CENTRAL BANK OF CYPRUS
EUROSYSTEM

WORKING PAPER SERIES

Forecasting Issues: Ideas of Decomposition and Combination

Marina Theodosiou

June 2010

Working Paper 2010-4

Address
80 Kennedy Avenue
CY-1076 Nicosia, Cyprus

Postal Address
P. O. Box 25529
CY-1395 Nicosia, Cyprus

E-mail
publications@centralbank.gov.cy

Website
http://www.centralbank.gov.cy

Fax
+357 22 378153

Papers in the Working Paper Series may be downloaded from:
http://www.centralbank.gov.cy/nqcontent.cfm?a_id=5755

Forecasting Issues: Ideas of Decomposition and Combination

Marina Theodosiou*

June 2010

Abstract

Combination techniques and decomposition procedures have been applied to time series forecasting to enhance prediction accuracy and to facilitate the analysis of data respectively. However, the restrictive complexity of some combination techniques and the difficulties associated with the application of the decomposition results to the extrapolation of data, mainly due to the large variability involved in economic and financial time series, have limited their application and compromised their development. This paper is a re-examination of the benefits and limitations of decomposition and combination techniques in the area of forecasting, and a contribution to the field with a new forecasting methodology. The new methodology is based on the disaggregation of time series components through the STL decomposition procedure, the extrapolation of linear combinations of the disaggregated sub-series, and the reaggregation of the extrapolations to obtain estimation for the global series. With the application of the methodology to the data from the NN3 and M1 Competition series, the results suggest that it can outperform other competing statistical techniques. The power of the method lies in its ability to perform consistently well, irrespective of the characteristics, underlying structure and level of noise of the data.

Keywords: ARIMA models, combining forecasts, decomposition, error measures, evaluating forecasts, forecasting competitions, time series.

JEL Classification: C53.

---

*Central Bank of Cyprus. The opinions expressed in this paper are those of the author and do not necessarily reflect the views of the Central Bank of Cyprus or the Eurosystem.

Correspondence: Marina Theodosiou, Economic Research Department, Central Bank of Cyprus, 80 Kennedy Avenue, P.O.Box 25529, 1395 Nicosia, Cyprus. E-mail: MarinaTheodosiou@centralbank.gov.cy

1 Introduction

> "*Better predictions remain the foundation of all science...*" (Makridakis and Hibon, 2000)

Forecast accuracy has been a critical issue in areas of financial, economic and scientific modeling, which enthused the proliferation of a vast literature on the development and empirical application of forecasting models (Gooijer and Hyndman, 2006). Nevertheless, these models are just "intentional abstractions of a much more complicated reality" (Diebold and Lopez (1996, p.22)) and rely on historical data to draw upon conclusions about the future. Consequently, they are always prone to estimation error due to model misspecification. Combination techniques have been developed to address this issue of misspecification by exploiting the capabilities of the various forecasting models in capturing specific aspects of the data.

Combination techniques operate by pooling forecasts from various models, in order to enhance and robustify prediction accuracy. The integration of information from different models into one forecast can reduce the estimation error in the prediction significantly (Clemen, 1989; Stock and Watson, 2004; Timmermann, 2006). Nonetheless, the restrictive complexity of some existing combination methods and lack of comprehensive guidelines for their application have been admitted flaws in the literature (Armstrong, 1989; Menezes et al., 2000).

Decomposition procedures on the other hand, can facilitate the analysis by disaggregation of the time series into feature-based sub-series. As suggested in this paper, the isolation of the more important features of the data in distinct sub-series can enhance the forecasting performance of the models used for their estimation. As a consequence, the estimation error obtained from the aggregation of the extrapolated sub-series is reduced relative to the estimation error obtained for the series as a whole. The improvement in accuracy is mainly due to the elimination of any residual variability within the sub-series, which may affect the structure of the individual components and consequently the performance of the forecasting method.

In this paper, such a forecasting method is developed which extrapolates the global series through the individual extrapolations of linear combinations of the sub-series returned from the application of a decomposition procedure, including the residual error component. The new forecasting method makes use of both decomposition procedures and combination techniques. A decomposition procedure from the literature is employed to disaggregate the data into three dominant components namely trend, seasonality and residual error, while a linear combination technique is used to obtain an estimation for the global series. The main underlying idea of the method is that better prediction accuracies can be achieved by subdividing the forecasting problem into smaller parts, and consequently also segregating the degree of complexity of the problem. Those parts are then easier to extrapolate, contributing to higher prediction accuracies, than those obtained from the direct forecast of the global series using a single model.

The new forecasting method is applied to the NN3 (Crone and Nikolopoulos, 2007) and M1 Competition (Makridakis et al., 1982) datasets. The results obtained are benchmarked against the results of four forecasting methods namely ARIMA, Theta, Holt's Damped Trend (hereafter HDT) and Holt-Winters (hereafter HW); to the simple combination of the forecasts obtained from these methods,as well as to a Classical Decomposition forecasting method.

The four statistical forecasting methods included in the analysis can be readily implemented in a software package, and were selected on the basis of their performance in previous forecasting competitions and empirical applications. The statistical software used in this paper is the R-Language (R Development Core Team, 2010) and is free to download from `www.r-project.org`. The `forecast` package (Hyndman, 2010) was used for the implementation of the ARIMA and Holt's Damped Trend method.

The paper unfolds as follows. Section two gives an overview of decomposition and combination techniques, while section three describes the data used in the paper. In section four, the various steps leading to the implementation of the new forecasting method are described in detail. Section five presents the results from the application of the new method on the NN3 and M1 competition data for three different forecast horizons, using rolling origins and recalibrating at each step. Concluding remarks are given in section six.

2 A Synopsis on Decomposition & Combination

2.1 Combination Techniques in Forecasting

Clemen (1989) reported that "forecast accuracy can be substantially improved through the combination of multiple individual forecasts". The same conclusion has been reached in many papers and surveys that followed (see for example Marcellino, 2004; Timmermann, 2006; Zou and Yang, 2004). Furthermore, as found in various forecasting competitions (M, M3 Competition), no single technique can perform consistently well across all time series and across all forecasting horizons (Fildes et al., 1998; Makridakis and Hibon, 2000). Therefore, by combining forecasts, one may reduce the misspecification bias in the individual models and increase prediction accuracy.

The gain in accuracy achieved through combination is due to the strengths and limitations of the individual forecasting methods. Hendry and Clements (2002) offer a formal explanation of this phenomenon. They suggest that forecast combining adds value when the individual forecasting methods are *differentially mis-specified*. This argument is supported in the work of Makridakis (1989), Diebold and Lopez (1996) and Stock and Watson (1999, 2004). Furthermore, by combining, the practitioner avoids the possibility of choosing the worst forecasting method for the particular point in time and, hence, robustifies the estimations across all forecasting horizons (Armstrong et al., 1983; Gooijer and Hyndman, 2006). Another explanation given by Pesaran and Timmermann (2007) and Timmermann (2006) is that individual models react differently to structural changes in the data. As a result, "combinations of forecasts from models with different degrees of adaptability to structural changes will outperform forecasts from individual models" (Timmermann, 2006).

In this paper,the idea of combination is employed on the extrapolated disaggregated subseries to obtain an estimation of the global series.

2.2 Overview of Decomposition

Decomposition techniques were initially developed by Persons (1919) to identify and isolate salient features of a time series such as trend seasonality and cyclical patterns. They have since been used for the analysis of economic data to produce official statistics by various governments and institutions. Some of the most prominent methods in the literature are moving averages-based techniques such as X-11 ARIMA/88 (Dagum, 1988), SABL (**S**easonal

Adjustment at **B**ell **L**aboratories) (Cleveland et al., 1981a,b) and STL (**S**easonal-**T**rend decomposition based on **L**oess smoothing) (Cleveland et al., 1990), multiple regression-based techniques, and methods based on time series ARIMA modeling (Box et al., 1978; Bell and Hillmer, 1984; Gómez and Maravall, 1996) (consult Fischer, 1995, for a well-documented survey on the various methods). There also exists a stream of literature which deals with the extraction of the trend component from seasonally adjusted time series. This includes the parametric methods based on the pivotal work of Kalman (1960), its extensions which incorporate state space representations, initially developed by Rauch63, and models based on the Wiener-Kolmogorov theory (Whittle, 1983) (discussions of such approaches can be found in the books by Harvey (1989); West and Harrison (1997); Kitagawa and Gersch (1996); Durbin and Koopman (2001). In addition, there exist semi-parametric methods based on spline smoothing and mixed models (Ruppert et al., 2003), nonparametric methods based on band-pass filters and wavelet methods (Percival and Walden, 2000); and methods based on kernel estimation and local polynomial modeling (Fan and Gijbels, 1996) (see Mills, 2003, for an extensive review).

Even though decomposition methods were not primarily developed to serve as prediction tools, the intuition behind their application in forecasting is nonetheless very appealing. Disaggregating the various components in the data and predicting each one individually can be viewed as a process of isolating smaller parts of the overall process which are governed by a strong and persistent element, and therefore separating them from any 'noise' and inconsistent variability. These processes are then easier to extrapolate due to their more deterministic nature. It should be therefore possible to obtain more accurate forecasts for the individual components than one is likely to obtain for the global series. This becomes important in the case of time series with a high degree of noise.

