuracy measures for single methods and the three weighting schemes in 2017. Methods are ranked in the same order according to MDE and iU. EW combination shows balanced MDE, iU, CR, and ER values in the middle of the range of the corresponding values of the five methods.

Table 7: Accuracy measures in the case of forecasting 2017 for the monthly low-high of the SP500 with several weighting schemes

Forecasting method	Name	MDE	iU	CR	ER
M1	iRW	41.3025	1	0.5347	0.53955
M2	mMM	31.9091	0.7715	0.66927	0.66001
M3	CM	31.6428	0.7655	0.66595	0.66423
M4	CRM	30.9652	0.7484	0.72348	0.65409
M5	M	24.6263	0.6194	0.61426	0.82609
Simple av. M1 to M5	EW	31.002	0.7544	0.65197	0.66946
iU-Weig. av. M1 to M5	iUW	29.2438	0.7144	0.6584	0.6925
iUSeq-Weig. Av. M1 to M5	SeqCW	29.8762	0.7428	0.5054	0.7715

Table 8 informs about the accuracy measures in the year 2018. In this case, EW combination is a bad approach due to the negative effect of the M model in the combination. Now, EW combination does not report balanced MDE, iU, CR, and ER values in the middle of the range of the corresponding values of the five methods.

Looking at the performance of the methods in 2017, there is the temptation of eliminating the iRW method due to is the worst one. However, 2017 is a year with an uptrend in the SP500 but 2018 is with sideways and downtrend. Analyzing the behavior of the methods in 2018, the temptation would be eliminating the M model due to is the worst one. In this case, forecasting equations with the five methods have not been modified at the beginning of 2018 incorporating the information of 2017. However, as 2017 remain with an uptrend in line with the period 2004-2016, it would be quite difficult to catch better the real behavior of the [SP500]. One solution to this problem is to have a rolling window with a short period of time (say 6 months in this case) to get the equations of the methods.

Table 8: Accuracy measures in the case of forecasting 2018 for the monthly low-high of the SP500 with several weighting schemes

Forecasting method	Name	MDE	iU	CR	ER
M1	iRW	107.1681	1	0.5871	0.6271
M2	mMM	108.8802	1.03117	0.6281	0.6420
M3	CM	109.0247	1.0334	0.6263	0.6432
M4	CRM	109.7366	1.0407	0.5118	0.7018
M5	M	137.9044	1.3407	0.25	0.7029
Simple av. (M1 to M5)	EW	110.4870	1.0478	0.5453	0.6806
iU-Weig. Av, M1 to M5	iUW	112.8250	1.0742	0.5120	0.6912
iUSeq-Weig. Av. M1 to M5	SeqCW	129.3678	1.2448	0.2660	0.7304

The conclusion is that when averaging several forecasting methods for one ITS, we usually obtain like a smoothed method with accuracy measures in the middle of the range from the worst one to the best one. But that order is with the dataset of the past, new data can modify that ranking which is happening in the case of the SP500 in 2018. See Figure 4 where it can be observed sideways behavior during the first three quarters of 2018 and downtrend in the last quarter.

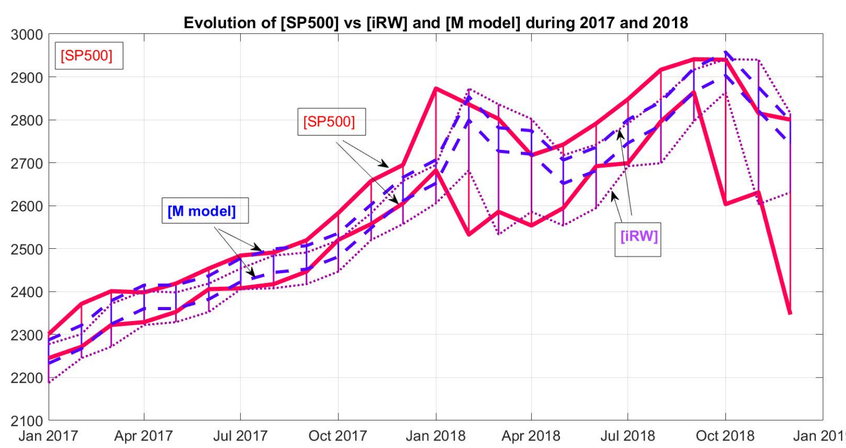

Figure 4: SP500 monthly ITS in 2017 and 2018 versus iRW and model M.

