COMPARATIVE STUDY ON RETAIL SALES FORECASTING BETWEEN SINGLE AND COMBINATION METHODS

Serkan ARAS[1], İpek DEVECİ KOCAKOÇ[2], Cigdem POLAT[3]

Dokuz Eylul Universitesi, Izmir, Turkey
E-mails: [1]serkanaras35@hotmail.com (corresponding author);
[2]ipekdeveci@yahoo.com; [3]cigdempolat35@hotmail.com

Abstract. In today's competitive global economy, businesses must adjust themselves constantly to ever-changing markets. Therefore, predicting future events in the market-place is crucial to the maintenance of successful business activities. In this study, sales forecasts for a global furniture retailer operating in Turkey were made using state space models, ARIMA and ARFIMA models, neural networks, and Adaptive Network-based Fuzzy Inference System (ANFIS). Also, the forecasting performances of some widely used combining methods were evaluated by comparison with the weekly sales data for ten products. According to the best of our knowledge, this study is the first time that the recently developed state space models, also called ETS (Error-Trend-Seasonal) models, and the ANFIS model have been tested within combining methods for forecasting retail sales. Analysis of the results of the single models in isolation indicated that none of them outperformed all the others across all the time series investigated. However, the empirical results suggested that most of the combined forecasts examined could achieve statistically significant increases in forecasting accuracy compared with individual models and with the forecasts generated by the company's current system.

Keywords: sales forecasting, neural networks, exponential smoothing, ARIMA, ARFIMA, ANFIS, combined forecasts, retail sales.

JEL Classification: M21, C53, C45.

Introduction

Retail businesses are forced to use their resources efficiently and to make sound strategic decisions for the future in order to survive and increase their revenues, especially as conditions become ever more competitive. Since all forecasts entail at least some degree of uncertainty, businesses need to make estimates with the aim of minimising uncertainty. Businesses must make predictions encompassing many variables, such as raw-material requirements, optimal stock levels, borrowing requirements, and personnel requirements. However, in order for any of these to be estimated, it is necessary first to predict the level of demand that will pertain in the market and, accordingly, the company's prospective sales. Thus, market demand forecasts are a necessary precursor of all other estimates required by a given operation. Accurate forecasts allow appropriate targeting of the company's activities (such as production, finance, R&D, purchasing,

and marketing) and facilitate them in reaching their targets (Mentzer, Bienstock 1998). Also, sales forecasting is of great importance to companies making strategic decisions regarding their future investments. For instance, sales amounts are used in combination with margin forecasts to evaluate a company's future income, and used together with turnover forecasts to assess a company's future assets (Curtis *et al.* 2014).

Retailing can be defined as consisting in the whole process of marketing goods and services directly to end users. As such, it can be thought of as a bridge between the producer and the final consumer. The capability of retailing managers to predict future sales volumes correlates with increased customer satisfaction, reduced resource waste, increased sales revenues, and more efficient and effective production plans (Chen, Ou 2011). Barksdale and Hilliard (1975) have studied the relationship between retail stocks and sales, and have concluded that the successful management of stock depends primarily on accurate prediction of retail sales. Accurate forecasting of future retail sales can contribute to gains in the efficiency and efficacy of operations carried out in retail businesses and supply chains. Therefore, forecasts play a crucial role in business management and strategic planning. The management decisions taken at every level of a business are either directly or indirectly related to the forecasts it has made. Without useful predictions, planning and control activities cannot be accomplished effectively. Poor forecasting adversely affects the capacity of organisations and companies to meet their goals, as it results in problems such as increased stock costs and inability to meet demand, which can in turn lead to losses of market share (Agrawal, Schorling 1997).

Given the essential role of retail sales and the utmost importance of forecasting them accurately, this study seeks to model sales volumes by using time series techniques. Time series sales-forecasting models can be divided into two categories: linear models and nonlinear models. Among the linear models, exponential smoothing methods (Gardner 2006) and the autoregressive integrated moving averages (ARIMA) model (Box, Jenkins 1970) have numerous effective applications (Schilling, Jarrett 2008; Nurunnabi 2012; Çatik, Karçuka 2012). Consequently, they have played special roles in sales forecasting for a long time. ARFIMA (Autoregressive Fractionally Integrated Moving Average) is a long-memory version of an ARIMA model. The ARFIMA model allows for the series to be fractionally integrated and thereby captures the long memory of the series. Many time series exhibit long-range dependence and ARFIMA models sometimes produce significantly better forecasts than non-ARFIMA models in various financial and macroeconomic data sets, as well as aggregate retail sales (Bhardwaj, Swanson 2006; Aye *et al.* 2015). It would therefore be interesting to compare their usefulness with other methods used in this study.

The artificial neural networks (ANN) model has stood out in the field of nonlinear modelling as a result of its beneficial properties, such as the capacity to model any functional form at the desired level, without having to assume any relationships in advance (Cybenko 1989). Hence, the ANN model has attracted the attention of many researchers and practitioners, who have identified a number of ANN forecasting applications that exhibit the potential to produce better forecasts (Adhikari, Agrawal 2014; Claveria, Torra 2014; Keles *et al.* 2016; Lolli *et al.* 2017). There are various types of ANN model such as Support Vector Machine (SVM), Radial Basis Function Neural

Networks (RBFNN), and the ANFIS model successfully implemented in the economic time series forecasting applications (Kim 2003; Shen *et al.* 2011; Bagheri *et al.* 2014). However, as far as we know, the ANFIS model has not been applied yet for forecasting retail sales. The ANFIS is a neuro-fuzzy approach proposed by Jang (1993). It integrates the benefits of the ANN model (learning ability of complex relationships between input and output data) and fuzzy systems (modelling ability of human knowledge and inference processes in the case of uncertain and noisy data). Thus, the disadvantages of both methods of artificial intelligence can be overcome using this hybrid system. ANFIS has led to significant results in various forecasting applications (Boyacioglu, Avci 2010; Yazdani *et al.* 2012; Svalina *et al.* 2013; Lotfi *et al.* 2016).

Another way of obtaining more accurate forecasts is to use combination methods, which are based on the use of more than one model to produce forecasts. While a single method can be any forecasting approach which has its own unique mechanism to produce forecasts, combination methods combine only the forecasts of each selected single method. Hybrid methods differ from combination methods in the context of using single methods. Hybrid methods integrate single methods with the intent of overcoming the drawbacks of each single model. Also, in many studies the term "ensemble" is used in place of "combination". Actually, the two methods do exactly the same job but under different names. Researchers from the field of machine learning use "ensemble" while researchers from statistics use "combination". We prefer "combination" in this study. The idea behind combining techniques is that each model's approach to identifying patterns is different, and combining the predictions of these single models to form the final forecast can provide advantages, as the combination can capture a broader cross section of patterns in the data (Zhang 2003). In light of the benefits that combined forecasts offer, many studies relating to retail sales forecasting have focused on methods involving the integration of some artificial intelligence techniques into neural networks, or on combining neural networks with traditional models, in order to improve forecasts (Chen, Ou 2011; Lu 2014; Du *et al.* 2015; Yang, Li 2016). However, it should be kept in mind that using combining methods is not a panacea. They require estimating multiple parameters based on the forecast values, especially when sophisticated combining methods are employed. Hence, this second round of the parameter estimation process can adversely affect the accuracy of the combined forecasts. Another drawback of using combining methods can be encountered when it is known in advance that one of the single forecasting methods significantly outperforms others. In that case, the best single method may have better performance than combining methods. The motivation for including combining techniques in this study derives from numerous published studies in the literature (Clemen 1989; De Menezes *et al.* 2000), which have asserted the superiority of combined methods over component models.

The objectives of this study are threefold. In the first place, methods such as state space models and the ANFIS model have not yet been examined for retail sales forecasting. Therefore, this study's original contribution to the existing literature is for the first time to investigate the forecasting performances of these models for retail sales. Secondly, the study analyses forecasts of the sales quantities of a retail furniture store in order to

evaluate the forecasting performance of five methods: namely, state space models, the ARIMA model, the ARFIMA model, the ANN model, and the ANFIS model. Based on the observed accuracy of each of these forecasting techniques, employees in the relevant departments can derive a general understanding of which forecasting methods conform best with which product sales, and thereby identify opportunities to ameliorate the existing supply chain system. Thirdly, we aim to determine which of the most commonly employed combined methods generates the greatest statistical improvements in the forecast accuracy of the relevant series; this will be assessed in terms of several error measures. As a consequence, the study should enable companies to identify which combining methods can be used more or less confidently to forecast the sales quantities of given products.

1. Literature review

There are numerous studies in the literature that employ quite different methods, from fairly simple ones to complex hybrids, to model sales data in order to forecast future sales. In this section we review some of them, in chronological order.

Ansuj *et al.* (1996) analysed a medium-sized enterprise's sales data for the period from January 1979 to December 1989, in terms of both the ARIMA model with interventions and the ANN model. They reached the conclusion that the ANN model's forecasts were superior to those generated by the ARIMA model. Alon *et al.* (2001) conducted a comparative study of ANN models and traditional methods, including ARIMA models, Winter's exponential smoothing, and multivariate regression for aggregate retail sales in the US. They found that ANN models outperform their traditional counterparts, and that exponential smoothing is a viable method under relatively stable economic conditions. Frank *et al.* (2003) modelled women's apparel sales using single seasonal exponential smoothing, Winters's three-parameter model and the ANN model, referring to four years' sales data. The obtained results indicated that the ANN model performs better than the other two models in terms of R^2 evaluation statistics. These three studies have concluded that ANN models outperform the classical methods used predominantly in sales forecasting. Apart from the aforementioned studies, Aburto and Weber (2007) exploited a two-stage hybrid methodology to establish a replenishment system for a Chilean supermarket. They found that the hybrid method produces better forecasts than both component models – namely, the ARIMA and ANN models – and concluded that their new system results in fewer sales failures and lower inventory levels.

In another study by Au *et al.* (2008) investigated the performance of evolutionary neural networks for sales forecasting in fashion retailing, and compared their performance with that of a fully connected neural network and the seasonal ARIMA model. This study indicated that evolutionary neural networks produce more accurate forecasts than fully connected networks, and show still greater gains in accuracy over the seasonal ARIMA model in some cases. Pan *et al.* (2013) proposed a hybrid method that integrates empirical mode decomposition with a neural network (EMD-NN) to forecast retail sales. They claimed that this EMD-NN exhibits superiority over the classical ANN model and

seasonal ARIMA, and that the proposed hybrid method can result in better performance especially when economic conditions are volatile. Taking a different approach from previous studies, Dwivedi *et al.* (2013) compared the performances of a neuro-fuzzy modelling approach, ANFIS, with ANN and linear regression, and found the ANFIS method to be best. Aye *et al.* (2015) assessed the performance of 26 models, including ARIMA, ANN, and ARFIMA, in forecasting South Africa's aggregate seasonal retail sales. They noted that, contrary to previous findings, the ANN model widely employed nonlinear model in the literature of sales forecasting is consistently outperformed by other nonlinear models. They also give an extensive literature review on sales forecasting.