There exists a number of papers in the literature which deal with the extrapolation of time series through the extrapolation of the individual components, obtained from the application of averaging techniques (Damrongkulkamjorn and Churueang, 2005; Temraz et al., 1996). This approach to forecasting is known as the classical decomposition technique and was developed by Macaulay (1938) and is described for example in Makridakis et al. (1998). However, in all applications of the classical decomposition technique, the residual component after the elimination of any trend, cyclical and seasonal variations, is always assumed to be a random variable with constant variance and is therefore excluded from the forecasting process.

In the current paper, a new approach to decomposition in forecasting is developed which achieves the forecasting of a time series through the linear combination of its components, including that of the residual error component.

3 Data Description

3.1 NN3 Competition Dataset

The dataset of 111 time series distributed for the NN3 competition was used for the implementation of the new forecasting method. This can be obtained from `http://www.neural-forecasting- competition.com/NN3/datasets.htm`. The competition organizers have not disclosed the source of the dataset, and the only information available is that this is composed of empirical business time series. The data are monthly, with positive observations and structural characteristics which vary widely across the time series. Many series are dominated by a strong seasonal structure, and for some (NN59, NN102, NN103), the seasonality is exhibited with almost zero noise. There are also series exhibiting both trending and seasonal behavior, while in some cases outliers can be detected (e.g. NN108, NN110). Nevertheless, the majority of time series is characterized by a high level of noise, and in some instances this appears to be the dominant component in the series (NN78, NN95, NN96, NN97, NN99, NN108, NN110). The length of the various data ranges from 68 to 144 monthly observations. From these, the last 18,12 and 1 observations are withheld for evaluating the predictive ability of the new forecasting method across three different forecasting horizons. The time series are not subjected to any data preprocessing prior to the implementation of the forecasting method.

The large variability of structural characteristics within the 111 time series underlines the need for a single forecasting method that could predict all series with a relatively high level of accuracy, and consequently, remain unaffected by structural changes and persistent trending or seasonal behavior in the data. In the following section, such a method is described, which is based on the individual unobserved components within each observed time series and thus possesses the capability of attaining high levels of predictive accuracy irrespective of the structural attributes of the underlying data.

3.2 M1 Competition Dataset

The performance of this new method is also tested on the complete and reduced datasets of the M1 Competition (Makridakis et al., 1982). The complete dataset consists of 1001 time series of economic and financial indicators (micro, macro and demographic), from which 181 are of annual frequency, 203 of quarterly frequency and 617 of monthly frequency. The reduced dataset consists of 111 series, analyzed in Makridakis et al. (1982) and is composed of 20 annual, 23 quarterly and 68 monthly series. These datasets have been extensively documented in the literature and have become a standard test data for the evaluation of forecasting techniques. Figure 1 depicts some example time series from the three datasets.

4 Methodology Description

In this section, the various steps for the implementation of the new forecasting method are described in detail. The subdivision of the forecasting problem into smaller parts is achieved through a decomposition procedure which disaggregates the global series x_t into three additive components, namely trend (m_t), seasonality (s_t) and error (e_t), i.e.

$$x_t = m_t + s_t + e_t \tag{1}$$

4.1 The Decomposition Procedure

The *Seasonal and Trend Decomposition using Loess* (STL) procedure (Cleveland et al., 1990) is used for the additive decomposition of the global time series. STL performs additive decomposition of the data through a sequence of applications of the Loess smoother, which

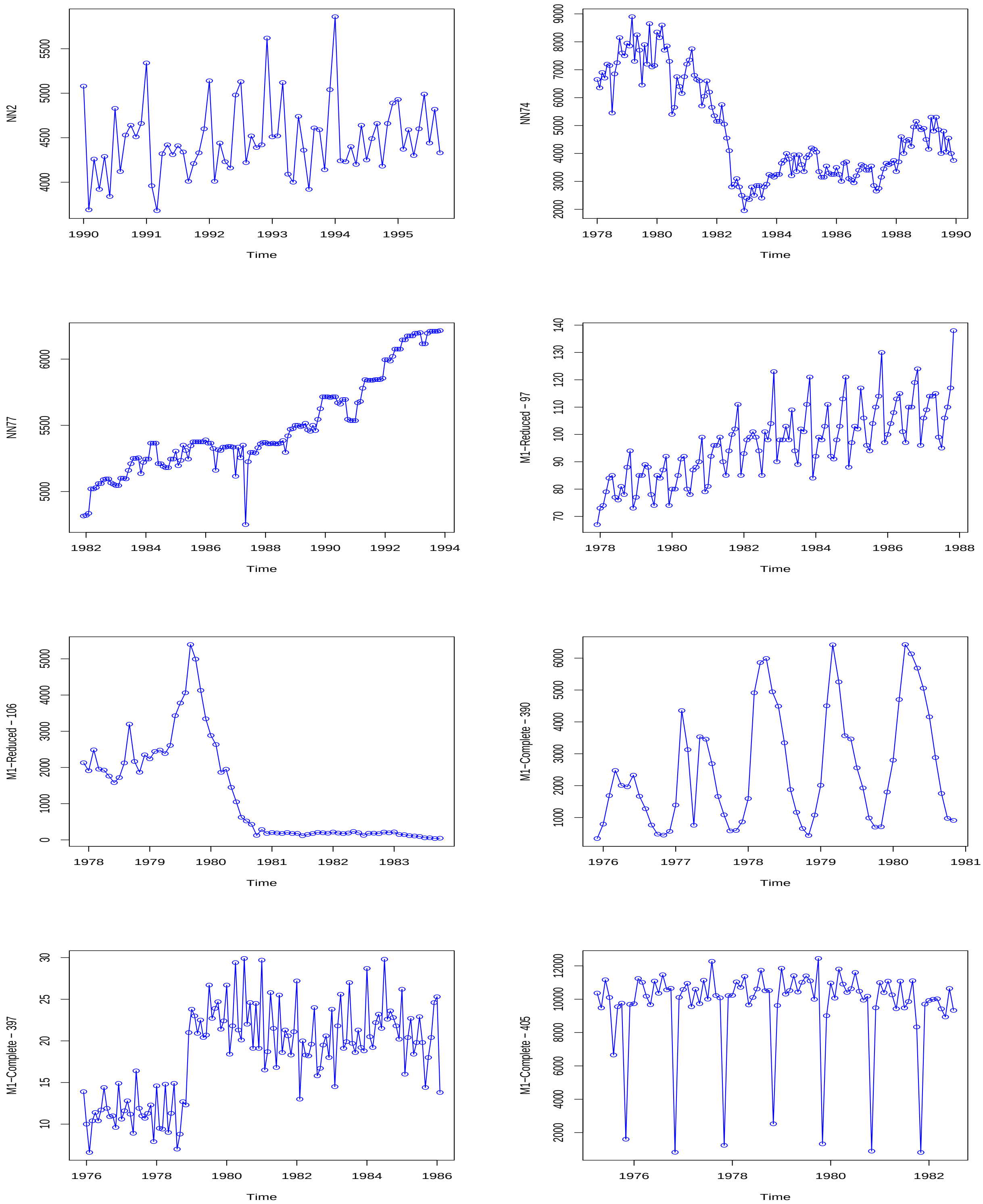


Figure 1: Time series plots for the 2^{nd}, 74^{th} and 77^{th} time series of the NN3 Competition, 97^{th} and 106^{th} time series of the M1 Competition reduced dataset, and 390^{nd}, 397^{th} and 405^{th} time series of the M1 Competition complete dataset.

applies locally weighted polynomial regressions at each point in the data set, with the explanatory variables being the values close to the point whose response is estimated. The STL decomposition procedure was chosen amongst other decomposition techniques in the literature as it presents important advantages for extensive applications to a large number of time series, which is also the scope of the paper. The most attractive feature of the STL, relative to other decomposition procedures, is its strong resilience to outliers in the data, and consequently its resulting in robust component sub-series. This is a particularly important attribute for the scope of the forecasting method proposed in the paper, as robustness of components series can translate to enhanced predictive accuracy for the forecasting methods applied to these sub-series. Furthermore, the procedure treats the results from neighboring time points as independent and therefore, does not constrain the seasonal pattern to take a particular form. In addition, unlike other decomposition techniques, STL is capable of handling seasonal time series with any seasonal frequency greater than one, and thus it is not restricted only to monthly and quarterly data. This is again a very appealing quality of the decomposition method, given that the motivation behind the new forecasting method was for it to be applicable to a wide range of time series with different characteristics, and hence a diverse set of sampling frequencies. Moreover, the implementation of the STL procedure is based purely on numerical methods and does not require any mathematical modeling. This contributes to the method being very easy to implement for a large number of time series. This attribute also suggests that the STL method can be applied to a large number of time series without requiring time invested in modeling the properties of each of the time series involved in the analysis.

The procedure is carried out in an iterated cycle of detrending and then updating the seasonal component from the resulting sub-series. At every iteration the robustness weights are formed based on the estimated irregular component; the former are then used to down-weight outlying observations in subsequent calculations. Therefore, the iterated cycle is composed of two recursive procedures, the inner and the outer loop. Each pass of the inner loop applies seasonal smoothing that updates the seasonal component, followed by trend smoothing that updates the trend component.