Figure 5 shows that the iU combination reduces the volatility problem of the M model and maintains some of the essences of the iRW model.

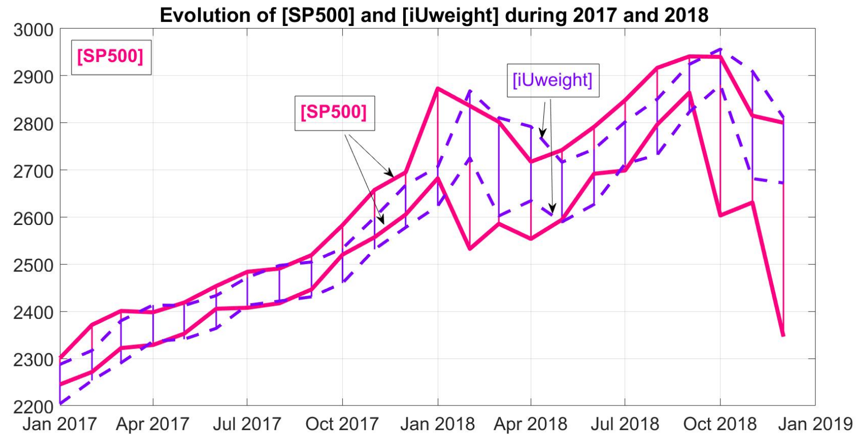

FIGURE 5: SP500 monthly ITS in 2017 and 2018 versus the combination with iU weights.

7.2.5. Comparing Predictive Accuracy of the Three Weighting Schemes

Diebold & Mariano (1995) propose and evaluate explicit tests of the null hypothesis of no difference in the accuracy of two competing forecasts in the so-called Diebold-Mariano (DM) test. In the framework of ITS has been used, for example, in Arroyo et al. (2011*b*) or Buansing et al. (2020). In the context of classic forecasts combination, there are a lot of references using the DM test. For example, (den Butter & Jansen, 2013) compare different approaches to assess the performance of several long-term interest rate forecast approaches, including some combinations such as EW and OW, where the RW is the benchmark.

In this example, we find a case of competing combination schemes whose predictive accuracy can be analyzed through the DM test. We carry out this test with the three competing combination schemes considered in Table 7 (2017) and Table 8 (2018), both for lows and highs.

TABLE 9: Comparing forecasts according to the statistic value (P-value) of the DM test in the case of forecasting 2017 and 2018 for the monthly low-high of the SP500 with three weighting schemes

Year Competing forecasts	EW vs iUW	EW vs SeqCW	iUW vs SeqCW
2017 (lows)	3.0676 (0.0107)	-2.1355 (0.0560)	-1.8839 (0.0863)
2017 (highs)	2.7605 (0.0185)	3.7564 (0.0032)	3.9067 (0.0024)
2018 (lows)	-1.6849 (0.1201)	2.2523 (0.0457)	2.3632 (0.0376)
2018 (highs)	1.5865 (0.1409)	-0.1235 (0.9039)	0.0662 (0.9484)

Table 9 collects results for the DM test. Positive (negative) values indicate the first (second) combination scheme is less (more) accurate than the second (first) one. We conclude that in 2017 the three approaches are significantly different, at 10% for lows and 2% for highs, being, in this case, the winner approach iUW. However, in 2008 these three combination schemes are not significantly different at any usual percentage for highs and, at 5%, the SeqCW combination is the best way to proceed with lows.

8. Discussion and Research Agenda

The combination of several ITS forecasts is a new research field in statistical forecasting which contains a lot of challenges. Until our knowledge goes, this is the first paper in this field and provides several solutions and guidelines to the problem of combining two or several ITS forecasts with constant or varying weights.

8.1. Guidelines to Combine Several ITS Forecasts

In the classic context of combining several forecasts for one time series, various papers provide guidance such as de Menezes et al. (2000), or simple explanations of the forecast combination puzzle such as Smith & Wallis (2009). According to the above knowledge and the research developed in this paper, at least the following guidelines can be proposed.