Ramos *et al.* (2015) drew comparisons between the performance of state space models and the ARIMA model in forecasting retail sales of some women's footwear products. The results implied that there is no significant difference in performance between these two models for one-step and multi-step forecasts, when automatic algorithms are implemented. Fabianová *et al.* (2016) analysed refrigerator sales from a retail store. This study has an advantage over previous studies in modelling the input data uncertainty, to find the probability of achieving the level of total revenue by Monte Carlo simulation, and used sensitivity analysis to identify the variables that most influenced the target. Kolassa (2016) took discrete predictive distributions into consideration to forecast daily retail sales and explained why forecast accuracy measures are inappropriate for count data. Ma *et al.* (2016) proposed a four-step methodological framework to examine the case of SKU level retail store sales forecasting with both intra- and inter-category promotional information. The obtained results indicated that most of the improvements were achieved by exploiting the intra-category information rather than the inter-category information. They provided a detailed review on sales forecasting. Finally, Jiménez *et al.* (2017) suggested a novel feature selection methodology to obtain more accurate forecasts for online sales and to find out the relevant features of the sold products that affect the sales.

2. Forecasting and combining methods under investigation

In this section, we focus on describing the single models and forecast combination methods employed in this analysis. First, the relationship between state space models and exponential smoothing methods is presented. Next, ARIMA, ARFIMA, ANN, and ANFIS modelling processes are briefly explicated. Finally, the existing literature relating to combined forecasts and the most commonly used methods is summarised. As stated by Ahmed *et al.* (2010), for each considered model there are myriads of variations proposed in the literature, and it would be a hopeless task to consider all existing varieties. Therefore, we considered the basic version of each model. A short description of the models investigated in the paper is as follows:

2.1. The single models

2.1.1. State space models for exponential smoothing
Exponential smoothing methods have been employed for more than five decades, owing to their simplicity and to the property that they do not require complex calculations

(Granger, Newbold 2014). This is especially advantageous in the frequently encountered cases that require a forecasting technique that is reliable and capable of producing forecasts for many variables in the shortest time possible. Exponential smoothing methods consist in determining separately the forms taken by the two components of the data: namely, trends and seasonality. By taking different trend and seasonality structures into account, it is possible to define five trend components and three seasonal components, yielding fifteen combinations, each of which represents a different exponential smoothing method. These methods, presented in Table 1, are denoted by a pair of letters (T,S) specifying the forms of their components. Some of these methods are very popular, and are generally known by their special names such as simple exponential smoothing, denoted by (N,N); Holt's linear method, denoted by (A,N); the additive Holt-Winters method, denoted by (A,A); and the multiplicative Holt-Winters method, denoted by (A,M).

Table 1. All the combinations of trend and seasonality components

Trend Component	Seasonal Component		
	N (None)	A (Additive)	M (Multiplicative)
N (None)	N, N	N, A	N, M
A (Additive)	A, N	A, A	A, M
A_d (Additive damped)	A_d, N	A_d, A	A_d, M
M (Multiplicative)	M, N	M, A	M, M
M_d (Multiplicative damped)	M_d, N	M_d, A	M_d, M

Though they have been used for many years, exponential smoothing methods had not been placed within a statistical framework until the recent emergence of new methodological developments. Following a study by Ord *et al.* (1997) that linked state space formulations with exponential smoothing methods, Hyndman *et al.* (2002) extended the approach and provided likelihood calculations, model selection criteria, and computations of prediction intervals for all the methods under consideration. The detailed theoretical foundation of this approach is presented in a book by Hyndman *et al.* (2008). Each exponential smoothing method corresponds with two state space models: one with additive errors and one with multiplicative errors. Although the point forecasts obtained by the models are the same, they give rise to different prediction intervals. With the intent of differentiating between models with additive and multiplicative errors, a third component – in addition to trend and seasonality – is included in the analysis. Adding this error component raises the number of state space models under investigation to a total of 30. To distinguish different state space models, the acronym ETS (Error-Trend-Seasonal or Exponen*T*ial Smoothing) is used. Thus, the meaning of ETS (A,A,N) is that the model is composed of additive errors, additive trends and no seasonality.

State space models are formed by two groups of equations. The first group comprises measurement (or observation) equations that define the observed data. The second group comprises state (or transition) equations that define how the unobserved components or

states (level, trend, seasonal) change over time (Hyndman, Athanasopoulos 2014). To illustrate this, the state space equations of the ETS model (A,A$_d$,A) are given as follows:

$$y_t = l_{t-1} + \phi b_{t-1} + s_{t-m} + \varepsilon_t; \tag{1}$$

$$\text{Level: } l_t = l_{t-1} + \phi b_{t-1} + \alpha \varepsilon_t; \tag{2}$$

$$\text{Growth: } b_t = \phi b_{t-1} + \beta \varepsilon_t; \tag{3}$$

$$\text{Seasonal: } s_t = s_{t-m} + \gamma \varepsilon_t, \tag{4}$$

where y_t, l_t, b_t, and s_t denote the observed value, the series level, the slope, and the seasonal component of the series at time t, respectively, and m represents the period of seasonality. ε_t is normally an independently distributed random variable with zero mean and variance σ^2. Smoothing parameters are represented by the symbols α, β, ϕ, and γ.

There are certain restrictions on the values that the smoothing parameters can take: some are traditional constraints, intended to produce weighted averages from the equations, while others are imposed in order to obviate numerical instabilities when estimating the model. More details regarding the estimation process of the smoothing parameters and the selection of initial states to maximise the likelihood function can be found in a book written by Hyndman *et al.* (2008).

2.1.2. ARIMA and ARFIMA models

The ARIMA modelling procedure introduced in a pioneering study conducted by Box and Jenkins (1970) has been one of the linear models most commonly used in time series forecasting. ARIMA models are based on identifying the structure of the autocorrelations in the data. It is assumed that the variable to be forecasted consists of a linear combination of its own lagged values and errors. This assumption also indicates the weakness of ARIMA models: they cannot capture nonlinear patterns. However, they have achieved considerable success in many forecasting applications (Fildes, Makridakis 1995). The ARIMA model consists in an advanced theoretical approach that includes an iterative three-step model-building process: model identification, parameter estimation and diagnostic checking.

A non-seasonal ARIMA model can be expressed by the following equation:

$$y_t' = c + \phi_1 y_{t-1}' + \ldots + \phi_p y_{t-p}' + \theta_1 e_{t-1} + \ldots + \theta_q e_{t-q} + e_t, \tag{5}$$

where y_t' and e_t are the differenced series, and random error is specified at time t. The equation contains the lagged values of the differenced series and errors on the right-hand side. $c, \phi_i \left(i = 1, 2, \ldots, p \right)$ and $\theta_j \left(j = 1, 2, \ldots, q \right)$ are the model parameters to be estimated after determining the values of p and q.

The abbreviation ARIMA (p,d,q) is used to represent different models. p, d, q denotes the order of the autoregressive part, the degree of first differencing, and the order of the moving average part, respectively. When estimating more complex models, using the backward shift operator B provides an advantage in expressing and understanding them.

The equation given above can be expressed by means of the notation of the backward shift operator B, defined by $By_t = y_{t-1}$, as follows:

$$\left(1 - \phi_1 B - \ldots - \phi_p B^p\right)\left(1 - B\right)^d y_t = c + (1 + \theta_1 B + \ldots + \theta_q B^q)\varepsilon_t. \tag{6}$$

Or in a more simple yet equivalent formulation:

$$\phi_p\left(B\right)\left(1 - B\right)^d y_t = c + \theta_q\left(B\right)\varepsilon_t, \tag{7}$$

where $\phi_p\left(B\right)$ and $\theta_q\left(B\right)$ are polynomials of the orders p and q, and ε_t is a white-noise process with zero mean and variance σ^2.

Similarly, a long-memory model, denoted by ARFIMA (p,d,q), can be written using operator notation:

$$\phi_p\left(B\right)\left(1 - B\right)^d (y_t - \mu) = \theta_q\left(B\right)\varepsilon_t, \tag{8}$$

where $\left(1 - B\right)^d$ is the fractional differencing operator defined by:

$$\left(1 - B\right)^d = \sum_{k=0}^{\infty} \frac{\Gamma\left(k - d\right) B^k}{\Gamma\left(k + 1\right)\Gamma\left(-d\right)}, \tag{9}$$

with $\Gamma\left(.\right)$ indicating the gamma function. The parameter d can take any real value.

The stochastic process y_t is covariance stationary for $|d| < 0.5$. In that case, Hosking (1981) showed that the autocorrelation of an ARFIMA process decays hyperbolically. When $0 < d < 0.5$, the process displays long memory. For $-0.5 < d < 0$ the process is said to exhibit intermediate memory, and for $d \geq 0.5$ the process is mean reverting, hence there is no long memory.

Having determined an appropriate model order (i.e., the values of p, d, q), the parameters are then estimated so as to maximise the log likelihood of the related model. For further information regarding this procedure, see Montgomery *et al.* (2015).

2.1.3. Artificial neural networks

ANN models are inspired by the way the human brain works and are essentially a very simplified version of it, imitating its parallel processing capabilities using mathematical artificial neurons distributed in parallel layers. Though primitive versions of neural networks first emerged back in the early 1940s, the pioneering study by Rumelhart *et al.* (1986) has expanded their use widely in almost every field of scientific endeavour, and has made them a serious rival to classical statistical methods. In contrast to traditional time series models, ANNs constitute a model-free and data-driven approach which entails two significant features: namely, flexibility and generalisability. The trade-off for these appealing features is ANNs' lack of any systematic methodology by which uncertainties regarding model construction and parameter setting could be overcome.

ANNs comprise an input layer, one or more hidden layers, and one output layer, each of which contains the corresponding neurons. ANNs with a single hidden layer are generally preferred, as they are less liable to produce an overfitted model that fits well with in-sample data, but performs poorly in terms of generalisation to out-of-sample data (Zhang *et al.* 1998). It is worth noting that feedforward multilayer networks, also known

as multilayer perceptrons (MLPs), are used in the scope of this study. For time series forecasting, the relationship between the output (y_t) and the inputs ($y_{t-1}, y_{t-2}, \ldots, y_{t-p}$) has the following mathematical representation:

$$y_t = w_0 + \sum_{j=1}^{q} w_j f\left(w_{0j} + \sum_{i=1}^{p} w_{ij} y_{t-i} \right) + \varepsilon_t, \tag{10}$$

where, w_0 and w_{0j} $(j = 1, 2, \ldots, q)$ are the biases on the neurons, w_j $(j = 1, 2, \ldots, q)$ and w_{ij} $(i = 1, 2, \ldots, p, j = 1, 2, \ldots, q)$ are the connection weights between the layers of the model, $f(.)$ is the activation (transfer) function of the hidden layer, p is the number of input neurons, and q is the number of hidden neurons.

The ANN model represented in Eq. (10) basically performs a nonlinear functional mapping from the lagged values ($y_{t-1}, y_{t-2}, \ldots, y_{t-p}$) to the future value y_t, which is expressed as below. Therefore, the ANN model can be regarded as a general nonlinear autoregressive model:

$$y_t = g\left(y_{t-1}, y_{t-2}, \ldots, y_{t-p}, w \right) + \varepsilon_t, \tag{11}$$

where w consists of all parameters, and g is a function specified by the network architecture.