An iteration of the outer loop consists of one iteration of the inner loop with resulting estimates of the trend and seasonal components used to calculate the irregular component. Any large values in e_t are identified as extreme values and a weight is calculated. This concludes the outer loop. Further iterations of the inner loop use the weights to down-weight the effect of extreme values, identified in the previous iteration of the outer loop. For a more detailed description of the STL decomposition procedure, the reader is referred to Cleveland et al. (1990).

In the current application, each time series was tested for outliers prior to the implementation of the STL procedure. The detection of outliers was based on the equation:

$$\left| \frac{X_t - \mu_t}{\sigma_t} \right| > 2 \tag{2}$$

where μ_t and σ_t denote the mean and standard deviation of the time series X_t. If no outliers are detected, the number of iterations for the outer loop is set to 0.

Thus, for every time series x_t, STL returns, m_t, s_t and e_t, as in equation (1). The STL decomposition procedure can be readily implemented in R-Language (R Development Core

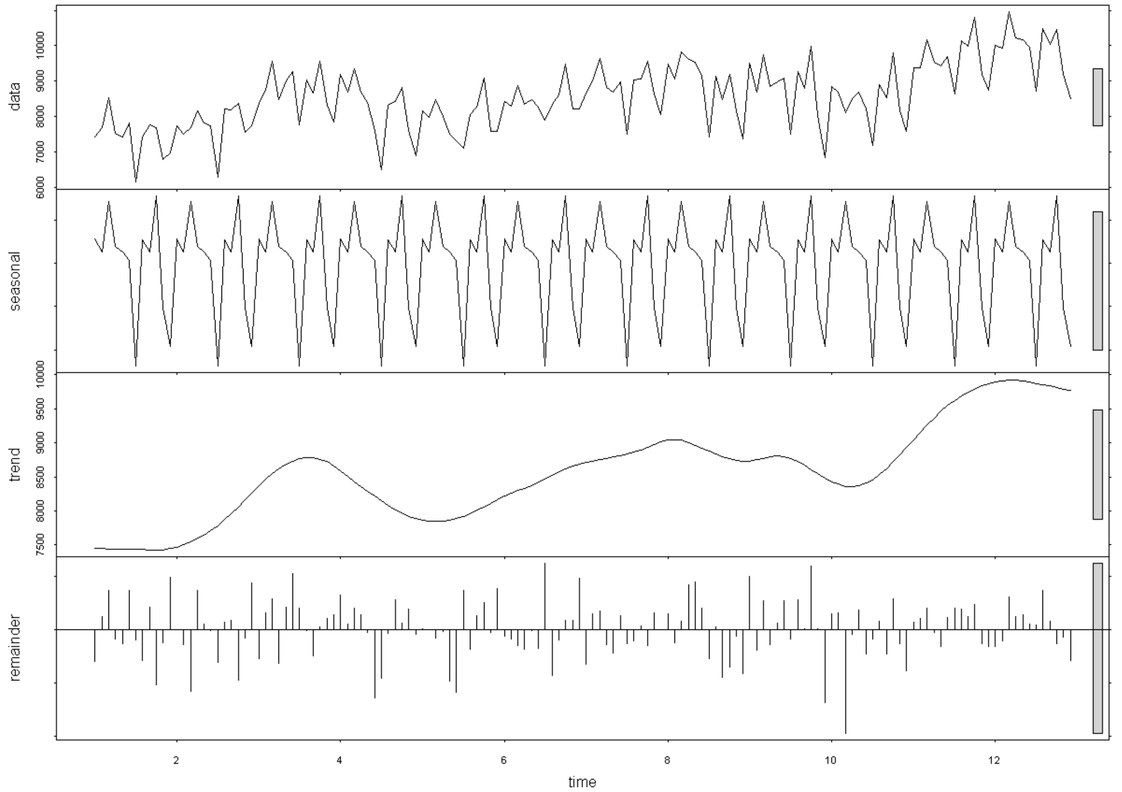


Figure 2: Results from the application of the STL decomposition procedure on time series NN52.

Team, 2010) using the function `stl()`. Figure 2 depicts the results from the application of the STL decomposition procedure on series NN52 from the NN3 Competition dataset.

4.2 Forecasting Methods for the Disaggregated Components

The success of the new forecasting method therefore relies on the successful interpolation of linear combinations of the additive components. Below, a description of the analysis carried out for choosing suitable forecasting models for the extrapolation of the individual components is given together with the main conclusions from the analysis.

In order to obtain some guidance as to which forecasting method is optimal for the extrapolation of each individual component, four forecasting methods namely ARIMA (Box and Jenkins, 1976), Theta (Assimakopoulos and Nikolopoulos, 2000), HDT (Holt, 2004) and HW (Winters, 1960), were applied to the raw data using rolling origins to predict 6, 12 and18 observations ahead, using only the first 36 observations (3 years) in the sample. Therefore, only the first 42, 48 and 54 observations from each time series are used in this analysis for each forecast horizon respectively. Their performance was evaluated based on prediction error and relative to the dominant component and the level of noise in the data.

As mentioned before, these forecasting methods were selected based on their performance in forecasting competitions and other empirical applications, as well as on their ability to capture salient features of the data. Exponential smoothing methods such as HDT and HW have been examined extensively in the literature and were reported to perform well for a wide range of data (Satchell and Timmermann, 1995; Hyndman et al., 2002; Chatfield et al., 2001; Hyndman et al., 2005). The Theta method which, as shown by Hyndman and Billah (2003) is simple exponential smoothing with drift, was the best performing method in the M3-Competition (Makridakis and Hibon, 2000), and was reported as the second best statistical method for the NN3 Competition after Wildi (Crone and Nikolopoulos, 2007). Finally, ARIMA models are very popular in the literature for their robustness to model misspecification (Chen, 1997). Here, the stepwise selection procedure described in Hyndman and Khandakar (2008) was used for choosing the optimal ARIMA model for each of the time series considered. In addition, the automatic algorithms described in the same paper, were used to choose the optimal parameters for the implementation of the other three forecasting techniques.

The 'best' method for each time series, in terms of mean absolute scaled error (MASE) (Hyndman and Koehler, 2006), was recorded and examined relative to the structural components in the time series. Firstly, in order to determine the strength of each component in the time series, these are regressed against the original data and the coefficients of determination from each individual regression are obtained, i.e., x_t is regressed against m_t, s_t and e_t and the coefficients of determination, R^2_m,R^2_s and R^2_e, are obtained respectively. R^2 provides an indication of the 'strength' of each component in the series. Therefore, the higher the R^2 the greater the power of the component in predicting x_t. Hence, the following regressions were carried out:

$$\begin{array}{ccccccl} x_t & = & \alpha_m & + & \beta_m m_t & + & \epsilon_{t,m} \implies R^2_m \\ x_t & = & \alpha_s & + & \beta_s s_t & + & \epsilon_{t,s} \implies R^2_s \\ x_t & = & \alpha_e & + & \beta_e e_t & + & \epsilon_{t,e} \implies R^2_e \end{array} \tag{3}$$

Secondly, the time series are classified into four groups based on the best forecasting method for each time series. Hence, those time series for which method M, for $M = 1, \ldots, 4$ (HW, HDT, Theta and ARIMA respectively), was found to be the best method in terms of MASE were grouped together. The coefficients of determination of the three components from each time series in each group are then recorded into a matrix G_M. Therefore, each group of time series for which method M recorded the smallest MASE was associated with a matrix, G_M, of $n \times 3$ coefficients of determination, n being the number of time series for which method M had the smallest MASE amongst the four forecasting methods examined, i.e.,

$$G_M = \begin{pmatrix} R^2_{s,1} & R^2_{m,1} & R^2_{e,1} \\ \vdots & \vdots & \vdots \\ R^2_{s,n_M} & R^2_{m,n_M} & R^2_{e,n_M} \end{pmatrix} \tag{4}$$

The purpose of this classification was to determine the relationship between the performance of each individual forecasting method with respect to the structural features of the series. Hence, the column means of the G_M matrix provide an indication of the most dominant components in the time series for which method M returned the smallest error. The numeric results obtained for the four forecasting methods and the three forecasting horizons investigated are shown in the Table 1. From the table some important conclusions can be drawn:

	$\overline{R}_s^2$	$\overline{R}_m^2$	$\overline{R}_e^2$
	$h = 18$		
G_{HW}	0.50	0.12	0.34
G_{HDT}	0.27	0.39	0.31
G_{TH}	0.28	0.39	0.31
G_{AR}	0.52	0.14	0.27
	$h = 12$		
G_{HW}	0.43	0.19	0.33
G_{HDT}	0.26	0.39	0.32
G_{TH}	0.25	0.26	0.39
G_{AR}	0.43	0.21	0.30
	$h = 6$		
G_{HW}	0.36	0.20	0.36
G_{HDT}	0.22	0.46	0.27
G_{TH}	0.25	0.22	0.43
G_{AR}	0.44	0.26	0.31

Table 1: Column means for each of the four matrices G_M.

- Time series, for which the ARIMA and HW methods were optimal, exhibited a stronger seasonal component.
- Time series, for which the Theta, HDT and ARIMA methods were optimal, exhibited a stronger trend component.
- Time series, for which the Theta and HW methods were optimal, exhibited a stronger error component.