1. EW combination of several ITS forecasts is one approach to consider when the involved methods are not quite different among them. As a result, EW combination shows balanced MDE, iU, CR, and ER values in the middle of the range of the corresponding values of the involved methods. EW combination is preferred to another combination scheme when the corresponding ITS mixes periods of uptrend, downtrend, and sideways.

2. EW combination of several ITS forecasts is one approach to avoid when the involved methods can show important differences between them in terms of centers, radii or both. As a result, the EW combination can show MDE, iU, CR, and ER values far from the middle of the range of the corresponding values of the involved methods.

3. Weighting schemes based on iU Theil values developed in this paper can outperform EW combination when the corresponding ITS shows uptrend or downtrend.

4. Sequential weighting scheme based on iU Theil values developed in this paper needs to be monitored in shorter periods of time than EW or iU combination schemes.

8.2. Research Agenda

ITS is an alternative way to crisp or classic time series to analyze the time evolution of key meteorological variables such as temperatures, wind speed, and so on; or key financial variables such as stock prices, exchange rates, commodity prices, and so on.

There are several methods to forecast ITS and the combination of several forecasts gives the impression is the best approach.

It seems the forecast combination puzzle remains in some way in this framework. Taking into account the advances in interval-valued data analysis methods, we can suggest the following research agenda.

- The problem of optimal weights with k ITS forecasts.

  The optimal weights procedure developed in this paper for two methods M1 and M2 needs to be extended to three or, in general, k forecasts obtained by different methods or models M1, M2,..., Mk used with an ITS. This is not a direct or trivial extension due to there are several equations for the covariance matrix with interval-valued data. See, for example, Le-Rademacher & Billard (2012) from the symbolic data analysis approach and Sinova et al. (2012) from the random interval approach.

- The forecast combination puzzle with k ITS forecasts.

  Complicated weighting schemes obtained in the above sections need to be compared with the simple average. Additional empirical practical cases need to be run to evaluate the effect of the ITS combined forecast in coverage and efficiency rates and other accuracy measures.

- The linear regression way.

  Linear regression methods for interval valued-data have acquired maturity in the scientific community in the last 10 years and new developments are coming every year. Hence, the analysis of OLS and CLS methods to combine forecasts obtained through several ITS methods suggests a clear way, taking as a benchmark the arithmetic average with EW proposed in this paper.

- The principal component analysis (PCA) way.

  PCA methods for interval valued-data have acquired maturity in the scientific community in the last 15 years. Hence, to combine forecasts obtained through several ITS methods using a weighting scheme based on PCA sounds interesting, taking again as a benchmark the arithmetic average with EW.

- The Bayesian way.

  Bayesian model averaging (BMA) for interval-valued data was firstly introduced in Maté (2012). To develop a BMA scheme for forecasts combination of ITS forecasts could be an interesting research proposal.

  The Bayesian approach is particularly adequate when there are a lot of models and a hierarchy can be introduced into the set of models. Hence, a hierarchical Bayesian scheme to combine ITS seems a promising research proposal. Zhang et al. (2015) propose to analyze interval data from a Bayesian point of view. This paper could give some advice about additional alternatives on how to use Bayesian methods in the problem of combining several forecasts obtained from an ITS.

- The machine learning approach.

  The use of artificial neural networks (ANN) has been proposed by Aladag et al. (2010) to provide a new forecast combination approach. Development of procedures to forecast ITS with ANN is being quite succesful (see, as an example, the multi-layer perceptron for interval-valued data (iMLP) proposed in Muñoz et al. (2007)). Therefore, combining several ITS forecasts through different ANN frameworks seems quite promising. For example, Adhikari & Agrawal (2014) could be extended to interval-valued data using the models iMLP and iRW.

- The Big Data paradigm. Chen et al. (2016) state, "the big data paradigm needs theories to guide its development". As part of that development, procedures to combine information in big data systems are key. This paper puts the first stone in that direction providing different alternatives to combine several forecasts for one ITS and opens research avenues in large datasets contexts, some of them connected to the above points.