The number of input neurons, or lagged observations, is probably the most important factor in time series forecasting because it determines the autocorrelation structure of the time series, while the number of hidden neurons plays a key role in determining the degree of nonlinearity and flexibility possessed by a given ANN model (Zhang *et al.* 2001; Aras, Kocakoc 2016). Readers interested in a more detailed description and mathematical computations of ANN models can refer to a book written by Hagan *et al.* (1996).

2.1.4. Adaptive network-based fuzzy inference system (ANFIS)

Among neuro-fuzzy systems, the ANFIS draws more attention because of its outstanding features. It uses a feed-forward network to generate fuzzy decision rules regarding the problem at hand, where the parameters of the membership functions are adjusted by means of the relationship between input and output data to be learned. The ANFIS employs a Takagi-Sugeno type fuzzy system in which the output is a real number. It consists of two sets of parameters, namely, premise parameters (nonlinear parameters) belonging to the fuzzy inference membership functions and consequent parameters (linear parameters) belonging to output functions. The ANFIS's general architecture, with two inputs and one output, is presented in Figure 1. The tasks of each layer within the ANFIS architecture are as follows (Jang 1993):

Layer 1: This layer is formed by linguistic nodes, each of which corresponds to a linguistic input variable. Fuzzy membership functions are used to generate the outputs of these adaptive nodes described in Eq. (12):

$$O_{1,i} = \mu_{Ai}(x), \text{ for } i = 1, 2; \quad O_{1,i} = \mu_{Bi-2}(y), \text{ for } i = 1, 2, \tag{12}$$

where x and y represent inputs, A_i and B_i denote the linguistic labels for the inputs, and

$\mu_{Ai}(x)$ and $\mu_{Bi}(y)$ are the chosen membership functions. The membership function most often used is a bell-shaped function presented in the following equation:

$$\mu_{Ai}(x), \mu_{Bi-2}(y) = \exp\left(\left(-\frac{x_i - c_i}{a_i}\right)^2\right), \tag{13}$$

where a_i and b_i are the parameter set called premise parameters.

Layer 2: Each node in this layer multiplies the incoming signals to produce the output as follows:

$$O_{2,i} = w_i = \mu_{Ai}(x).\mu_{Bi-2}(y) \text{ for } i = 1,2,3,4, \tag{14}$$

where the node output w_i is the firing strength of a rule.

Layer 3: The ith node in this layer normalises its firing strength by calculating the ratio of the ith node firing strength to the sum of all firing strengths:

$$O_{3,i} = \overline{w}_i = w_i / (w_1 + w_2 + w_3 + w_4) \text{ for } i = 1,2,3,4. \tag{15}$$

Layer 4: Each node i indicated by a square node in this layer has the following node function:

$$O_{4,i} = \overline{w}_i f_i = w_i.\left(p_i x + q_i y + r_i\right) \text{ for } i = 1,2,3,4, \tag{16}$$

where p_i, q_i and r_i are the parameters set called consequent parameters.

Layer 5: This layer is composed of one fixed node indicated by a circle. The task of this single node is to compute the overall output as follows:

$$O_{5,i} = f_{out} = \sum_i \overline{w}_i f_i = \frac{\sum_i w_i f_i}{\sum_i w_i} = \text{overall output.} \tag{17}$$
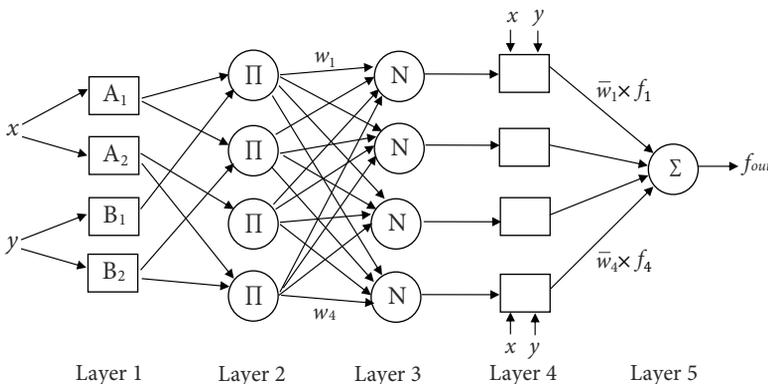

Fig. 1. ANFIS architecture with two inputs and four rules

2.2. Forecast combination methods

The study of Bates and Granger (1969) is the first to mention the combining of methods for business forecasting. A large number of studies have observed that combined forecasts exhibit better performance than both component models and single models (Clemen 1989; Makridakis 1989). Combining different models or methods as much as possible may pave the way to attaining more accurate forecasts. This claim has been supported by both theoretical and empirical findings (Wallis 2011; Bordignon *et al.* 2013; Firmino *et al.* 2015). There are many studies that promote different aspects of combining for time series data. Granger (1989), Clemen (1989) and Wallis (2011) gave detailed reviews of the developments of their times. Wallis (2011) also investigated the impact of forecasters' different information sets on the original point forecast combination result, and properties of different methods of combining density forecasts. De Gooijer and Hyndman (2006) emphasized the place of combining in forecasting, and briefly list the literature. Yu *et al.* (2005) presented combining as a dilemma, since selecting the models to be combined is a tedious task. They proposed a a double-phased procedure of time series forecasting to decide whether to select a single model or to combine models. Zou and Yang (2004) addressed the idea that combining (or mixing) forecasts from very similar models is also important, and that combining has great potential to reduce the variability that arises in the forced action of selecting a single model.

The difficulty faced by combining methods is in determining how, specifically, single models should be combined so as to result in more accurate forecasts. Many methods of resolving this difficulty have been proposed. There are a number of sophisticated statistical approaches, but it has been observed that simple methods tend to yield forecasts better than or nearly as accurate as sophisticated ones (Stock, Watson 1999; Armstrong 2001). Armstrong (2001) suggested that simple combining methods would be a reasonable approach when the problem in question involves substantial uncertainty. Some attractive features of simple combining methods are that they are easy to understand and implement, less prone to mistakes, and do not depend on any assumptions: these features have made them a frequently applied component of the forecaster's toolbox. The combining methods taken into consideration by this study are briefly described as follows.

Simple Mean: All forecasts at time t generated by single models are assigned the same weight. Many studies have shown that this is an efficient and robust method (Jose, Winkler 2008; Timmermann 2006).

Trimmed Mean: This averages forecasts of single models, but excludes the highest t % and lowest t % of the forecasts. The benefits of a 20% trimmed mean are extensively examined in Jose, Winkler (2008), Wilcox (2012).

Simple Median: This method is formed as the median of the forecasts generated by single models. The median combination is less prone to be affected by extreme values than the simple mean combination, so it tends to produce good results (Stock, Watson 2004; Wallis 2011).

Least Square Weights: In this method (LS weights), the forecasts generated by single models are regressed against the actual values. After that, the coefficients from the resulting regression are employed as the weights of the single models in combination (Granger, Ramanathan 1984; Timmermann 2006).

Mean Square Error Weights: The method of MSE weights, devised by Stock and Watson (1999), uses the MSE values computed from the period of model estimation of single models to calculate the weights of single models, via Eq. (18), as seen below:

$$w_i = \left(1 / MSE_i\right)^k / \sum_{j=1}^{n} \left(1 / MSE_j\right)^k, \qquad (18)$$

where k denotes the power to be raised. If k is set to $k = 0$, it places equal weight on single models. When k is increased, more emphasis is put on those models performing well. $k = 1$ is the most popular choice (Armstrong 2001).

Mean Square Error Ranks: The method of MSE ranks, proposed by Aiolfi and Timmermann (2006), sorts single models into clusters with respect to their MSE values by a k-means algorithm, then forecasts are pooled within each cluster. Finally, this method determines the weights of each cluster in combination so as to be inversely proportional to the models' ranks. This combination method can be expected to be more robust than the methods of LS weights and MSE weights, because it is less sensitive to outliers (Timmermann 2006).

3. Empirical study

3.1. Data description and experimental setup

The retail sector has a rapidly changing and growing structure, and retail areas are expanding every year in Turkey. The sales forecasting application conducted in this study is based on the weekly sales data of an international furniture company, which has operated in Turkey's retail sector for many years. It is one of the leading companies in its sector, preferred by many customers. The company is required to formulate effective plans in order to survive in Turkey's market, which is characterised by high levels of uncertainty and challenging competition. Like many other companies, this company exploits historical data to forecast future sales with the aim of minimising the adverse effects of fluctuations in sales.

Table 2. Description of the series under investigation

Series	Products	Series	Products
Data 1	Three-seat sofa bed	Data 6	Paper napkin
Data 2	Drawer mat, transparent	Data 7	Armchair cushion
Data 3	Pendant lamp	Data 8	Photo frame
Data 4	Coffee table	Data 9	Slatted bed base
Data 5	Tea towel	Data 10	Cord set

The sales data of ten products selected from different product groups are investigated in this study. Table 2 gives information regarding these data sets and the corresponding description for each. The selection of products was based on both the differences between products in terms of randomness of sales patterns (see Fig. 2) and expert opinions from the retailer's sales department on their difficulty of forecasting the sales.
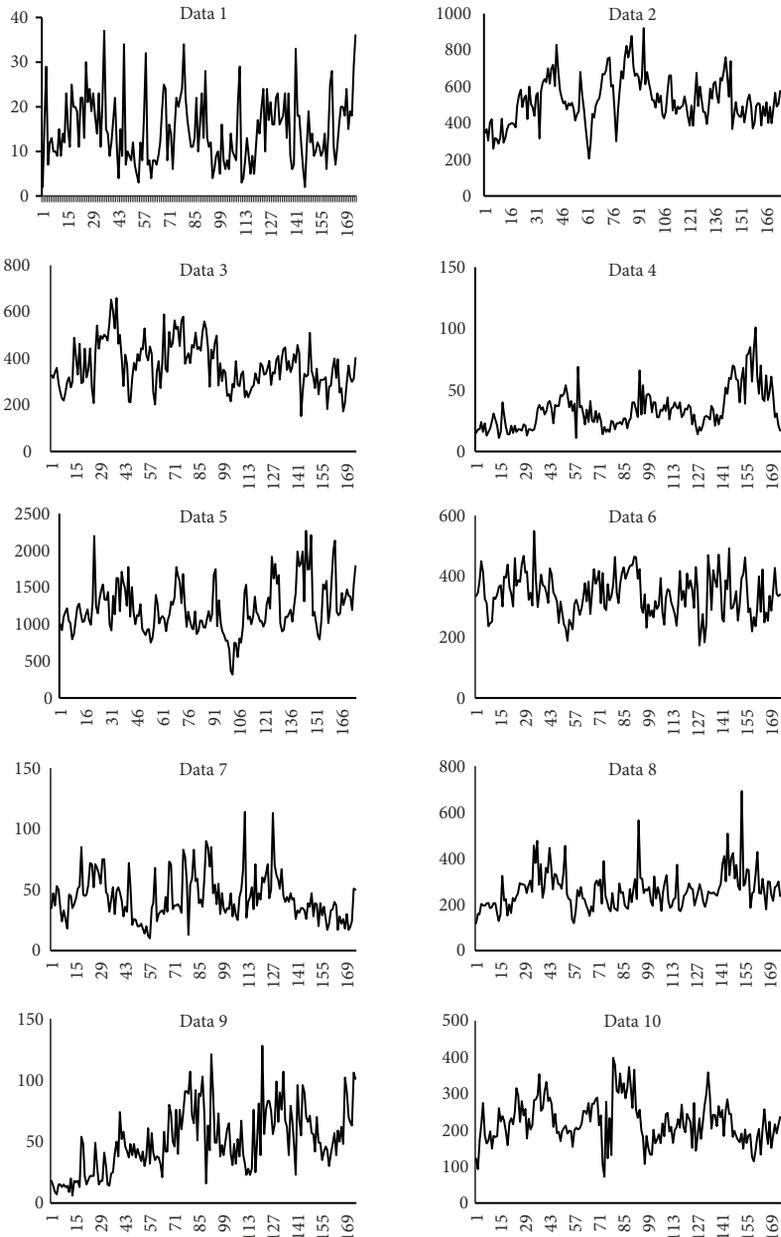