4.3 Extrapolating the Error Component

The most important step in the application of the new forecasting method lies in the estimation of the error component. Being the residual variability after the elimination of any structural component in the data (trend and seasonality), it is a very noisy series and therefore very difficult to predict. To our knowledge, there exists no published work in the literature that deals with the extrapolation of the irregular component obtained through the application of a decomposition procedure, using statistical techniques.

It is therefore customary in the literature to assume that the error component, obtained from the application of a decomposition procedure, is white noise, and always exclude it from the forecasting procedure. Nevertheless in the current application, the error component is believed to contain predictive information in its sub-series and that discarding it completely can negatively affect estimation accuracy. As seen from Table 1, the coefficient of determination of the error component was greater than 0.25 for all four groups of time series and for all forecasting horizons considered, implying that the error component can account for approximately 25% of the predictability in the series. Consequently, one can assume that the error component, obtained from the application of the STL decomposition procedure, is not residual noise. The information contained in the error component might be in the form of residual autocorrelation in its series, or of conditional dependence on the other decomposed features of the original time series.

Based on this intuition, the error component is also included in the estimation of the global series, through a combination technique which is based on the extraction of the error component from the extrapolated detrended and deseasonalised series, $\widehat{se}_{t+1}$ and $\widehat{me}_{t+1}$. These are obtained by adding together the seasonality and error, and trend and error components respectively, i.e:

$$se_t = s_t + e_t \tag{5}$$

$$me_t = m_t + e_t \tag{6}$$

The combinations of $\widehat{se}_{t+1}$ with the trend component, and $\widehat{me}_{t+1}$ with the seasonal component both give an estimation for the global series, i.e.:

$$\hat{x}_{t+1} = \widehat{se}_{t+1} + \hat{m}_{t+1} \tag{7}$$

$$\hat{x}_{t+1} = \widehat{me}_{t+1} + \hat{s}_{t+1} \tag{8}$$

4.4 Forecasting Method

In this paper, the HW method was used for the estimation of se_t, the seasonality component was extrapolated using the ARIMA method and the trend component using the Theta method. The choice of these methods was supported by the preliminary analysis carried out in section 4.2. From the analysis, it was found that the aforementioned methods were optimal for time series which exhibited a stronger trend component (Theta) and a stronger seasonal component (HW and ARIMA), in comparison with the time series for which the other employed methods were optimal.

For the estimation of the deseasonalized sub-series, me_t, the ARIMA method was used even though HDT and Theta methods were also found to be optimal for time series which exhibited a stronger trend component than that found in time series for which HW was optimal. The choice of the ARIMA method was guided by the fact that the number of time series for which HDT and Theta methods were optimal, in terms of MASE, was significantly smaller relative to the number of time series for which ARIMA was optimal.

Hence, by combining the extrapolated seasonal, trend, seasonal and error and trend and error components, one can obtain the estimation for the global series x_t:

$$\hat{x}_{t+1} = (\hat{m}^{(Th)}_{t+1} + \hat{s}^{(AR)}_{t+1} + \hat{me}^{(AR)}_{t+1} + \hat{se}^{(HW)}_{t+1})/2 \tag{9}$$

The new forecasting method is therefore based on the linear combination of the extrapolated sub-series. Accordingly, there is an element of originality in the method developed. That is, the forecasts included in the combination are not direct forecasts of the target series, but are forecasts of sub-series of the individual components, which approximate its behavior. Therefore, each sub-series is governed by a different structural characteristic and hence, a different forecasting method is used for its estimation. This aspect of distinguishability in the individual sub-series is what creates value in the combination framework; a conclusion which is also supported in the literature (Hendry and Clements, 2002). The method can be implemented in the R-language using the `stlforecast` function in the `forecast` package (Hyndman, 2010).

5 Application

5.1 Performance Evaluation

The predictive performance of the proposed forecasting method, described in equation (9), is evaluated using data from the NN3 and M1 forecasting Competitions. Both data are characterized by a wide range of time series with different structural characteristics.

The analysis of the new forecasting method on the M1 Competition data is limited to the quarterly and monthly time series. Annual data is excluded from the analysis, as two is the minimum time series frequency required by the STL decomposition method. Series with less than 36 observations are also excluded, on the basis that 36 is the minimum number of observations required by HW method for estimating a seasonal time series. The resulting M1 Competition data consists of 729 and 76 time series for the complete and reduced sample, respectively.

The performance of the new forecasting method is benchmarked against that of the four aforementioned statistical forecasting methods namely HW, HDT, Theta and ARIMA. The performance evaluation is carried out using the last 18, 12 and 1 observations in the sample, for monthly data, and the last 8, 4 and 1 observations, for quarterly data. The predictions are calculated using rolling origins and recalibrating the forecasts at each step (Tashman, 2000).

The results from the application of the new forecasting method are also compared to those obtained from the simple combination of forecasts of the four constituent forecasting methods (HW, HDT, Theta and ARIMA). It is widely stated in the literature, and also evidenced by past forecasting competitions, that an equal weighted combination of forecasts can outperform, in terms of forecast accuracy, the individual methods involved in the combination. In the NN3 Competition for example, a combination of a single, Holt and damped exponential smoothing was ranked amongst the top 10 performers of the competition.

Moreover, the performance of the proposed method is compared to that of a Classical Decomposition method obtained from the simple summation of the extrapolated trend and seasonal components; thus excluding from the forecast procedure the estimation of the residual component, as this is achieved using the method described in section 4.3. Therefore, the h-step ahead forecast is computed by the simple summation of the extrapolated trend and seasonal components obtained from the application of the Theta and ARIMA models respectively:

$$\hat{x}_{t+h} = \hat{m}_{t+h}^{(Th)} + \hat{s}_{t+h}^{(AR)} \tag{10}$$

This comparison was carried out to investigate the gains, in terms of predictive accuracy, from including the error component in the forecasting process.

A set of measures were adopted to evaluate the performance of the various forecasting methods. These can be categorized in scale-independent, scale-dependent, scaled, symmetric and relative. Table 2 gives a list of error measures examined under the five evaluation categories. x_t denotes the real observation and $\hat{x}_t$ the predicted observation at time t. Also,

$$\begin{aligned} \epsilon_t &= x_t - \hat{x}_t && (11) \\ q_t &= \frac{\epsilon_t}{\frac{1}{n-1}\sum_{i=2}^{n}|x_i - x_{i-1}|} && (12) \\ r_t &= \frac{\epsilon_t}{\epsilon_t^*} \end{aligned}$$

n is the number of observations in the data and ϵ_t^* is the forecast error obtained from a benchmark model. In this paper, the benchmark model was taken to be the random walk model, where $\hat{x}_{t+h}$ is equal to the last observation, x_t, h being the forecast horizon.

A. Scale-Independent Measures		
MAPE	Mean Absolute Percentage Error	$mean(100\lvert\frac{\epsilon_t}{x_t}\rvert)$
MdAPE	Median Absolute Percentage Error	$median(100\lvert\frac{\epsilon_t}{x_t}\rvert)$
RMSPE	Root Mean Square Percentage Error	$\sqrt{mean((100\frac{\epsilon_t}{x_t})^2)}$
RMdSPE	Root Median Square Percentage Error	$\sqrt{median((100\frac{\epsilon_t}{x_t})^2)}$
B. Scale-Dependent Measures		
MAE	Mean Absolute Error	$mean(\lvert\epsilon_t\rvert)$
MdAE	Median Absolute Error	$median(\lvert\epsilon_t\rvert)$
MSE	Mean Square Error	$mean(\epsilon_t^2)$
RMSE	Root Mean Square Error	$\sqrt{MSE}$
C. Scaled Errors		
MASE	Mean Absolute Scaled Error	$mean(\lvert q_t\rvert)$
MdASE	Median Absolute Scaled Error	$median(\lvert q_t\rvert)$
RMSSE	Root Mean Squared Scaled Error	$\sqrt{mean(q_t^2)}$
D. Symmetric Errors		
sMAPE	Sym. Mean Abs. Perc. Error	$mean\left(200\frac{\lvert x_t-\hat{x}_t\rvert}{(x_t+\hat{x}_t)}\right)$
sMdAPE	Sym. Median Abs. Perc. Error	$median\left(200\frac{\lvert x_t-\hat{x}_t\rvert}{(x_t+\hat{x}_t)}\right)$
E. Relative Error Measures		
MRAE	Mean Relative Absolute Error	$mean(\lvert r_t\rvert)$
MdRAE	Median Relative Absolute Error	$median(\lvert r_t\rvert)$
GMRAE	Geometric Mean Rel. Abs. Error	$gmean(\lvert r_t\rvert)$

Table 2: List of error measures employed for the performance evaluation of the new forecasting method

Scale-independent measures are suitable for measuring forecast performance between different data series. These have been recommended among others by Tashman (2000). However, as also pointed out by Hyndman (2006), this class of error measures can take infinite or undefined values if there are zero values in a series, and therefore, its distribution can be badly skewed. Scale-dependent measures are based on the variability of the predictions when compared to the real observations, and therefore, are useful when comparing methods for the same data set. Nonetheless, these should not be used when comparing across data sets with different scales (Hyndman and Koehler, 2006).