9. Conclusions

Three original approaches to combine linearly two ITS forecasting methods have been presented. The first one using the MDE with Euclidean distance, the second one considering the iARV statistics and the third one evaluating the U statistics for ITS (iU). Both approaches provide weights with a similar structure and they behave as an extension of the optimal weights obtained for two crisp forecasts in the seminal paper of Bates & Granger (1969). The iU approach is new in the combination of several forecasts of one time series (classic and interval-valued) and has the advantage of discarding some combination schemes or ranking several combination approaches. The optimal weights (OW) procedures developed in this paper for two methods M1 and M2 have been analyzed in one example with two periods of time, one where OW outperform EW and the other where both OW do not apply. These procedures need to be extended to 3 or, in general, k forecasts obtained by different methods or models M1, M2,. . . , Mk used with one ITS.

Linear combination of several forecasting methods for one ITS has been addressed through statistical, performance-based, and sequential combinations weighting schemes. The way to proceed is shown in one example concerning forecasting the monthly low-high price of the SP500 in different scenarios of uptrend, downtrend, and sideways which is the case of years 2017 and 2018. It seems the forecast combination puzzle remains in some way in the case of forecasting ITS with several methods. As part of future developments diverse ITS forecasting models and methods as well as ITS data sets from different fields of knowledge such as economics, finance, energy, health, tourism, weather, and so on; need to be considered. This paper has also traced an agenda for further research on combining several forecasts for ITS. Some pages of that agenda are: the problem of optimal weights with k ITS forecasts, the forecast combination puzzle with k ITS forecasts, the linear regression avenue, the principal component analysis way, the Bayesian framework, the machine learning route (mainly with neural networks) and the big data paradigm.

Acknowledgements

A preliminary draft of this paper was an invited talk at the Conference SIDM2015 (1st International Symposium on Interval Data Modelling: Theory and Applications) celebrated in Beijing July, 3 and 4 and organized by the Academy of Mathematics and Systems Science (AMSS) and the Chinese Academy of Sciences (CAS). We thank the invitation to the organizers of SIDM2015: Professor Shouyang Wang, Professor Yongmiao Hong, Professor Ai Han, and Professor Haiqiang Chen. And we also thank the comments of participants. Special acknowledgment to Editor Professor Ramón Giraldo and the three reviewers for all comments and suggestions.

[Received: February 2020 — Accepted: December 2020]

References

Adhikari, R. & Agrawal, R. (2014), 'A combination of artificial neural network and random walk models for financial time series forecasting', *Neural Computing and Applications* **24**(6), 1441–1449.

Aladag, C. H., Egrioglu, E. & Yolcu, U. (2010), 'Forecast combination by using artificial neural networks', *Neural Processing Letters* **32**(3), 269–276.

Armstrong, J. S. (2001), *Combined forecasts. In Principles of Forecasting: A Handbook for Researchers and Practitioners*, Kluwer Academic Publishers, Norwell, MA.

Arroyo, J., Espínola, R. & Maté, C. (2011*a*), 'Different approaches to forecast interval time series: a comparison in finance', *Computational Economics* **37**(2), 169–191.

Arroyo, J., Espínola, R. & Maté, C. (2011*b*), 'Different approaches to forecast interval time series: a comparison in finance', *Computational Economics* **37**(2), 169–191.

Arroyo, J., González-Rivera, G. & Maté, C. (2010), 'Forecasting with interval and histogram data. some financial applications', *Handbook of empirical economics and finance* pp. 247–280.

Arroyo, J. & Maté, C. (2006), Introducing interval time series: accuracy measures, *in* 'Compstat, proceedings in computational statistics', Heidelberg: Physica-Verlag, pp. 1139–1146.

Arroyo, J. & Maté, C. (2009), 'Forecasting histogram time series with k-nearest neighbours methods', *International Journal of Forecasting* **25**(1), 192–07.

Atiya, A. F. (2020), 'Why does forecast combination work so well?', *International Journal of Forecasting* **36**(1), 197–200.

Avci, E., Ketter, W. & van Heck, E. (2018), 'Managing electricity price modeling risk via ensemble forecasting: The case of turkey', *Energy Policy* **123**, 390–403.