Fig. 2. Time plots of the series

The data was obtained weekly, starting 2 January 2012 and ending 2 May 2015, giving a total of $n = 174$ observations, as illustrated in Figure 2. The forecasting models are estimated through the first 144 observations. The the last 30 observations of each series are reserved so they can be used to calculate forecast evaluation statistics, revealing the models' unseen-data generalisation accuracy. For the ANN and ANFIS models, the first 144 observations are divided into two parts to make room for a validation set. The last 30 of these observations are reserved for validation purposes when constructing these models. There are some missing values in several data sets, as certain products were out of stock for one or more weeks at certain times. Missing sales are predicted by the average of the sales of the previous five weeks. Since these lacunae never affect more than a few weeks' worth of data for any one product, the missing values' impact on this analysis is very low. Details concerning the model construction and parameter selection processes employed for the single models are discussed in the following paragraphs.

The statistical framework provided by state space models enables us to employ well-known information criteria for model selection. For this study, the AIC corrected for small-sample bias (AIC_C) given below is utilised in order to select the model with the smallest AIC_C among 30 possible ETS models:

$$AIC_C = AIC + \frac{2(k+1)(k+2)}{T-k},$$ (19)

where T denotes the total number of observations kept for model building, and k is the sum of the number of parameters and initial states used for the model estimation procedure.

30 possible models, some of the ETS model combinations – namely, ETS(M,M,A), ETS(M,M_d,A), ETS(A,N,M), ETS(A,A,M), ETS(A,A_d,M), ETS(A,M,N), ETS(A,M,A), ETS(A,M,M), ETS(A,M_d,N), ETS(A,M_d,A), and ETS(A,M_d,M) – can cause numerical instabilities (Hyndman, Khandakar 2008). As such, these ETS models are excluded from consideration. The *ets* function from the *forecast* package in the *R* programming language is used to estimate all possible ETS models.

Building ARIMA models requires that one assumes homoscedasticity and stationarity for the series prior to embarking on the modelling procedure. If and when the time series exhibits heteroscedasticity, a Box-Cox transformation is employed to stabilise variance for the analysis. In regards to whether the stationarity assumption holds, we made use of unit root tests based on a null hypothesis of no unit root. We did not adopt unit root tests with the null hypothesis that a unit root exists, because those tend to produce the result that more differences than necessary are taken to make the series stationary. Keeping the number of differences as small as possible improves the quality of forecasts and leads to narrower prediction intervals (Smith, Yadav 1994). In keeping with this, the degree of first differencing d is determined by means of successive KPSS unit root tests (Kwiatkowski *et al.* 1992). This means that if we find a unit root for the series, then the KPSS test is performed on the differenced series at that time, and this process is repeated until the result of the test is insignificant. To find the degree of seasonal dif-

ferencing D, the Canova-Hansen test (Canova, Hansen 1995) is applied. Once the orders of D and d have been found, in turn, via the aforementioned tests, the values of p, q, P and Q are determined by an information criterion. In order to select the appropriate ARIMA model, an AIC_C formulation (formed as described below) has been employed in the scope of this study. The model with the lowest AIC_C value of all candidates is chosen as the final ARIMA model. Even if the final ARIMA model appears to be the best with respect to the AIC_C value, the residuals from the chosen model must be checked to ensure that they are not distinguishable from a white-noise series. We have used a portmanteau test (the Ljung-Box test) for this purpose:

$$AIC_C = AIC + \frac{2(p+q+P+Q+k+1)(p+q+P+Q+k+2)}{T-p-q-P-Q-k-2}, \qquad (20)$$

where $k = 1$ if $c \neq 0$ and $k = 0$ if $c = 0$.

The paremeters of the ARFIMA(p,d,q) model were estimated by the exact maximum likelihood estimation procedure in this study. Firstly, the Hydman and Khandakar (2008) algorithm was exploited to determine the p and q parameters, after which the Haslett and Raftery (1989) algorithm was employed to calculate the maximum likelihood estimators of all parameters of the full ARFIMA(p, d, q) model including d. All computations for the ARIMA and ARFIMA models were implemented via the *Arima* and *arfima* functions from the *forecast* package in *R*.

As for the details of the experimental design formed for the ANN models, the backpropagation algorithm was implemented, as it is the most popular and widely used. There were two motivations for using a validation set for a neural network. The first is to halt iterations when neural networks start to lose their generalisation ability. Secondly, in selecting the final ANN model it is useful to identify which has the smallest error measure on the validation set. The general choices made for the activation functions are as follows: no activation function for the neurons of the input layer, nonlinear activation functions (logistic or hyperbolic) for the hidden neurons, and the linear function for the single output neuron in order to obviate the unwanted effects a nonlinear activation function would have on the predicted output (Zhang *et al.* 1998). The logistic function is employed as the activation function for all hidden neurons. The selection of appropriate numbers of hidden neurons and input neurons is data-dependent, and there is no generally accepted rule to determine these parameters. Hence, experiments are often carried out to determine these parameters. To this end, the numbers of input neurons and hidden neurons are varied from one to ten, with the result that 100 different models are considered for this analysis in the ANN model building process. One-step-ahead forecasts are the focus of this study. Owing to the choice of one-step-ahead forecasting, the output layer is made up of only one output neuron. Each of the ANN models is replicated 30 times using different random initial weights. In total, 3000 neural networks are constructed to model each time series. All the ANN models were trained using the Levenberg-Marquardt optimisation algorithm, with assistance from the MATLAB software package.

According to the KPSS test performed in the ARIMA models, data series 2, 4, and 9 are not stationary. Some studies have demonstrated that forecast accuracy is improved by converting a non-stationary series into a stationary series prior to modelling with neural networks (Nelson *et al.* 1999; Zhang, Qi 2005). Therefore, the differenced series are used to build the ANN and the ANFIS models for these data sets. For the ANFIS models, an initial Sugeno-type fuzzy inference system was generated using membership functions and rules were extracted from data clusters found by fuzzy c-means clustering. After many trials, some parameters related to the ANFIS model were determined in a way to minimise the validation error. The types of membership function were selected as gaussian for the input nodes and linear for the output node. The best cluster size for each of the lagged values was found to be 2. The value of the exponent for the fuzzy partition matrix was specified as 2. The maximum number of the training epoch was set at 30. The selected training algorithm for the ANFIS model was a hybrid method that used the backpropagation algorithm to compute the parameters of the input membership functions and the least squares method to compute the parameters of the output membership function. All ANFIS computations were carried out in MATLAB.

3.2. Forecast evaluation statistics

It is a well-known fact that there is no consensus on the best error measure for forecasting (Makridakis *et al.* 1982; Armstrong, Fildes 1995). As such, a variety of different forecast evaluation statistics are employed to evaluate the forecast accuracy of the ten time series in this study: Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Symmetric Mean Absolute Percentage Error (sMAPE), and Theil's U, respectively:

$$\text{RMSE} = \sqrt{\sum_{t=1}^{n} \frac{\left(\hat{y}_t - y_t\right)^2}{n}}; \tag{21}$$

$$\text{MAE} = \frac{1}{n}\sum_{t=1}^{n}\left|\hat{y}_t - y_t\right|; \tag{22}$$

$$\text{MAPE} = \frac{100\%}{n}\sum_{t=1}^{n}\left|\frac{y_t - \hat{y}_t}{y_t}\right|; \tag{23}$$

$$\text{sMAPE} = \frac{100\%}{n}\sum_{t=1}^{n}\frac{\left|\hat{y}_t - y_t\right|}{\left(\left|y_t\right| + \left|\hat{y}_t\right|\right)/2}; \tag{24}$$

$$\text{Theil-U} = \sqrt{\frac{\sum_{t=1}^{n}\left(\dfrac{\hat{y}_{t+1} - y_{t+1}}{y_t}\right)^2}{\sum_{t=1}^{n}\left(\dfrac{y_{t+1} - y_t}{y_t}\right)^2}}, \tag{25}$$

where y_t is the observation value at time t, $\hat{y}_t$ denotes the forecast value at time t, and n is the number of data points in the test set.

RMSE and MAE are widely used scale-dependent error measures, whereas MAPE and sMAPE are scale-independent error measures, which have the advantage of enabling comparisons between the forecast performance of different data sets. However, when $y_i = 0$s for any i in the period of interest, MAPE and sMAPE bear the risk of producing infinite or extreme values. Theil's U, developed by Theil (1966), is a relative accuracy measure which also penalises models with large errors. It allows comparisons to be made between the model under investigation and the naive method. If the U value equals 1, there is no difference between a naive forecasting and the model used. If the U value is greater than 1, it means that using the model would be pointless because the naive method would lead to better forecasts. If the U value is less than 1, the forecasting model is superior to the naive method.

3.3. Results

Two values are taken into consideration by the retailer – namely, AWS (Average Weekly Sales) and EWS (Expected Weekly Sales) – to predict future sales quantities and to determine the reorder points for its products. These are the outputs of the current system, which are supplied to the employees responsible for the operational activities of stock control. AWS is computed with the help of the exponential smoothing method, using the prespecified α value once the weekly sales have been realised. In other words, AWS represents an average of a product's weekly sales and is used as an input for calculating EWS values, whereas EWS represents a product's expected weekly sales and is used by the company as its final forecast for sales of the product in question. The stock planners simply enter weekly sales data into the system, and it returns the AWS and EWS values. To assess the relative effectiveness of the forecasts in use, the forecasts given by the AWS and EWS values are compared with the results of the forecasting techniques examined in this study. Hence, we have thirteen different forecasts in total for each time series regarding sales of the product in question; that is, two forecasts produced by the system under investigation (AWS, EWS), five single models (ETS, ARIMA, ANN, ARFIMA, ANFIS), and six combining methods (Simple mean, Trimmed mean, Simple median, LS weights, MSE weights, and MSE ranks).