Relative error measures compare the error in the forecasts with the error of a benchmark model. These have been supported in the literature as the most reliable error measures for a large number of applications (Armstrong and Collopy, 1992; Fildes, 1992; Thompson, 1990, 1992). Nevertheless, this class of error measures would result in infinite values if very small errors are computed for the benchmark method. In the current application, the relative error measures are winsorized to avoid this problem. Scaled measures scale the error based on the in-sample MAE from the naïve method and are independent of the scale of the data. They have been recommended by Hyndman and Koehler (2006). Specifically, they recommended MASE (Mean Absolute Scaled Error) "to become the standard measure for forecast accuracy" due to the fact that this is always defined and finite, unlike other measures under certain conditions.

Symmetric errors were the principal error measures employed in the NN3 competition to evaluate forecast performance across each forecasting horizon and across time series. However, it should be noted that this class of error measures has been criticized for not being as symmetric as its name implies. On the contrary, it has been stated in the literature that positive and negative errors have different effects under this measure. In addition, it was also shown by Goodwin and Lawton (1999) that 'symmetric' error measures suffer from instabilities in that, under certain conditions, larger actual errors lead to smaller sMAPE (see also Koehler, 2001).

The results from the performance evaluation analysis of the proposed forecasting method are reported in table 3 for the three datasets and for four error measures, namely MdAPE, MASE, sMAPE and MdRAE, representing the scale-independent, scaled, symmetric and relative error measure categories, respectively (results for the other error measures are available online). In order to avoid the possibility of negative values in the denominator of the sMAPE measure, in the current application the absolute value of the elements in the denominator is considered instead, i.e.

$$mean\left(200\frac{|x_t - \hat{x}_t|}{(|x_t| + |\hat{x}_t|)}\right). \tag{13}$$

The average ranking of each method across the 111, 729 and 76 time series, for the NN3, M1 complete and M1 reduced datasets respectively, is reported for all three forecasting horizons investigated, namely 18, 12 and 1 steps-ahead for monthly data and 8, 4 and 1 steps-ahead for quarterly data. *Comb* denotes the simple combination of the four statistical forecasting methods (HW, HDT, Theta and ARIMA), *Classic* denotes the method described in equation (10), which is based on a Classical Decomposition forecasting method, and *New* denotes the proposed forecasting method given in equation (9). For each error measure investigated, the entry with the smallest value is set in boldface and marked with an asterisk, and the entry with the second smallest entry is highlighted in bold.

The results obtained for the NN3 Competition dataset suggest that the proposed forecasting method can outperform HW, HDT, Theta, ARIMA and the Classical Decomposition method, for 18 and 12 steps-ahead forecasts. The gains from the implementation of the new forecasting method are particularly emphasized in its out-performance of the simple combination method, which appears to be the most accurate forecasting method out of the four other statistical methods investigated. In a pairwise comparison of the new forecasting method with the simple combination method, in terms of the percentage of times one method is more accurate than the other across the 111 times series, it was shown that the proposed method resulted in a larger percentage of more accurate time series than the simple combination method; with the exception of the results obtained under the MASE measure, for which the proposed method is only slightly outperformed by the simple combination method (see table 1 in the web appendix). HW, HDT and the Theta method were ranked the worst performing methods for the two longest forecast horizons investigated.

For the one-step ahead forecast horizon, the new method is outperformed by the other methods, with the exception of the Classical Decomposition method which appears to be the worst performing method for the shortest forecast horizon. The out-performance of the Classical Decomposition method by the proposed method further emphasizes the importance of including the residual error term in the forecasting process. The simple combination method achieved the highest ranking for the one-step ahead forecasts, followed by HW method. Consequently, one can deduce that the comparative advantage of the new forecasting method in predicting long-lead times does not hold for short-term horizons.

The conclusions drawn from the performance evaluation analysis of the proposed method on the NN3 Competition data are further reinforced by the results obtained on the M1 Competition reduced and complete datasets. It is deduced from the rankings that, for the longest forecast horizons of 18 (monthly) and 8 (quarterly), and 12 (monthly) and 4 (quarterly) steps-ahead forecasts, the new forecasting method returned consistently a smaller error in comparison to the other six forecasting methods considered. This holds true for both the reduced and complete datasets investigated and for all four error measure employed, with the exception of the MdAPE measure for the 18 (monthly) and 8 (quarterly) steps-ahead horizon for the reduced dataset, for which the proposed method is ranked second after the ARIMA method.

For the shortest horizon of one-step ahead forecast, the proposed method is ranked third, following the combination and HW method, and the combination and ARIMA method for the complete and reduced dataset respectively. HDT, Theta and the Classical Decomposition forecasting method appear to be the worst performing methods for the M1 Competition datasets for all three forecast horizons considered in the analysis.

It should be noted that the good performance of ARIMA contrasts with claims made in the forecasting competition literature (e.g Makridakis and Hibon, 2000; Armstrong, 2006). An explanation for the conflict in findings is that the performance of the ARIMA method strongly relies on the "correct" identification of the underlying process in the data. In the current application, the automatic model selection algorithm proposed by Hyndman and Khandakar (2008) was used. This automatic model selection algorithm was also implemented by Athanasopoulos et al. (2011) for the tourism forecasting competition with the authors claiming that it was "the first time in the empirical forecasting literature that an ARIMA based algorithm has produced forecasts as accurate as, or more accurate than, those of its competitors".

Figures 3, 4 and 5 present a graphical depiction of the performance of each forecasting method on each of the NN3 Competition, M1 Competition complete and M1 Competition reduced datasets, respectively. For each figure, the distribution of the MdAPE measure is summarized in seven notched boxplots, one for each of the seven methods included in the analysis, and in three different plots, one for each of the three forecasting horizons employed.

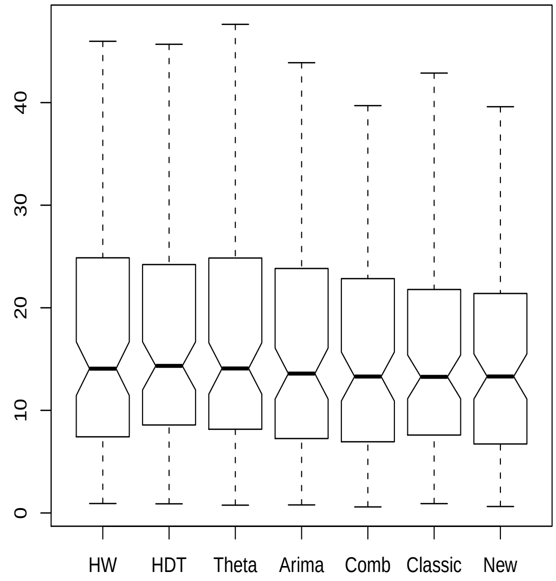
Outliers were omitted from the graph to better facilitate the graphical comparison between the various forecasting methods.

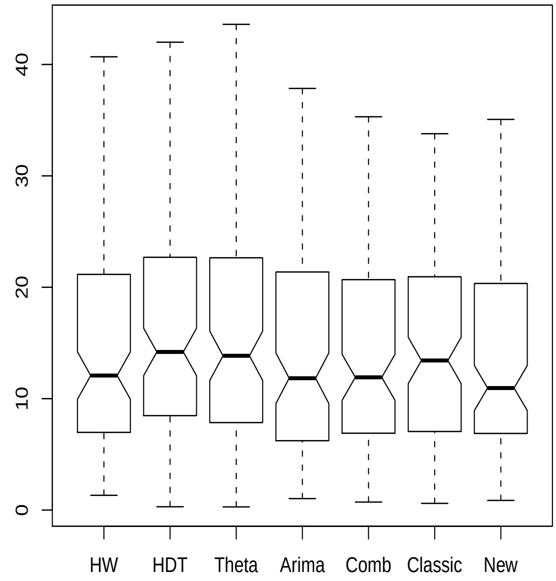
The various graphs confirm the conclusions drawn from the numerical results above. It is evident from figure 3 that the new forecasting method is the most accurate method in terms of the MdAPE measure for long forecasting horizons (18 and 12 steps-ahead), resulting in the lowest median, upper and lower quartiles for the distribution of errors amongst the six methods investigated, while it is the third best method after the simple combination and ARIMA methods for one step-ahead forecasts. HW, HDT and Theta are the least accurate methods in the sample. Furthermore, the HW method returned consistently the largest outliers. For the M1 Competition reduced and complete datasets, figures 4 and 5 indicate that the new method competes closely with ARIMA and HW methods as the most accurate methods for these datasets.

It can be concluded from the analysis that the new forecasting method provides important gains for long-term forecasts, in comparison to the other forecasting methods investigated. As evidenced by the results above, the new forecasting method can predict with a relatively high level of accuracy a larger range of time series than standard forecasting methods in the literature. Being based on the individual extrapolations of the trend and seasonal components in the data, the proposed method can ascertain enhanced robustness for longer lead times and for a wider sample of time series with varying characteristics. Additionally, the out-performance of the Classical Decomposition method confirms the assumption of the proposed method, in that the inclusion of the error component in the forecasting process of a combination technique, based on component sub-series, can improve forecast accuracy.