Bassetti, F., Casarin, R. & Ravazzolo, F. (2020), Density forecasting, *in* 'Macroeconomic Forecasting in the Era of Big Data', Springer, pp. 465–494.

Bates, J. M. & Granger, C. W. (1969), 'The combination of forecasts', *Journal of the Operational Research Society* **20**(4), 451–468.

Billard, L. & Diday, E. (2003), 'From the statistics of data to the statistics of knowledge: Symbolic data analysis', *Journal of the American Statistical Association* **98**(462), 470–487.

Billard, L. & Diday, E. (2006), *Symbolic Data Analysis: Conceptual Statistics and Data Mining John Wiley*, Chichester.

Blanco-Fernández, A., Corral, N. & González-Rodríguez, G. (2011), 'Estimation of a flexible simple linear model for interval data based on set arithmetic', *Computational Statistics & Data Analysis* **55**(9), 2568–2578.

Buansing, T. T., Golan, A. & Ullah, A. (2020), 'An information-theoretic approach for forecasting interval-valued sp500 daily returns', *International Journal of Forecasting* **36**, 800–813.

Chen, Y., Chen, H., Gorkhali, A., Lu, Y., Ma, Y. & Li, L. (2016), 'Big data analytics and big data science: a survey', *Journal of Management Analytics* **3**(1), 1–42.

Clemen, R. T. (1989), 'Combining forecasts: A review and annotated bibliography', *International Journal of Forecasting* **5**(4), 559–583.

de Menezes, L., Bunn, D. & Taylor, J. (2000), 'Review of guidelines for the use of combined forecasts', *European Journal of Operational Research* **120**(1), 190–204.

den Butter, F. A. & Jansen, P. W. (2013), 'Beating the random walk: a performance assessment of long-term interest rate forecasts', *Applied Financial Economics* **23**(9), 749–765.

Dickinson, J. (1975), 'Some comments on the combination of forecasts', *Journal of the Operational Research Society* **26**(1), 205–210.

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

Fama, E. F. (1995), 'Random walks in stock market prices', *Financial Analysts Journal* **51**(1), 75–80.

Fama, E. F. & Blume, M. E. (1966), 'Filter rules and stock-market trading', *The Journal of Business* **39**(1), 226–41.

Galicia, A., Talavera-Llames, R., Troncoso, A., Koprinska, I. & Martínez-Álvarez, F. (2019), 'Multi-step forecasting for big data time series based on ensemble learning', *Knowledge-Based Systems* **163**, 830–841.

Gao, Y., Shang, H. L. & Yang, Y. (2019), 'High-dimensional functional time series forecasting: An application to age-specific mortality rates', *Journal of Multivariate Analysis* **170**, 232–43.

García-Ascanio, C. & Maté, C. (2010), 'Electric power demand forecasting using interval time series: A comparison between var and imlp', *Energy Policy* **38**(2), 715–725.

Genre, V., Kenny, G., Meyler, A. & Timmermann, A. (2013), 'Combining expert forecasts: Can anything beat the simple average?', *International Journal of Forecasting* **29**(1), 108–121.

Gibbs, C. G. (2017), 'Forecast combination, non-linear dynamics, and the macroeconomy', *Economic Theory* **63**(3), 653–686.

Glennon, D., Kiefer, H. & Mayock, T. (2018), 'Measurement error in residential property valuation: An application of forecast combination', *Journal of Housing Economics* **41**, 1–9.

Gneiting, T. (2011), 'Making and evaluating point forecasts', *Journal of the American Statistical Association* **106**(494), 746–762.

Gneiting, T. & Katzfuss, M. (2014), 'Probabilistic forecasting', *Annual Review of Statistics and Its Application* **1**, 125–151.

Hallman, J. & Kamstra, M. (1989), 'Combining algorithms based on robust estimation techniques and co-integrating restrictions', *Journal of Forecasting* **8**(3), 189–198.

Han, A., Hong, Y., Lai, K. & Wang, S. (2008), 'Interval time series analysis with an application to the sterling-dollar exchange rate', *Journal of Systems Science and Complexity* **21**(4), 550–565.