Table 3 contains the details of the selected forecasting models for each of the time series. We used ANN(I×H×O) and ANFIS(I×T×O) notations to represent the number of lagged values of the series (I), hidden nodes (H), output nodes (O), and the total number of nodes (T) used in the ANFIS model. As can be seen from the table, d parameters for the ARFIMA models lie in the range $0 > d > 0.5$ for Data 1, 3, 7, and 9. Hence, one can say that sales quantities for these data sets exhibit a long memory.

The forecasting performance of the aforementioned techniques for all the series are presented in Tables 4–8 in terms of RMSE, MAE, MAPE, sMAPE, and Theil's U statistics, computed on the test sets. Bold type is used in these tables to emphasise the forecasting methods that exhibit the lowest values in terms of any of the error measures. As the tables show, no single forecasting technique is the best for all data sets with respect to all statistics considered. This observation supports the generally accepted principle that no individual forecasting model is the best for all situations under all circumstances

Table 3. Forecasting models selected for the ten time series under investigation

Time Series	ETS	ARIMA	ARFIMA	ANN	ANFIS
Data 1	ETS(A,N,N)	ARIMA (3, 0, 0)	ARFIMA (1, 0.17, 0)	2×1×1	2×17×1
Data 2	ETS(A,N,N)	ARIMA (0, 1, 4)	ARFIMA (3, 0.00, 2)	3×4×1	2×17×1
Data 3	ETS(A,N,N)	ARIMA (1, 0, 3)	ARFIMA (0, 0.33, 3)	3×1×1	2×17×1
Data 4	ETS(M,N,N)	ARIMA (0, 1, 1)	ARFIMA (2, 0.00, 0)	2×6×1	2×17×1
Data 5	ETS(A,N,N)	ARIMA (1, 0, 1)	ARFIMA (1, 0.00, 1)	1×1×1	5×32×1
Data 6	ETS(A,N,N)	ARIMA (3, 0, 0)	ARFIMA (3, 0.00, 0)	1×8×1	5×32×1
Data 7	ETS(A,N,N)	ARIMA (1, 0, 0)	ARFIMA (1, 0.22, 0)	2×1×1	3×22×1
Data 8	ETS(M,N,N)	ARIMA (1, 0, 3)	ARFIMA (1, 0.00, 3)	1×2×1	1×12×1
Data 9	ETS(M,A,N)	ARIMA (0, 1, 1)	ARFIMA (1, 0.06, 2)	2×2×1	2×17×1
Data 10	ETS(A,N,N)	ARIMA (3, 0, 2)	ARFIMA (3, 0.00, 2)	2×10×1	2×17×1

(Makridakis *et al.* 1982; Chatfield 1988). It should be noted that the Theil's U values obtained for Data 4 are all higher than 1, except in the case of the ARIMA and ARFIMA models. This means that the naive method leads to better forecasts for Data 4. When we look at the time series plot regarding this data set, it can be seen that the last 30 observations – which form the test set – have different patterns than those of all the other observations. The fact that the forecasting techniques did not incorporate these different patterns into their estimating process may explain their inferior forecasting performance relative to the naive model for this data set.

The tables demonstrate that the combined forecasts – whichever method is used – tend to outperform the worst forecasts of the single models, in terms of all forecast evaluation statistics for all the data sets considered. Hence, the risk of choosing a poorly performing single model might be minimised by means of combination methods. Granger (1989) has observed that when the component models are suboptimal, the combining method can outperform its component models. With regards to the selection of the final model for each of the single models in this study, we cannot guarantee that the selected models are optimal. Information criteria were used to select the final model for the ETS, ARIMA and ARFIMA models, and the ANN and ANFIS models selected were the models with the lowest error value on the validation set. Therefore, it is to be expected that the overall performances of the combining methods should be superior to those of the single models. The obtained results confirm this expectation, except in the case of Data 7. For this series, the selected ETS is probably the optimal model: it has exhibited clear superiority over all other single models and combining methods, in terms of all the forecast evaluation statistics considered.

Also, to investigate the impact of change in forecasting path on the accuracy of the studied methods, five forecasting horizons, i.e. $h = 6, 12, 18, 24, 30$, each of which had a different starting point, were considered, as in Çatik and Karçuka (2012). In other words, to obtain forecasts, for instance, for $h = 6$, all models were re-estimated using

168 observations and the last six weeks were reserved for out of sample forecasting comparison. The results of these comparisons are presented separately for single and combining methods in Figure 3. The mean ranks of the forecasting methods computed from the RMSE statistics were taken into consideration to examine the effect of the different forecasting paths on the methods considered. The mean ranks computed from other error statistics are similar to the one reported here, therefore they are not included here to save space. As can be seen in Figure 3, the patterns are somewhat mixed. There is no clear evidence of any effect of the different forecasting paths on the forecasting performances for the single methods. But it does show that the ANFIS is the most volatile method across different forecasting horizons. As for the combining methods, there is an increasing trend for the LS method. It can be said that this method is only effective in short forecasting horizons.
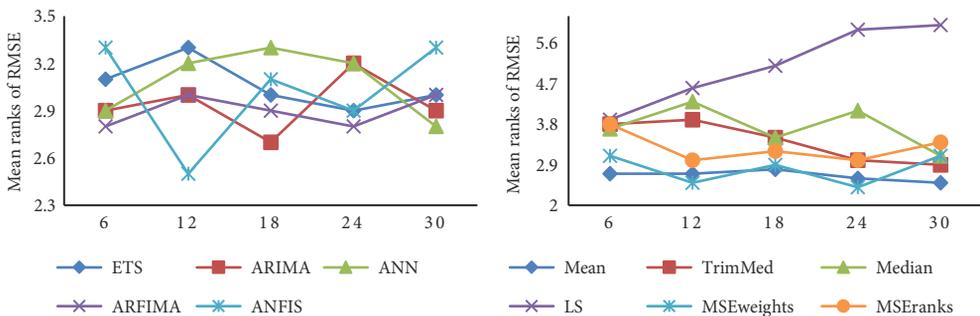

Fig. 3. Graph of the mean ranks of the methods for five forecasting horizons

Table 4. Forecasting results of the single models and combining methods for Data 1 and 2

Techniques	Data 1					Data 2				
	RMSE	MAE	MAPE	sMAPE	Theil-U	RMSE	MAE	MAPE	sMAPE	Theil-U
AWS	7.38	5.74	63.68	39.22	0.86	146.9	108.6	24.87	20.38	1.76
EWS	7.84	6.21	71.12	42.51	1.01	136.4	102.1	22.65	18.90	1.44
ETS	6.96	5.29	55.51	36.55	0.60	90.07	68.39	15.06	14.06	0.86
ARIMA	6.84	5.36	54.47	37.05	0.54	**84.25**	63.94	14.45	13.02	0.84
ANN	6.80	5.20	52.79	**36.05**	**0.51**	87.50	65.12	14.76	13.25	0.89
ARFIMA	6.72	5.28	53.54	36.50	0.53	92.42	70.48	15.78	14.41	0.90
ANFIS	6.84	5.23	**51.29**	37.40	0.71	87.66	66.29	14.66	13.59	0.83
Simple mean	6.74	5.26	55.87	36.44	0.56	86.98	63.03	14.56	12.83	0.85
Trimmed mean	6.80	5.29	55.58	36.58	0.56	85.63	61.27	14.04	12.51	0.82
Simple median	6.73	5.24	54.44	36.21	0.53	85.41	**61.10**	**13.85**	**12.45**	**0.79**
LS weights	7.00	5.27	55.00	36.43	0.53	91.60	70.42	16.14	14.25	0.96
MSE weights	6.74	5.23	55.02	36.25	0.55	84.25	62.44	14.23	12.76	0.81
MSE ranks	**6.72**	**5.20**	53.45	36.08	0.52	84.67	62.24	14.26	12.71	0.82

Note: Bold numbers indicate the best forecasting techniques in terms of the error measure considered.

Table 5. Forecasting results of the single models and combining methods for Data 3 and 4

Techniques	Data 3					Data 4				
	RMSE	MAE	MAPE	sMAPE	Theil-U	RMSE	MAE	MAPE	sMAPE	Theil-U
AWS	76.67	61.57	23.00	19.38	1.06	18.53	15.67	37.78	30.13	1.33
EWS	98.04	79.14	26.40	23.18	1.11	28.55	24.01	56.17	44.96	2.18
ETS	72.70	57.19	19.48	18.54	0.89	17.19	15.42	33.49	29.19	1.03
ARIMA	72.70	57.47	19.67	18.23	0.85	**16.89**	**15.24**	32.78	**28.70**	0.99
ANN	76.62	57.65	20.87	18.13	0.95	25.40	18.88	37.00	32.13	1.24
ARFIMA	77.80	60.58	21.15	19.06	0.94	17.46	15.58	**31.13**	28.98	**0.98**
ANFIS	71.27	55.51	20.04	17.49	0.86	19.63	16.48	33.37	29.91	1.05
Simple mean	66.63	52.54	18.71	16.67	0.81	17.33	15.49	34.56	29.47	1.11
Trimmed mean	67.51	52.00	18.64	16.49	0.81	17.29	15.60	33.69	29.39	1.04
Simple median	68.01	52.60	18.80	16.69	0.82	17.35	15.61	33.16	29.25	1.02
LS weights	85.49	72.64	24.54	21.88	0.98	31.95	22.37	43.14	36.34	1.57
MSE weights	**66.45**	**51.28**	18.27	**16.31**	0.80	18.00	15.89	33.87	29.60	1.03
MSE ranks	66.67	51.29	**18.22**	16.32	**0.80**	18.53	16.08	33.85	29.72	1.03

Note: Bold numbers indicate the best forecasting techniques in terms of the error measure considered.

Table 6. Forecasting results of the single models and combining methods for Data 5 and 6

Techniques	Data 5					Data 6				
	RMSE	MAE	MAPE	sMAPE	Theil-U	RMSE	MAE	MAPE	sMAPE	Theil-U
AWS	568.2	434.0	34.10	29.62	2.31	71.01	61.99	21.15	18.66	0.98
EWS	659.6	536.3	38.22	32.74	2.13	85.94	67.57	21.91	19.02	1.07
ETS	380.5	305.5	22.52	21.22	1.06	72.23	60.83	18.74	18.27	0.90
ARIMA	363.7	281.5	19.64	19.57	0.96	73.52	60.60	19.08	18.08	0.87
ANN	**342.7**	**261.1**	**17.87**	**18.07**	**0.88**	69.69	56.35	**17.66**	16.79	0.83
ARFIMA	369.4	287.3	20.31	20.00	0.98	70.35	56.54	17.81	16.89	0.81
ANFIS	383.9	314.2	21.46	22.08	1.03	72.08	59.73	18.72	17.93	0.87
Simple mean	364.8	296.8	21.40	20.83	1.03	**66.15**	**55.84**	17.94	**16.70**	**0.80**
Trimmed mean	358.0	281.5	19.65	19.67	0.96	67.64	56.12	17.84	16.79	0.80
Simple median	356.3	281.2	19.59	19.64	0.95	69.35	57.82	18.26	17.29	0.82
LS weights	369.9	276.5	18.14	19.12	0.93	301.9	268.5	53.9	44.87	2.62
MSE weights	359.0	281.8	19.92	19.64	0.97	68.87	56.96	18.01	17.06	0.82
MSE ranks	359.7	282.0	19.85	19.65	0.97	68.50	56.62	17.93	16.94	0.81

Note: Bold numbers indicate the best forecasting techniques in terms of the error measure considered.