	NN3 Competition				**M1 Competition (Complete)**				**M1 Competition (Reduced)**			
	MdAPE	MASE	sMAPE	MdRAE	MdAPE	MASE	sMAPE	MdRAE	MdAPE	MASE	sMAPE	MdRAE
	h=18				**h=18 (monthly)/h=8 (quarterly)**				**h=18 (monthly)/h=8 (quarterly)**			
HW	4.58	4.82	4.92	4.14	3.85	3.91	3.96	3.94	3.51	3.70	3.70	3.68
HDT	4.63	4.50	4.48	4.50	4.84	4.83	4.82	4.63	5.01	4.95	4.91	4.71
Theta	4.55	4.39	4.42	4.15	4.54	4.43	4.41	4.39	4.97	4.86	4.82	4.55
Arima	3.89	3.68	3.78	**3.66**	**3.54**	3.68	**3.66**	3.80	**3.21***	**3.41**	**3.42**	3.78
Comb	**3.37**	**3.14***	**3.14***	**3.12***	3.62	**3.61**	**3.66**	**3.70**	3.54	3.50	3.51	**3.66**
Classic	3.89	3.90	3.89	3.70	4.34	4.23	4.24	4.03	4.36	4.24	4.24	4.03
New	**3.09***	**3.57**	**3.36**	4.73	**3.27***	**3.29***	**3.26***	**3.51***	**3.39**	**3.36***	**3.41***	**3.59***
	h=12				**h=12 (monthly)/h=4 (quarterly)**				**h=12 (monthly)/h=4 (quarterly)**			
HW	4.27	4.49	4.45	3.92	3.67	3.75	3.72	3.75	**3.47**	3.59	**3.50**	3.50
HDT	4.58	4.37	4.37	4.39	4.90	4.88	4.90	4.75	5.05	5.05	5.08	4.92
Theta	4.40	4.04	4.13	4.16	4.69	4.65	4.65	4.48	4.86	4.70	4.75	4.54
Arima	3.86	3.89	4.03	**3.58**	**3.42**	**3.45**	**3.45**	**3.60**	3.49	3.58	3.58	3.83
Comb	**3.53**	**3.37***	**3.42***	**3.45**	3.65	3.57	3.60	3.65	3.66	**3.55**	3.51	**3.49**
Classic	3.94	4.02	4.04	3.81	4.42	4.41	4.41	4.35	4.30	4.34	4.37	4.42
New	**3.43***	**3.83**	**3.57**	4.69*	**3.25***	**3.30***	**3.26***	**3.41***	**3.17***	**3.18***	**3.21***	**3.30***
	h=1				**h=1**				**h=1**			
HW	**3.88**	**3.88**	**3.86**	**3.91**	**3.61**	**3.61**	**3.61**	**3.57**	3.66	3.66	3.68	3.67
HDT	3.95	3.95	3.94	3.99	4.50	4.50	4.49	4.51	4.39	4.39	4.38	4.36
Theta	3.98	3.98	3.96	4.03	4.56	4.56	4.56	4.61	4.89	4.89	4.92	4.88
Arima	3.96	3.96	3.98	**3.91**	3.67	3.67	3.67	3.67	**3.54**	**3.54**	**3.53**	**3.53**
Comb	**3.46***	**3.46***	**3.46***	**3.46***	**3.58***	**3.58***	**3.58***	**3.56***	**3.50***	**3.50***	**3.50***	**3.48***
Classic	4.63	4.63	4.63	4.62	4.43	4.43	4.44	4.40	4.33	4.33	4.32	4.37
New	4.13	4.13	4.16	4.08	3.66	3.66	3.65	3.66	3.68	3.68	3.67	3.71

Table 3: Average ranking for MdAPE, MASE, sMAPE and MdRAE measures across three forecasting horizons (18, 12 and 1).





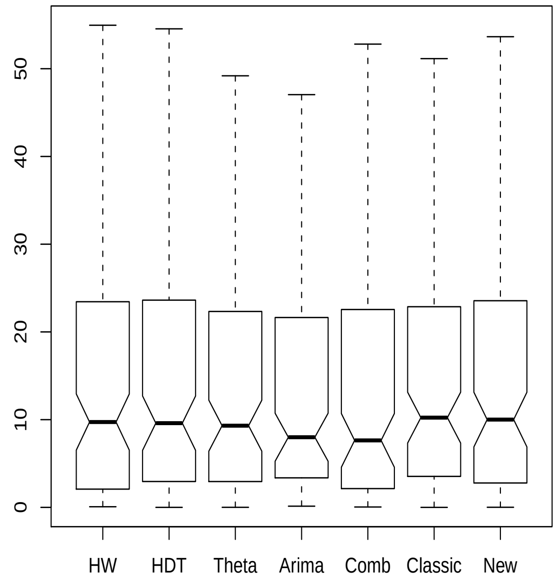
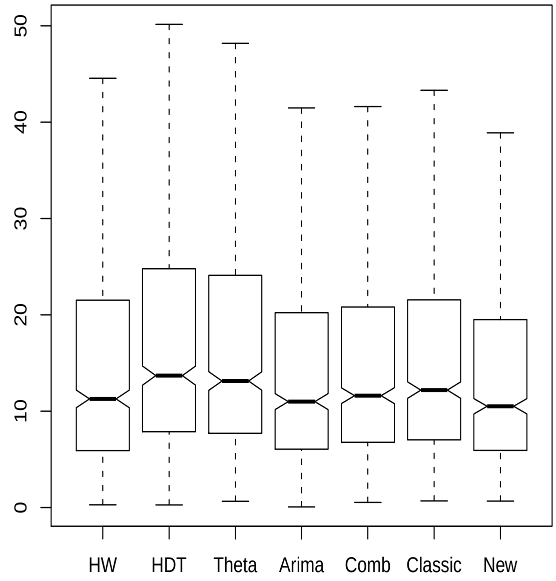
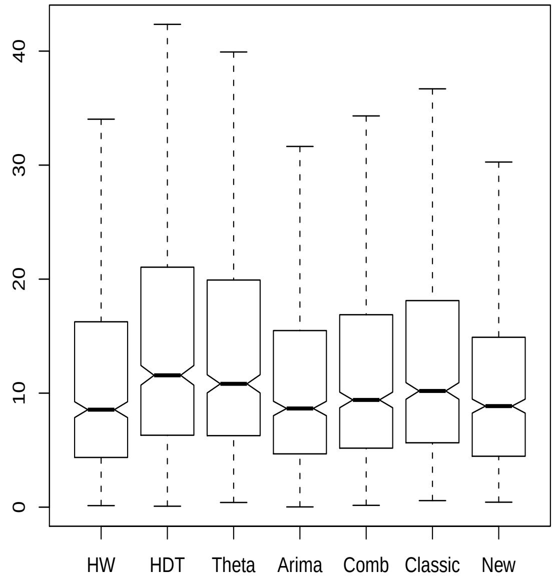


Figure 3: **NN3 Competition:** Boxplots of the MdAPE measures obtained across the 111 time series for the seven forecasting methods and for 18 (upper left plot), 12 (upper right plot) and 1 (bottom plot) steps-ahead forecasts.





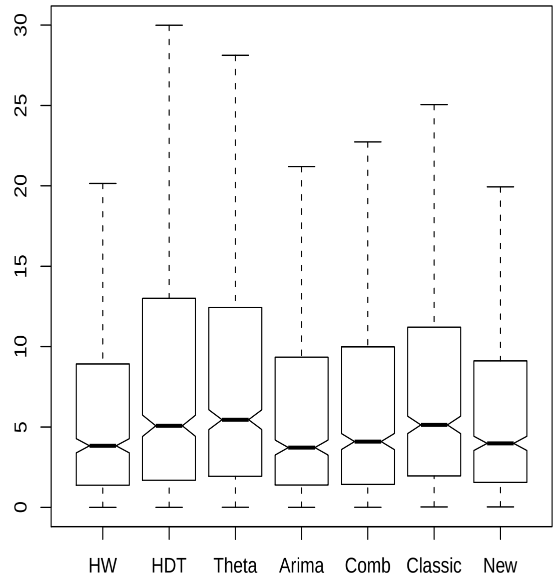


Figure 4: **M1 Complete:** Boxplots of the MdAPE measures obtained across the 729 time series for the seven forecasting methods and for 18 and 8 (upper left plot), 12 and 4 (upper right plot) and 1 (bottom plot) steps-ahead forecasts for monthly and quarterly data respectively.