Han, A., Lai, K. K., Wang, S. & Xu, S. (2012), 'An interval method for studying the relationship between the australian dollar exchange rate and the gold price', *Journal of Systems Science and Complexity* **25**(1), 121–132. http://link.springer.com/article/10.1007/s11424-012-8116-x

Hendry, D. F. & Clements, M. P. (2004), 'Pooling of forecasts', *The Econometrics Journal* **7**(1), 1–31.

Hsu, H.-L. & Wu, B. (2008), 'Evaluating forecasting performance for interval data', *Computers & Mathematics with Applications* **56**(9), 2155–163.

Hyndman, R. J. & Koehler, A. B. (2006), 'Another look at measures of forecast accuracy', *International journal of forecasting* **22**(4), 679–688.

Hyndman, R. J. & Shang, H. L. (2009), 'Forecasting functional time series', *Journal of the Korean Statistical Society* **38**(3), 199–11.

Irpino, A. & Verde, R. (2008), 'Dynamic clustering of interval data using a wasserstein-based distance', *Pattern Recognition Letters* **29**(11), 1648–1658.

Kao, C.-H., Nakano, J., Shieh, S.-H., Tien, Y.-J., Wu, H.-M., Yang, C.-K. & Chen, C.-h. (2014), 'Exploratory data analysis of interval-valued symbolic data with matrix visualization', *Computational Statistics & Data Analysis* **79**, 14–29.

Kourentzes, N., Barrow, D. & Petropoulos, F. (2019), 'Another look at forecast selection and combination: Evidence from forecast pooling', *International Journal of Production Economics* **209**, 226–235.

Kubica, B. J. & Malinowski, K. (2006), Interval random variables and their application in queueing systems with long-tailed service times., *in* 'SMPS', Springer, pp. 393–403.

Kwiatkowski, D., Phillips, P., Schmidt, P. & Shin, Y. (1992), 'Testing the null hypothesis of stationarity against the alternative of a unit root: how sure are we that economic time series have a unit root?', *Journal of Econometrics* **54**(1-3), 159–178.

Le-Rademacher, J. & Billard, L. (2012), 'Symbolic covariance principal component analysis and visualization for interval-valued data', *Journal of Computational and Graphical Statistics* **21**(2), 413–432.

Li, H., Wang, J., Lu, H. & Guo, Z. (2018), 'Research and application of a combined model based on variable weight for short term wind speed forecasting', *Renewable Energy* **116**, 669–684.

Lima Neto, E. & De Carvalho, F. (2010), 'Constrained linear regression models for symbolic interval-valued variables', *Computational Statistics and Data Analysis* **54**, 333–347.

Maia, A. L. S. & de Carvalho, F. d. A. (2011), 'Holts exponential smoothing and neural network models for forecasting interval-valued time series', *International Journal of Forecasting* **27**(3), 740–759.

Maia, A. L. S., de Carvalho, F. d. A. & Ludermir, T. B. (2008), 'Forecasting models for interval-valued time series', *Neurocomputing* **71**(16-18), 3344–3352.

Makridakis, S., Spiliotis, E. & Assimakopoulos, V. (2020), 'The m4 competition: 100,000 time series and 61 forecasting methods', *International Journal of Forecasting* **36**(1), 54–74.

Maté, C. (2012), 'The bayesian model averaging approach for interval-valued data. 3 rd workshop in symbolic data analysis', pp. 47–48.

Maté, C. G. (2011), 'A multivariate analysis approach to forecasts combination. application to foreign exchange (fx) markets', *Revista Colombiana de Estadística* **34**(2), 347–375.

Moral-Benito, E. (2015), 'Model averaging in economics: An overview', *Journal of Economic Surveys* **29**(1), 46–75.

Muñoz, A., Maté, C., Arroyo, J. & Sarabia, A. (2007), 'iMLP: applying multilayer perceptrons to interval-valued data', *Neural Processing Letters* **25**(2), 157–169.

Naseer, M., bin Tariq, Y. et al. (2015), 'The efficient market hypothesis: A critical review of the literature', *IUP Journal of Financial Risk Management* **12**(4), 48–63.