Table 7. Forecasting results of the single models and combining methods for Data 7 and 8

Techniques	Data 7					Data 8				
	RMSE	MAE	MAPE	sMAPE	Theil-U	RMSE	MAE	MAPE	sMAPE	Theil-U
AWS	17.39	12.01	53.69	34.20	1.88	134.6	116.5	41.20	34.10	1.13
EWS	11.11	9.47	35.12	31.64	0.93	126.7	100.6	34.44	29.74	1.16
ETS	**9.99**	**7.91**	**28.72**	**26.23**	**0.81**	102.4	73.16	23.61	22.13	0.83
ARIMA	12.35	10.12	43.31	31.85	1.01	114.3	82.26	25.63	25.37	0.95
ANN	11.92	9.39	39.77	29.77	0.94	108.3	70.13	20.10	21.57	0.90
ARFIMA	10.70	8.43	34.59	27.30	0.85	108.7	73.15	22.42	22.00	0.89
ANFIS	11.16	8.35	34.81	26.81	0.86	105.3	**67.27**	**19.47**	**20.67**	0.89
Simple mean	10.59	8.62	36.38	28.15	0.90	**101.3**	71.61	22.81	21.76	**0.83**
Trimmed mean	10.51	8.38	35.21	27.35	0.86	102.7	69.39	21.50	21.02	0.84
Simple median	10.59	8.38	35.55	27.44	0.87	105.7	69.65	21.27	21.04	0.86
LS weights	11.34	9.12	39.27	29.38	0.97	109.0	72.96	21.92	22.38	0.91
MSE weights	10.64	8.59	36.15	27.99	0.88	101.9	69.26	21.42	20.99	0.83
MSE ranks	10.86	8.66	36.47	28.04	0.88	102.1	69.71	21.65	21.14	0.83

Note: Bold numbers indicate the best forecasting techniques in terms of the error measure considered.

Table 8. Forecasting results of the single models and combining methods for Data 9 and 10

Techniques	Data 9					Data 10				
	RMSE	MAE	MAPE	sMAPE	Theil-U	RMSE	MAE	MAPE	sMAPE	Theil-U
AWS	17.03	13.80	25.58	22.70	0.95	**32.93**	**24.98**	**15.53**	**13.76**	0.88
EWS	19.38	13.90	24.09	23.79	1.05	66.73	46.35	23.53	21.62	1.25
ETS	17.33	14.53	25.90	23.52	0.95	35.43	30.06	17.09	16.43	0.87
ARIMA	16.95	12.89	21.30	21.02	0.90	35.36	27.58	16.96	14.96	0.85
ANN	16.88	13.79	22.66	21.86	0.89	41.35	34.00	20.16	17.79	0.99
ARFIMA	16.95	12.42	19.99	20.19	0.88	35.65	27.94	17.07	15.17	0.86
ANFIS	17.07	13.75	22.35	21.82	0.89	36.60	28.67	17.20	15.35	0.94
Simple mean	15.95	12.66	21.57	20.61	0.86	34.24	28.00	16.60	15.05	**0.83**
Trimmed mean	16.24	12.87	21.68	20.90	0.87	34.39	27.69	16.62	14.97	0.84
Simple median	**15.91**	**12.33**	20.69	**20.12**	**0.85**	34.92	27.31	16.46	14.77	0.84
LS weights	19.70	13.53	**19.68**	22.78	1.03	39.11	31.20	19.14	16.58	0.97
MSE weights	16.32	13.32	22.58	21.47	0.87	35.01	28.49	17.03	15.33	0.86
MSE ranks	16.21	13.12	22.30	21.25	0.87	35.20	28.61	17.14	15.39	0.86

Note: Bold numbers indicate the best forecasting techniques in terms of the error measure considered.

It cannot be judged from Tables 4–8 whether there is a statistically significant difference in performance between the single and combining methods when considering all data sets, even if it is possible to apply some interpretations. To identify the best overall performance attained by the methods under investigation, statistical tests are performed on the five performance criteria described above. For this purpose, Friedman's test – a non-parametric alternative to one-way ANOVA with repeated measures – is performed for all evaluation statistics. This test evaluates the null hypothesis (H_0) – that is, the hypothesis that the forecasting performances of the methods under examination do not differ with respect to the corresponding error measure – against the alternative hypothesis (H_1) that there are differences in performance among them (Hollander *et al.* 2013). The Friedman's test results that were obtained are presented in Table 9. In this table, p-value is the probability that the H_0 hypothesis is true. Based on these small p-values, all H_0 hypotheses are rejected with a 95% confidence level. This means that the various forecasting methods exhibit statistically different performances with respect to RMSE, MAE, MAPE, sMAPE, and Theil's U statistics when all data sets are taken into account.

Table 9. Results of the Friedman's test statistics in terms of forecast evaluation measures

	Friedman statistics	
	χ_2	*p*-values
RMSE	70.05	0.00000
MAE	50.70	0.00001
MAPE	40.64	0.00006
sMAPE	52.92	0.00000
Theil-U	62.59	0.00000

After finding statistically significant differences in forecasting performances, identifying the source of these differences is of importance for this study to determine specifically which forecasting methods exhibit which specific differences. A multiple-comparison test can provide this information. Multiple-comparison procedures identify which forecasting methods are significantly different from which others, drawing on the differences between the mean ranks of the methods in question. Detailed information about these procedures is given in a book by Hochberg and Tamhane (2009). The results of the multiple-comparison procedures employed in this study are demonstrated in Figure 4. The value of the mean rank of each forecasting method is represented by a circle, and the line extending out from each circle represents the comparison interval. The performances of two forecasting methods are significantly different if their intervals are disjoint. Conversely, overlapping intervals indicate that the relevant forecasting performances do not differ significantly. When multiple-comparison tests are performed, the risk of making type-I error increases. To avoid this kind of error, the Bonferroni method (Corder, Foreman 2014) is exploited in this study.
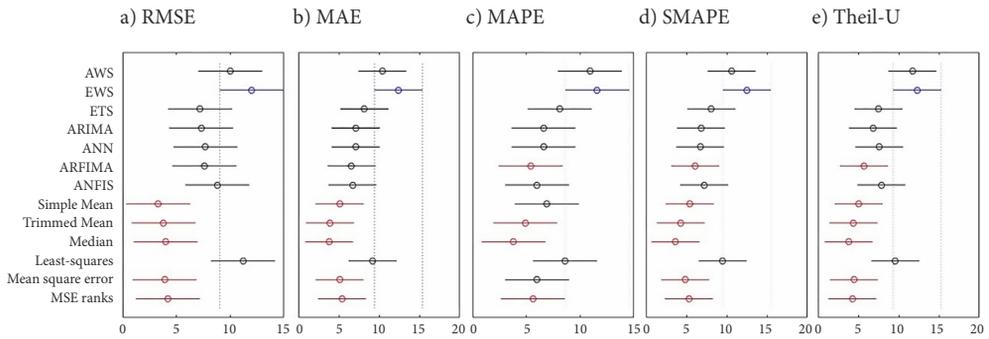
Fig. 4. Results of the multiple-comparison procedures in terms
of RMSE, MAE, MAPE, sMAPE, and Theil's U

Generally speaking, these figures provide statistical comparisons of RMSE, MAE, MAPE, sMAPE and Theil's U across different forecasting techniques and the EWS and AWS methods. These comparisons indicate that the forecasting performances of all the techniques considered are superior to those of the forecasts produced by the retailer's existing system, for all ten products investigated. Though there are some differences in performance among the combining methods, it is apparent that combined forecasts improve forecast accuracy with respect to all the error measures considered across all the series. In particular, some combining methods provide statistical proof of this superiority. Examination of Figure 4 reveals clearly that there are no statistically significant differences in performance among the single models. If a comparison is made between the combining methods and the single models, it can be seen that the combining methods – other than the least-square weights – generally tend to produce more accurate forecasts.

Discussion and conclusions

In today's market conditions, with competition increasing day by day and shortened product life spans, companies need to be fast, flexible and agile in order to compete. This ultra-competitive environment means that the power is now in the hands of the end consumer, so businesses that are able to anticipate the customer's wishes will always be one step ahead. Companies striving to maintain their market presence must be able to analyse and interpret market movements and variations in customers' needs well in advance, enabling them to direct their business activities appropriately. It is now a necessity for businesses to anticipate the goods or services that will be requested by customers, and to make the necessary preparations for supplying them in advance. Accordingly, researchers have turned their attention to demand forecasting methods, and have already developed an array of methods for that purpose. In this study, we have presented an experimental analysis of five single models and of the most popular methods of combining forecasts, as applied to sales forecasting for a furniture retailer operating in Turkey. Based on the findings reported in this study, some remarkable conclusions may be drawn.

First, in time series applications it is almost impossible to know in advance which forecasting model will perform best for a given data set. This corresponds with the generally accepted fact that no single model is the best for all situations under all circumstances. For this reason, the meaning of 'best' can be interpreted in various ways, such as that the best model is the model whose forecasts both perform well and exhibit little variability in performance, whichever time series is analysed. From this point of view, the empirical results obtained in this study indicate that combining methods can be called 'best' performer in this context and can confidently be used to forecast the sales of any product of a retail company, as they typically produce robust forecasting performances whichever time series is considered. This study also found that the existing forecasts of the company could be statistically improved by using combining methods, except the LS weights combining method. This had the worst overall performance among the combining methods examined, and even performed worse than the single methods. A possible reason for this is that the combining method of LS weights is the most elaborate method requiring the estimation of many parameters. This finding supports the claim in the literature that simple combining methods tend to produce forecasts better than the sophisticated ones. When the combining methods employed in the scope of this study are examined, it is seen that they consist of three simple methods (simple mean, trimmed, median) and three more sophisticated ones (LS weights, MSE weights, MSE ranks). It is observed that simple methods yield similar performances or even better than more complicated combining methods, and this finding is in line with the widely accepted claim in the existing literature. Second, it is observed that making use of only a single model involves more risk of producing inaccurate forecasts than is associated with the use of combining methods. This conclusion indicates that combining forecasts is less risky than choosing any single forecasting model. This study accordingly suggests that the tendency of combination methods to produce more accurate, safer forecasts should be taken advantage of by retail sector decision makers when planning related activities.