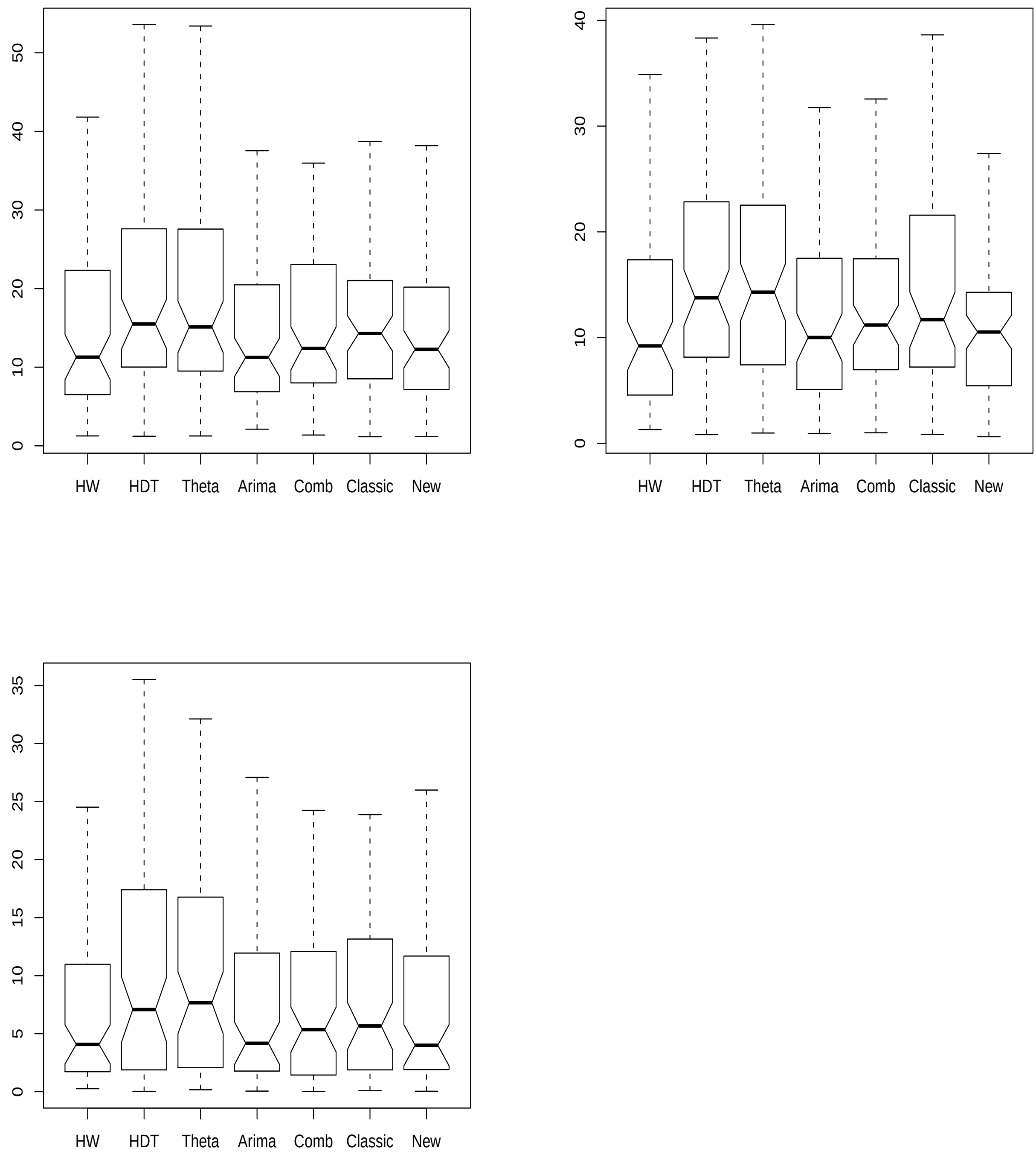


Figure 5: **M1 Reduced:** Boxplots of the MdAPE measures obtained across the 76 time series for the seven forecasting methods and for 18 and 8 (upper left plot), 12 and 4 (upper right plot) and 1 (bottom plot) steps-ahead forecasts for monthly and quarterly data respectively.

6 Conclusions

A new decomposition based forecasting method is developed and applied to time series from the NN3 Competition and the M1 Competition complete and reduced datasets. The proposed method achieves the extrapolation of the target data through the individual extrapolation of the auxiliary sub-series, returned from the application of a decomposition procedure, including that of the irregular component. The performance evaluation results, obtained from the implementation of the method on three different datasets, for three different forecast horizons and using four different error measures, indicate that this can perform comparatively well with standard statistical techniques in the literature for long lead times. It performs persistently well across a wide range of time series with varied characteristics, underlying structure and level of noise in the data. The relative increase in prediction accuracy, the stability of the results across the three datasets examined, and the simplicity of the underlying method are some of the strengths underlying this novel approach to forecasting.

However, an admitted weakness of the method lies in the fact that this can only be applied to time series with frequency greater than one and with observations spanning a minimum of three years, thus rendering its application to short and annual data infeasible. Furthermore, the relatively weak performance of the method documented for the shortest forecast horizon indicates that the comparative advantages of the proposed method, exhibited for long term forecasts, do not translate to short-term forecasts.

The explanation for the weak performance of the proposed method for the one-step ahead forecast horizon lies in the combination technique involved in the forecasting process. It has been shown in the literature that the gains from the application of a forecast combining method increase with increasing uncertainty, and thus with increasing forecast horizon (Makridakis and Winkler, 1983; Lobo, 1992). Consequently, given that the proposed method employs a combination technique to achieve the extrapolation of the global series, it is anticipated that the longer the forecast horizon, the more competitive the method will be. For the shortest forecast horizon of one step ahead, on the other hand, methods which employ a single model for the extrapolation of the global time series are expected to be relatively more robust.

A possible extension to the proposed method could be the application of a multiplicative decomposition method for the disaggregation of the time series components. It is presumed that a method which can distinguish between additive and multiplicative seasonality, will further enhance predictive results, conditional on the fact that a more accurate estimation of the seasonal component could be obtained.

References

J. Scott Armstrong. Combining forecasts: The end of the beginning or the beginning of the end? *International Journal of Forecasting*, 5(4):585–588, 1989. ISSN 0169-2070. doi: http://dx.doi.org/10.1016/0169-2070(89)90013-7.

J. Scott Armstrong. Findings from evidence-based forecasting: Methods for reducing forecast error. *International Journal of Forecasting*, 22(3):583–598, 2006. ISSN 0169-2070. doi: http://dx.doi.org/10.1016/j.ijforecast.2006.04.006.

J. Scott Armstrong and Fred Collopy. Error measures for generalizing about forecasting methods: Empirical comparisons. *International Journal of Forecasting*, 8(1):69–80, 1992. ISSN 0169-2070. doi: http://dx.doi.org/10.1016/0169-2070(92)90008-W.

J.S. Armstrong, E.J. Lusk, E.S. Jr. Gardner, M.D. Geurts, L.L. Lopes, R.E. Markland, R.L. McLaughlin, P. Newbold, D.J. Pack, A. Andersen, R. Carbone, R. Fildes, H.J. Newton, E. Parzen, R.L. Winkler, and S. Makridakis. Commentary on the makridakis time series competition (m-competition). *Journal of Forecasting*, 2:259–311, 1983.

V. Assimakopoulos and K. Nikolopoulos. The theta model: a decomposition approach to forecasting. *International Journal of Forecasting*, 16(4):521–530, 2000. ISSN 0169-2070. doi: http://dx.doi.org/10.1016/S0169-2070(00)00066-2.

George Athanasopoulos, Rob J Hyndman, Haiyan Song, and Doris Wu. The tourism forecasting competition. *International Journal of Forecasting*, 27(2), 2011. to appear.

W. Bell and S.C. Hillmer. Issues involved with the seasonal adjustment of economic time series. *Journal of Business and Economic Statistics*, 2(4):291–349, 1984.

G.E.P. Box and G.M. Jenkins. *Time Series Analysis: Forecasting and Control (revised edition)*. Holden Day, San Francisco, 1976.

G.E.P. Box, S.C. Hillmer, and G.C. Tiao. *Seasonal Analysis of Economic time Series*, chapter Analysis and Modelling of Seasonal Time Series, pages 309–334. A. Zellner (Ed.), Washington D.C.: US Dept. of Commerce-Bureau of the Census, 1978.

C. Chatfield, A.B. Koehler, J.K. Ord, and R.D. Snyder. A new look at models for exponential smoothing. *The Statistician*, 50:147–159, 2001.

Chunhang Chen. Robustness properties of some forecasting methods for seasonal time series: A monte carlo study. *International Journal of Forecasting*, 13(2):269–280, 1997. ISSN 0169-2070. doi: http://dx.doi.org/10.1016/S0169-2070(97)00014-9.

Robert T. Clemen. Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting*, 5(4):559–583, 1989. ISSN 0169-2070. doi: http://dx.doi.org/10.1016/0169-2070(89)90012-5.

R.B. Cleveland, W.S. Cleveland, J.E. McRae, and I. Terpenning. Stl: A seasonal-trend decomposition procedure based on loess. *Journal of Official Statistics*, 6:3–73, 1990.

W. S. Cleveland, S.J. Devlin, and I.J. Terpenning. The sabl statistical and graphical methods. Computing Information Library, Bell Labs, 600 Mountain Ave., Murray Hill, N.J. 07974, USA, 1981a.

W. S. Cleveland, S.J. Devlin, and I.J. Terpenning. The details of the sabl transformation, decomposition and calendar methods. Computing Information Library, Bell Labs, 600 Mountain Ave., Murray Hill, N.J. 07974, USA, 1981b.

S. F Crone and K. Nikolopoulos. Nn3 artifficial neural networks and computational intelligence fore- casting competition. `www.neuralforecastingcompetition.com/NN3/results.htm`, 2007.

E.D. Dagum. *The X-11-ARIMA/88 Seasonal Adjustment Method Foundations and Users Manual.* Time Series Research and Analysis Division, Statistics Canada technical report, Canada, 1988.

P. Damrongkulkamjorn and P. Churueang. Monthly energy forecasting using decomposition method with application of seasonal arima. In *The 7th International Power Engineering Conference.*, 2005.

F.X. Diebold and J. Lopez. Forecast evaluation and combination. NBER Technical Working Papers 0192, National Bureau of Economic Research, Inc., 1996.