Nelson, C. R. & Plosser, C. R. (1982), 'Trends and random walks in macroeconomic time series: some evidence and implications', *Journal of Monetary Economics* **10**(2), 139–162.

Noirhomme-Fraiture, M. & Brito, P. (2011), 'Far beyond the classical data models: symbolic data analysis', *Statistical Analysis and Data Mining: the ASA Data Science Journal* **4**(2), 157–170.

Ordiano, J. Á. G., Bartschat, A., Ludwig, N., Braun, E., Waczowicz, S., Renkamp, N., Peter, N., Düpmeier, C., Mikut, R. & Hagenmeyer, V. (2018), 'Concept and benchmark results for big data energy forecasting based on apache spark', *Journal of Big Data* **5**(1), 11.

Phillips, P. C. & Perron, P. (1988), 'Testing for a unit root in time series regression', *Biometrika* **75**(2), 335–346.

Ramos-Guajardo, A. B., González-Rodríguez, G. & Colubi, A. (2020), 'Testing the degree of overlap for the expected value of random intervals', *International Journal of Approximate Reasoning* **119**, 1–19.

Rapach, D. & Zhou, G. (2013), Forecasting stock returns, *in* 'Handbook of economic forecasting', Vol. 2, Elsevier, pp. 328–383.

Riddington, G. (1993), 'Time varying coefficient models and their forecasting performance', *Omega* **21**(5), 573–583.

Rodrigues, P. M. & Salish, N. (2015), 'Modeling and forecasting interval time series with threshold models', *Advances in Data Analysis and Classification* **9**(1), 41–57.

Sarno, L. & Valente, G. (2005), 'Empirical exchange rate models and currency risk: Some evidence from density forecasts', *Journal of International Money and Finance* **24**(2), 363–385.

Shaub, D. (2020), 'Fast and accurate yearly time series forecasting with forecast combinations', *International Journal of Forecasting* **36**(1), 116–120.

Sinova, B., Casals, M. R., Colubi, A. & Gil, M. Á. (2010), The median of a random interval, *in* 'Combining Soft Computing and Statistical Methods in Data Analysis', Springer, pp. 575–583.

Sinova, B., Colubi, A., González-Rodrı, G. et al. (2012), 'Interval arithmetic-based simple linear regression between interval data: Discussion and sensitivity analysis on the choice of the metric', *Information Sciences* **199**, 109–124.

Sinova, B. & Van Aelst, S. (2015), 'On the consistency of a spatial-type interval-valued median for random intervals', *Statistics & Probability Letters* **100**, 130–136.

Smith, J. & Wallis, K. F. (2009), 'A simple explanation of the forecast combination puzzle', *Oxford Bulletin of Economics and Statistics* **71**(3), 331–355.

Song, H. & Liu, H. (2017), Predicting tourist demand using big data, *in* 'Analytics in smart tourism design', Springer, pp. 13–29.

Stock, J. H. & Watson, M. W. (2004), 'Combination forecasts of output growth in a seven-country data set', *Journal of Forecasting* **23**(6), 405–430.

Stock, J. H. & Watson, M. W. (2006), 'Forecasting with many predictors', *Handbook of economic forecasting* **1**, 515–554.

Tay, A. S. & Wallis, K. F. (2000), 'Density forecasting: a survey', *Journal of Forecasting* **19**(4), 235–54.

Thomson, M. E., Pollock, A. C., Önkal, D. & Gönül, M. S. (2019), 'Combining forecasts: Performance and coherence', *International Journal of Forecasting* **35**(2), 474–484.

Timmermann, A. (2006), 'Forecast combinations', *Handbook of Economic Forecasting* **1**, 135–196.

Wang, J. & Wu, J. J. (2012), 'The taylor rule and forecast intervals for exchange rates', *Journal of Money, Credit and banking* **44**(1), 103–144.

Winkler, R. L. & Clemen, R. T. (1992), 'Sensitivity of weights in combining forecasts', *Operations Research* **40**(3), 609–614.

Zhang, W., Liu, J., Cho, C. & Han, X. (2015), 'A hybrid parameter identification method based on bayesian approach and interval analysis for uncertain structures', *Mechanical Systems and Signal Processing* **60**, 853–865.