Third, it has been claimed by a number of studies that, when different single models are considered, using combined forecasting methods tends to result in better forecasts than using any given single model, and that this allows one to avoid the necessity of identifying specifically which individual model might in fact be the best for the given application. A variety of different forecasting models were analysed in this study, and we have reached the similar conclusion with the existing literature that it is unnecessary to identify the best for single models when seeking to improve the forecast accuracy achieved by combined forecasts. Another claim suggests that a combination of very similar models affects forecasting performance positively. This study uses both quite different models such as the ETS, the ANN and the ANFIS and similar models such as the ARIMA and the ARFIMA, producing close forecasts under certain circumstances. Hence, it takes advantage of the truth part of two claims in the literature. Finally, it should be noted that no clear differences in sales forecasting performance were observed among the single models under examination. This conclusion was proven statistically by means of the non-parametric tests carried out. Among the single models, the ETS and the ANFIS models were applied to forecasting retail sales for the first time in this

study. It found that these models produce forecasts that are comparable with the more frequently used ones. Most of the studies in the literature come to the conclusion that the ANN model exhibits better performance than traditional techniques such as the ARIMA models and exponential smoothing methods. Contrary to the common finding, this study has not found that there is a statistically significant difference in performance between the ANN model and traditional forecasting techniques when the overall forecasting performance is taken into consideration for all data sets used.

Some limitations of the study can be described to guide future research. One limitation is that the study does not include any macroeconomic variables informing about the economic conditions and purchasing power in the country where the study is carried out, and any indicators summarising the conditions in the relevant sector. The other limitation is that the company has disclosed the details of its existing forecasting system and information regarding promotional strategies that affect the sales of a given product substantially. Lastly, even if the results of the study have shown that it is possible to attain statistically more accurate sales forecasts by using the methods in the study, a decision support platform which collects all of the abovementioned methods and helps the people in charge in determining stock level is not built into the scope of this study. For future research directions, similar studies could be made in relation to products with pronounced seasonal sales patterns, and products of companies operating in a range of different sectors and markets. In addition, there are some recently developed neural network types such as emotional neural networks, deep neural networks, etc. that have not been examined for sales forecasting. Hence, the performance of these neural networks may be investigated. Further, the forecasting process should take into account the effects of the exogenous variables apart from the lagged values of the series to be forecast, as well as qualitative assessments describing the company's turnover and the share of and position in the market on the forecasts.

References

Aburto, L.; Weber, R. 2007. Improved supply chain management based on hybrid demand forecasts, *Applied Soft Computing* 7(1): 136–144. https://doi.org/10.1016/j.asoc.2005.06.001

Adhikari, R.; Agrawal, R. K. 2014. A combination of artificial neural network and random walk models for financial time series forecasting, *Neural Computing and Applications* 24(6): 1441–1449. https://doi.org/10.1007/s00521-013-1386-y

Agrawal, D.; Schorling, C. 1997. Market share forecasting: an empirical comparison of artificial neural networks and multinomial logit model, *Journal of Retailing* 72(4): 383–407. 0https://doi.org/10.1016/S0022-4359(96)90020-2

Ahmed, N. K.; Atiya, A. F.; Gayar, N. E.; El-Shishiny, H. 2010. An empirical comparison of machine learning models for time series forecasting, *Econometric Reviews* 29(5–6): 594–621. https://doi.org/10.1080/07474938.2010.481556

Aiolfi, M.; Timmermann, A. 2006. Persistence in forecasting performance and conditional combination strategies, *Journal of Econometrics* 135(1): 31–53. https://doi.org/10.1016/j.jeconom.2005.07.015

Alon, I.; Qi, M.; Sadowski, R. J. 2001. Forecasting aggregate retail sales: a comparison of artificial neural networks and traditional methods, *Journal of Retailing and Consumer Services* 8(3): 147–156. https://doi.org/10.1016/S0969-6989(00)00011-4

Ansuj, A. P.; Camargo, M. E.; Radharamanan, R.; Petry, D. G. 1996. Sales forecasting using time series and neural networks, *Computers & Industrial Engineering* 31(1): 421–424. https://doi.org/10.1016/0360-8352(96)00166-0

Aras, S.; Kocakoç, İ. D. 2016. A new model selection strategy in time series forecasting with artificial neural networks: IHTS, *Neurocomputing* 174: 974–987. https://doi.org/10.1016/j.neucom.2015.10.036

Armstrong, J. S. (Ed.). 2001. *Principles of forecasting: a handbook for researchers and practitioners*. Springer Science & Business Media. https://doi.org/10.1007/978-0-306-47630-3

Armstrong, J. S.; Fildes, R. 1995. Correspondence on the selection of error measures for comparisons among forecasting methods, *Journal of Forecasting* 14(1): 67–71. https://doi.org/10.1002/for.3980140106

Au, K. F.; Choi, T. M.; Yu, Y. 2008. Fashion retail forecasting by evolutionary neural networks, *International Journal of Production Economics* 114(2): 615–630. https://doi.org/10.1016/j.ijpe.2007.06.013

Aye, G. C.; Balcilar, M.; Gupta, R.; Majumdar, A. 2015. Forecasting aggregate retail sales: the case of South Africa, *International Journal of Production Economics* 160: 66–79. https://doi.org/10.1016/j.ijpe.2014.09.033

Bagheri, A.; Peyhani, H. M.; Akbari, M. 2014. Financial forecasting using ANFIS networks with quantum-behaved particle swarm optimization, *Expert Systems with Applications* 41(14): 6235–6250. https://doi.org/10.1016/j.eswa.2014.04.003

Barksdale, H. C.; Hilliard, J. E. 1975. A cross-spectral analysis of retail inventories and sales, *The Journal of Business* 48(3): 365–382. https://doi.org/10.1086/295762

Bates, J. M.; Granger, C. W. J. 1969. The combination of forecasts, *Operational Research Quarterly* 20(4): 451–468. https://doi.org/10.1057/jors.1969.103

Bhardwaj, G.; Swanson, N. R. 2006. An empirical investigation of the usefulness of ARFIMA models for predicting macroeconomic and financial time series, *Journal of Econometrics* 131(1): 539–578. https://doi.org/10.1016/j.jeconom.2005.01.016

Bordignon, S.; Bunn, D. W.; Lisi, F.; Nan, F. 2013. Combining day-ahead forecasts for British electricity prices, *Energy Economics* 35: 88–103. https://doi.org/10.1016/j.eneco.2011.12.001

Box, G. E. P.; Jenkins, G. 1970. *Time series analysis, forecasting and control*. San Francisco: Holden-Day, CA.

Boyacioglu, M. A.; Avci, D. 2010. An adaptive network-based fuzzy inference system (ANFIS) for the prediction of stock market return: the case of the Istanbul stock exchange, *Expert Systems with Applications* 37(12): 7908–7912. https://doi.org/10.1016/j.eswa.2010.04.045

Canova, F.; Hansen, B. E. 1995. Are seasonal patterns constant over time? A test for seasonal stability, *Journal of Business & Economic Statistics* 13(3): 237–252.

Çatik, A. N.; Karaçuka, M. 2012. A comparative analysis of alternative univariate time series models in forecasting Turkish inflation, *Journal of Business Economics and Management* 13(2): 275–293. https://doi.org/10.3846/16111699.2011.620135

Chatfield, C. 1988. What is the 'best' method of forecasting?, *Journal of Applied Statistics* 15(1): 19–38. https://doi.org/10.1080/02664768800000003

Chen, F. L.; Ou, T. Y. 2011. Sales forecasting system based on Gray extreme learning machine with Taguchi method in retail industry, *Expert Systems with Applications* 38(3): 1336–1345. https://doi.org/10.1016/j.eswa.2010.07.014

Claveria, O.; Torra, S. 2014. Forecasting tourism demand to Catalonia: neural networks vs. time series models, *Economic Modelling* 36: 220–228. https://doi.org/10.1016/j.econmod.2013.09.024

Clemen, R. T. 1989. Combining forecasts: a review and annotated bibliography, *International*

Journal of Forecasting 5(4): 559–583. https://doi.org/10.1016/0169-2070(89)90012-5

Corder, G. W.; Foreman, D. I. 2014. *Nonparametric statistics: a step-by-step approach*. John Wiley & Sons.

Curtis, A. B.; Lundholm, R. J.; McVay, S. E. 2014. Forecasting sales: a model and some evidence from the retail industry, *Contemporary Accounting Research* 31(2): 581–608. https://doi.org/10.1111/1911-3846.12040

Cybenko, G. 1989. Approximation by superpositions of a sigmoidal function, *Mathematics of Control, Signals and Systems* 2(4): 303–314. https://doi.org/10.1007/BF02551274

De Gooijer, J. G.; Hyndman, R. J. 2006. 25 years of time series forecasting, *International Journal of Forecasting* 22(3): 443–473. https://doi.org/10.1016/j.ijforecast.2006.01.001

De Menezes, L. M.; Bunn, D. W.; Taylor, J. W. 2000. Review of guidelines for the use of combined forecasts, *European Journal of Operational Research* 120: 190–204. https://doi.org/10.1016/S0377-2217(98)00380-4

Du, W.; Leung, S. Y. S.; Kwong, C. K. 2015. A multiobjective optimization-based neural network model for short-term replenishment forecasting in fashion industry, *Neurocomputing* 151: 342–353. https://doi.org/10.1016/j.neucom.2014.09.030

Dwivedi, A.; Niranjan, M.; Sahu, K. 2013. A business intelligence technique for forecasting the automobile sales using Adaptive Intelligent Systems (ANFIS and ANN), *International Journal of Computer Applications* 74(9): 7–13. https://doi.org/10.5120/12911-9383

Fabianová, J.; Kačmáry, P.; Molnár, V.; Michalik, P. 2016. Using a software tool in forecasting: a case study of sales forecasting taking into account data uncertainty, *Open Engineering* 6(1): 270–279. https://doi.org/10.1515/eng-2016-0033

Fildes, R.; Makridakis, S. 1995. The impact of empirical accuracy studies on time series analysis and forecasting, *International Statistical Review/Revue Internationale de Statistique* 63(3): 289–308.

Firmino, P. R. A.; de Mattos Neto, P. S.; Ferreira, T. A. 2015. Error modeling approach to improve time series forecasters, *Neurocomputing* 153: 242–254. https://doi.org/10.1016/j.neucom.2014.11.030

Frank, C.; Garg, A.; Sztandera, L.; Raheja, A. 2003. Forecasting women's apparel sales using mathematical modelling, *International Journal of Clothing Science and Technology* 15(2): 107–125. https://doi.org/10.1108/09556220310470097

Gardner, E. S. 2006. Exponential smoothing: the state of the art – part II, *International Journal of Forecasting* 22(4): 637–666. https://doi.org/10.1016/j.ijforecast.2006.03.005

Granger, C. W. 1989. Invited review combining forecasts – twenty years later, *Journal of Forecasting* 8(3): 167–173. https://doi.org/10.1002/for.3980080303

Granger, C. W. J.; Newbold, P. 2014. *Forecasting economic time series*. Academic Press.

Granger, C. W.; Ramanathan, R. 1984. Improved methods of combining forecasts, *Journal of Forecasting* 3(2): 197–204. https://doi.org/10.1002/for.3980030207

Hagan, M. T.; Demuth, H. B.; Beale, M. H.; De Jesús, O. 1996. *Neural network design*. Boston: PWS publishing company.