J. Durbin and S.J. Koopman. *Time Series Analysis by State Space Methods.* New York: Oxford University Press, 2001.

J. Fan and I. Gijbels. *Local Polynomial Modelling and its Applications.* New York: Chapman and Hall, 1996.

Robert Fildes. The practice of econometrics: Classical and contemporary : Ernst r. berndt, (addison-wesley publishing company, reading, mass., 1991), pp. 702, $18.95. *International Journal of Forecasting*, 8(2):269–270, 1992. ISSN 0169-2070. doi: http://dx.doi.org/10.1016/0169-2070(92)90124-R.

Robert Fildes, Michle Hibon, Spyros Makridakis, and Nigel Meade. Generalising about univariate forecasting methods: further empirical evidence. *International Journal of Forecasting*, 14(3):339–358, 1998. ISSN 0169-2070. doi: http://dx.doi.org/10.1016/S0169-2070(98)00009-0.

B. Fischer. Decompositions of time series. comparing different methods in theory and practice. VIROS (Virtual Institute for Research in Official Statistics, Luxembourg, 1995.

V. Gómez and A. Maravall. *Programs TRAMO and SEATS. Instructions for the User, (with some updates).* Servicio de Estudios, Banco de España, Spain, working paper 9628 edition, 1996.

Paul Goodwin and Richard Lawton. On the asymmetry of the symmetric mape. *International Journal of Forecasting*, 15(4):405–408, 1999. ISSN 0169-2070. doi: http://dx.doi.org/10.1016/S0169-2070(99)00007-2.

Jan G. De Gooijer and Rob J. Hyndman. 25 years of time series forecasting. *International Journal of Forecasting*, 22(3):443–473, 2006. ISSN 0169-2070. doi: http://dx.doi.org/10.1016/j.ijforecast.2006.01.001.

A.C. Harvey. *Forecasting Structural Time Series Models and the Kalman Filter.* Cambridge University Press, 1989.

D.F. Hendry and M.P Clements. Pooling of forecasts. *Econometrics Journal*, 5:1–26, 2002.

Charles C. Holt. Forecasting seasonals and trends by exponentially weighted moving averages. *International Journal of Forecasting*, 20(1):5–10, 2004. ISSN 0169-2070. doi: http://dx.doi.org/10.1016/j.ijforecast.2003.09.015.

R. Hyndman. Another look at forecast-accuracy metrics for intermittent demand. *Foresight: The International Journal of Applied Forecasting*, 4:43–46, 2006.

R.J. Hyndman and Y. Khandakar. Automatic time series forecasting: The forecast package for r. *Journal of Statistical Software*, 26(3), 2008.

R.J. Hyndman, A.B. Koehler, J.K. Ord, and R.D. Snyder. Prediction intervals for exponential smoothing state space models. *Journal of Forecasting*, 24(17-37), 2005.

Rob J Hyndman. *forecast: Forecasting functions for time series*, 2010. URL `http://CRAN.R-project.org/package=forecast`. R package version 2.07.

Rob J. Hyndman and Baki Billah. Unmasking the theta method. *International Journal of Forecasting*, 19(2):287–290, 2003. ISSN 0169-2070. doi: http://dx.doi.org/10.1016/S0169-2070(01)00143-1.

Rob J. Hyndman and Anne B. Koehler. Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22(4):679–688, 2006. ISSN 0169-2070. doi: http://dx.doi.org/10.1016/j.ijforecast.2006.03.001.

Rob J. Hyndman, Anne B. Koehler, Ralph D. Snyder, and Simone Grose. A state space framework for automatic forecasting using exponential smoothing methods. *International Journal of Forecasting*, 18(3):439–454, 2002. ISSN 0169-2070. doi: http://dx.doi.org/10.1016/S0169-2070(01)00110-8.

R.E. Kalman. A new approach to linear filtering and prediction problems. *Transactions of the ASME–Journal of Basic Engineering*, 82(Series D):35–45, 1960.

G. Kitagawa and W. Gersch. *Smoothness Priors Analysis of Time Series. Lecture Notes in Statistics.* New York: Springer-Verlag, 1996.

Anne B. Koehler. Time series analysis and forecasting with applications of sas and spss: Robert yaffee and monnie mcgee (contributor), (2000) san diego: Academic press. 496 pages. isbn 0 127 67870 0 hardback: 29.95, $69.95. *International Journal of Forecasting*, 17(2):301–302, 2001. ISSN 0169-2070. doi: http://dx.doi.org/10.1016/S0169-2070(01)00087-5.

G. J. D. Lobo. Analysis and comparison of financial analysts' time series, and combined forecasts of annual earnings. *Journal of Business Research*, 24:269–280, 1992.

F.C. Macaulay. The movements of interest rates, bond yields, and stock prices in the united states since 1856. Technical report, National Bureau of Economic Research, New York, 1938.

S. Makridakis and R. Winkler. Averages of forecasts: Some empirical results. *Management Science*, 29:987–996, 1983.

S. Makridakis, A. Andersen, R. Carbone, R. Fildes, M. Hibon, R. Lewandowski, J. Newton, E. Parzen, and R. Winkler. The accuracy of extrapolative (time series) methods: Results of a forecasting competition. *Journal of Forecasting*, 1(2):111–153, 1982.

S. Makridakis, S.C. Wheelwright, and R.J. Hyndman. *Forecasting: Methods and Applications.* New York: Wiley, 1998.

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

Spyros Makridakis and Michle Hibon. The m3-competition: results, conclusions and implications. *International Journal of Forecasting*, 16(4):451–476, 2000. ISSN 0169-2070. doi: http://dx.doi.org/10.1016/S0169-2070(00)00057-1.

M. Marcellino. Forecast pooling for short time series of macroeconomic variables. *Oxford Bulletin of Economic and Statistics*, 66:91–112, 2004.

L.M. Menezes, D.W. Bunn, and J.W. Taylor. Review of guidelines for the use of combined forecasts. *European Journal of Operational Research*, 120:190–204, 2000.

T.C. Mills. *Modelling Trends and Cycles in Economic Time Series.* Palgrave Macmillan, Basingstoke, 2003.

D.B. Percival and A.T. Walden. *Wavelet Methods for Time Series Analysis.* Cambridge University Press, 2000.

W.M. Persons. Indices of business conditions. *Review of Economics and Statistics*, 1:5–110, 1919.

Hashem M. Pesaran and Allan Timmermann. Selection of estimation window in the presence of breaks. *Journal of Econometrics*, 137(1):134–161, March 2007.

R Development Core Team. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria, 2010. URL `http://www.R-project.org`. ISBN 3-900051-07-0.

D. Ruppert, M.J. Wand, and R.J. Carroll. *Semi-parametric Regression.* Cambridge University Press, 2003.

Steve Satchell and Allan Timmermann. On the optimality of adaptive expectations: Muth revisited. *International Journal of Forecasting*, 11(3):407–416, 1995. ISSN 0169-2070. doi: http://dx.doi.org/10.1016/0169-2070(95)00588-7.

J.H. Stock and M. Watson. A comparison of linear and nonlinear univariate models for forecasting macroeconomic time series. In *Cointegration, Causality, and Forecasting A Festschrift in Honour of Clive W.J. Granger*, pages 1–44. R.F Engle and H. White (Eds.), Oxford University Press, 1999.

J.H. Stock and M. Watson. Combination forecasts of output growth in a seven-country data set. *Journal of Forecasting*, 23:405–430, 2004.

Leonard J. Tashman. Out-of-sample tests of forecasting accuracy: an analysis and review. *International Journal of Forecasting*, 16(4):437–450, 2000. ISSN 0169-2070. doi: http://dx.doi.org/10.1016/S0169-2070(00)00065-0.

H.K. Temraz, M.M.A. Salama, and V.H. Quintana. Application of the decomposition technique for forecasting the load of a large electric power network. *Generation, Transmission and Distribution, IEE Proceedings-*, 143(1):13 –18, jan. 1996. ISSN 1350-2360. doi: 10.1049/ip-gtd:19960110.

Patrick A. Thompson. An mse statistic for comparing forecast accuracy across series. *International Journal of Forecasting*, 6(2):219–227, 1990. ISSN 0169-2070. doi: http://dx.doi.org/10.1016/0169-2070(90)90007-X.

Patrick A. Thompson. A statistician in search of a population. *International Journal of Forecasting*, 8(1):103–104, 1992. ISSN 0169-2070. doi: http://dx.doi.org/10.1016/0169-2070(92)90013-Y.

A.G. Timmermann. *Handbook of Economic Forecasting*, chapter Forecast Combinations, pages 135–196. Elliott, Granger and Timmermann (Eds.), Elsevier, 2006.

M. West and J. Harrison. *Bayesian Forecasting and Dynamic Models.* London: Springer, 1997.

P. Whittle. *Prediction and Regulation by Linear Least-Square Methods.* University of Minnesota Press, 2nd edition, 1983.

P. Winters. Forecasting sales by exponentially weighted moving averages. *Management Science*, 6:324–342, 1960.

Hui Zou and Yuhong Yang. Combining time series models for forecasting. *International Journal of Forecasting*, 20(1):69–84, 2004. ISSN 0169-2070. doi: http://dx.doi.org/10.1016/S0169-2070(03)00004-9.