Haslett, J.; Raftery, A. E. 1989. Space-time modelling with long-memory dependence: assessing Ireland's wind power resource, *Applied Statistics* 38(1): 1–50. https://doi.org/10.2307/2347679

Hochberg, Y.; Tamhane, A. C. 2009. *Multiple comparison procedures*. Hoboken, NJ: John Wiley & Sons.

Hollander, M.; Wolfe, D. A.; Chicken, E. 2013. *Nonparametric statistical methods*. John Wiley & Sons.

Hosking, J. R. 1981. Fractional differencing, *Biometrika* 68(1): 165–176. https://doi.org/10.1093/biomet/68.1.165

Hyndman, R. J.; Athanasopoulos, G. 2014. *Forecasting: principles and practice*. OTexts.

Hyndman, R. J.; Koehler, A. B.; Ord, J. K.; Snyder, R. D. 2008. *Forecasting with exponential smoothing: the state space approach*. Berlin: Springer. https://doi.org/10.1007/978-3-540-71918-2

Hyndman, R. J.; Koehler, A. B.; Snyder, R. D.; Grose, S. 2002. A state space framework for automatic forecasting using exponential smoothing methods, *International Journal of Forecasting* 18(3): 439–454. https://doi.org/10.1016/S0169-2070(01)00110-8

Hyndman, R.; Khandakar, Y. 2008. Automatic time series forecasting: the forecast package for R, *Journal of Statistical Software* 27(3): 1–22. https://doi.org/10.18637/jss.v027.i03

Jang, J. S. 1993. ANFIS: adaptive-network-based fuzzy inference system, *IEEE Transactions on Systems, Man, and Cybernetics* 23(3): 665–685. https://doi.org/10.1109/21.256541

Jiménez, F.; Sánchez, G.; García, J. M.; Sciavicco, G.; Miralles, L. 2017. Multi-objective evolutionary feature selection for online sales forecasting, *Neurocomputing* 234: 75–92. https://doi.org/10.1016/j.neucom.2016.12.045

Jose, V. R. R.; Winkler, R. L. 2008. Simple robust averages of forecasts: some empirical results, *International Journal of Forecasting* 24(1): 163–169. https://doi.org/10.1016/j.ijforecast.2007.06.001

Keles, D.; Scelle, J.; Paraschiv, F.; Fichtner, W. 2016. Extended forecast methods for day-ahead electricity spot prices applying artificial neural networks, *Applied Energy* 162: 218–230. https://doi.org/10.1016/j.apenergy.2015.09.087

Kim, K. J. 2003. Financial time series forecasting using support vector machines, *Neurocomputing* 55(1): 307–319. https://doi.org/10.1016/S0925-2312(03)00372-2

Kolassa, S. 2016. Evaluating predictive count data distributions in retail sales forecasting, *International Journal of Forecasting* 32(3): 788–803. https://doi.org/10.1016/j.ijforecast.2015.12.004

Kwiatkowski, D.; Phillips, P. C.; 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. https://doi.org/10.1016/0304-4076(92)90104-Y

Lolli, F.; Gamberini, R.; Regattieri, A.; Balugani, E.; Gatos, T.; Gucci, S. 2017. Single-hidden layer neural networks for forecasting intermittent demand, *International Journal of Production Economics* 183: 116–128. https://doi.org/10.1016/j.ijpe.2016.10.021

Lotfi, E.; Darini, M.; Karimi-T, M. R. 2016. Cost estimation using ANFIS, *The Engineering Economist* 61(2): 144–154. https://doi.org/10.1080/0013791X.2015.1104568

Lu, C. J. 2014. Sales forecasting of computer products based on variable selection scheme and support vector regression, *Neurocomputing* 128: 491–499. https://doi.org/10.1016/j.neucom.2013.08.012

Ma, S.; Fildes, R.; Huang, T. 2016. Demand forecasting with high dimensional data: the case of SKU retail sales forecasting with intra-and inter-category promotional information, *European Journal of Operational Research* 249(1): 245–257. https://doi.org/10.1016/j.ejor.2015.08.029

Makridakis, S. 1989. Why combining works?, *International Journal of Forecasting* 5(4): 601–603. https://doi.org/10.1016/0169-2070(89)90017-4

Makridakis, S.; Andersen, A.; Carbone, R.; Fildes, R.; Hibon, M.; Lewandowski, R.; Winkler, R. 1982. The accuracy of extrapolation (time series) methods: results of a forecasting competition, *Journal of Forecasting* 1(2): 111–153. https://doi.org/10.1002/for.3980010202

Mentzer, J. T.; Bienstock, C. C. 1998. *Sales forecasting management: understanding the techniques, systems and management of the sales forecasting process*. SAGE Publications, Incorporated.

Montgomery, D. C.; Jennings, C. L.; Kulahci, M. 2015. *Introduction to time series analysis and forecasting*. John Wiley & Sons.

Nelson, M.; Hill, T.; Remus, W.; O'Connor, M. 1999. Time series forecasting using neural networks: should the data be deseasonalized first?, *Journal of Forecasting* 18(5): 359–367. https://doi.org/10.1002/(SICI)1099-131X(199909)18:5<359::AID-FOR746>3.0.CO;2-P

Nurunnabi, M. 2012. Testing weak-form efficiency of emerging economies: a critical review of literature, *Journal of Business Economics and Management* 13(1): 167–188. https://doi.org/10.3846/16111699.2011.620140

Ord, J. K.; Koehler, A. B.; Snyder, R. D. 1997, Estimation and prediction for a class of dynamic nonlinear statistical models, *Journal of the American Statistical Association* 92(440): 1621–1629. https://doi.org/10.1080/01621459.1997.10473684

Pan, Y.; Pohlen, T.; Manago, S. 2013, Hybrid neural network model in forecasting aggregate US retail sales, *Advances in Business and Management Forecasting* 9: 153–170. https://doi.org/10.1108/S1477-4070(2013)0000009013

Ramos, P.; Santos, N.; Rebelo, R. 2015. Performance of state space and ARIMA models for consumer retail sales forecasting, *Robotics and Computer-Integrated Manufacturing* 34: 151–163. https://doi.org/10.1016/j.rcim.2014.12.015

Rumelhart, D. E.; Hinton, G. E.; Williams, R. J. 1986. Learning representations by back-propagating errors, *Nature* 323: 533–536. https://doi.org/10.1038/323533a0

Schilling, J.; Jarrett, J. E. 2008. Daily variation and predicting stock market returns for the Frankfurter Börse (stock market), *Journal of Business Economics and Management* 9(3): 189–198. https://doi.org/10.3846/1611-1699.2008.9.189-198

Shen, W.; Guo, X.; Wu, C.; Wu, D. 2011. Forecasting stock indices using radial basis function neural networks optimized by artificial fish swarm algorithm, *Knowledge-Based Systems* 24(3): 378–385. https://doi.org/10.1016/j.knosys.2010.11.001

Smith, J.; Yadav, S. 1994. Forecasting costs incurred from unit differencing fractionally integrated processes, *International Journal of Forecasting* 10(4): 507–514. https://doi.org/10.1016/0169-2070(94)90019-1

Stock, J. H.; Watson, M. W. 1999. Cointegration, causality and forecasting, in R. F. Engle, H. White (Eds.). *A comparison of linear and nonlinear models for forecasting macroeconomic time series*. Oxford: Oxford University Press, 1–44.

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. https://doi.org/10.1002/for.928

Svalina, I.; Galzina, V.; Lujić, R.; ŠImunović, G. 2013. An adaptive network-based fuzzy inference system (ANFIS) for the forecasting: the case of close price indices, *Expert Systems with Applications* 40(15): 6055–6063. https://doi.org/10.1016/j.eswa.2013.05.029

Theil, H. 1966. *Applied economic forecasting*. Amsterdam: North-Holland Pub.

Timmermann, A. 2006. Handbook of economic forecasting 1, in G. Elliott, C. Granger, A. Timmermann (Eds.). *Forecast combinations*. North Holland Publishing Company, 135–196.

Wallis, K. F. 2011. Combining forecasts–forty years later, *Applied Financial Economics* 21(1–2): 33–41. https://doi.org/10.1080/09603107.2011.523179

Wilcox, R. R. 2012. *Introduction to robust estimation and hypothesis testing*. Academic Press.

Yang, L.; Li, B. 2016. The combination forecasting model of auto sales based on seasonal index and RBF neural network, *International Journal of Database Theory and Application* 9(1): 67–76. https://doi.org/10.14257/ijdta.2016.9.1.06

Yazdani-Chamzini, A.; Yakhchali, S. H.; Volungevičienė, D.; Zavadskas, E. K. 2012. Forecasting gold price changes by using adaptive network fuzzy inference system, *Journal of Business Economics and Management* 13(5): 994–1010. https://doi.org/10.3846/16111699.2012.683808

Yu, L.; Wang, S.Y.; Lai, K.K.; Nakamori, Y. 2005. Time series forecasting with multiple candidate models: selecting or combining?, *Journal of Systems Science and Complexity* 18(1): 1–18.

Zhang, G. P. 2003. Time series forecasting using a hybrid ARIMA and neural network model, *Neurocomputing* 50: 159–175. https://doi.org/10.1016/S0925-2312(01)00702-0

Zhang, G. P.; Patuwo, B. E.; Hu, M. Y. 2001. A simulation study of artificial neural networks for nonlinear time-series forecasting, *Computers & Operations Research* 28(4): 381–396. https://doi.org/10.1016/S0305-0548(99)00123-9

Zhang, G. P.; Qi, M. 2005. Neural network forecasting for seasonal and trend time series, *European Journal of Operational Research* 160(2): 501–514. https://doi.org/10.1016/j.ejor.2003.08.037

Zhang, G.; Patuwo, B. E.; Hu, M. Y. 1998. Forecasting with artificial neural networks: the state of the art, *International Journal of Forecasting* 14(1): 35–62. https://doi.org/10.1016/S0169-2070(97)00044-7

Zou, H.; Yang, Y. 2004. Combining time series models for forecasting, *International Journal of Forecasting* 20(1): 69–84. https://doi.org/10.1016/S0169-2070(03)00004-9

Serkan ARAS received his Bachelor degree in Econometrics in 2005 from Dokuz Eylul University, Izmir, Turkey and in 2008 obtained an MSc in Econometrics from Dokuz Eylul University. He got his PhD degree in Econometrics, in 2013 from Dokuz Eylul University. Also, he visited School of Computer Science, University of Birmingham, U.K. three semesters as academic researcher during his PhD. His research interests include neural networks, evolutionary algorithms, fuzzy logic and time series forecasting.

İpek DEVECİ KOCAKOÇ received her PhD in Econometrics with a major in Industrial Engineering from Dokuz Eylul University, Turkey. Her recent research interests include Artificial Intelligence, Scientific Programming and Quantitative methods. She is currently a full time professor in Dokuz Eylul University.

Çiğdem POLAT received her Bachelor degree in Industrial Engineering in 2010 from Sakarya University, Turkey and in 2016 obtanied an MSc in Quality Management from Dokuz Eylul University. Her recent interests include Total Quality Management and Quantitative Methods.