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 & 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 (Hyndman & De Gooijer (2006)). Nevertheless, these models are just "intentional abstractions of a much more complicated reality"[1] 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 and decomposition procedures 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 together 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 & Watson (2004), Timmermann (2006)). Nonetheless, the restrictive complexity of some existing combination methodologies and the lack of comprehensive guidelines for their application have been admitted flaws in the literature (Armstrong (1989), Menezes *et al.*(2000)).

Decomposition procedures 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 model.

In this paper, such a forecasting methodology 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

[1] Diebold & Lopez (1996, p22)

methodology 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, whilst a linear combination technique is used to obtain an estimation for the global series. The main underlying idea of the methodology 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 methodology is applied to the NN3 (Crone & Nikolopoulos (2007)) and M1 Competition (Makridakis *et al.* (1982)) datasets. The results obtained are benchmarked against the results of four forecasting methodologies namely ARIMA, Theta, Holt's Damped Trend (hereafter HDT) and Holt-Winter's (hereafter HW). These methods can be readily implemented in a software package[2], and were selected on the basis of their performance in previous forecasting competitions and empirical applications.

The paper unfolds as follows. Section two gives an overview of decomposition and combination techniques. In section three, the various steps leading to the implementation of the new methodology are described in detail. Section four presents the results from the forecasting application of the new methodology on the NN3 competition data. The power of the method in forecasting a large range of time series with different characteristics is tested in section five, with its application on the complete dataset of time time series from the M1 Competition. 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)). 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*

[2]The statistical software used in this paper is the **R-Language** and is free to download from `www.r-project.org`.

al.(1998), Makridakis & 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 & Clements (2002) offer a formal explanation of this phenomenon. They suggest that forecast combining adds value when the individual forecasting models are *differentially mis-specified*. This argument is supported in the work of Makridakis (1989), Diebold & Lopez (1996) and Stock & Watson (1999, 2004). Furthermore, by combining, the practitioner avoids the possibility of choosing the worst forecasting model for the particular point in time and, hence, robustifies the estimations across all forecasting horizons (Armstrong *et al.*(1983), De Gooijer & Hyndman (1998)). Another explanation given by Pesaran & Timmermann (2005) 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, a simple linear combination technique is used on the extrapolated disaggregated subseries to obtain an estimation of the global series.

2.2. An Overview on Decomposition

Decomposition techniques have been primarily developed by Persons (1919) to identify and isolate salient features of a time series. They have since been used for the analysis of economic data to produce official statistics by various governments and institutions (see Fischer (1995) for a well-documented survey on the various methods).

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 who deal with the extrapolation of time series through the extrapolation of the individual components, obtained from the application of averaging techniques (Damrongkulkamjorn & Churueang, (2008), Temraz *et al.* (1996)). This approach to forecasting is known as the classical decomposition technique and was developed by Macauley (1938) and later described in Makridakis Wheelwright & Hyndman (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[3] was used for the implementation of the new methodology. 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, whilst 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 observations are withheld for evaluating the predictive ability of the new methodology. The time series are not subjected to any data preprocessing prior to the implementation of the new forecasting methodology.

The large variability of structural characteristics within the 111 time series underlines the need for a single forecasting methodology that could predict all series with a relatively high level of accuracy, and consequently, remain unaffected by structural changes and persistent trending or

[3]The data used for the analysis can be obtained from:
`http://www.neural-forecasting-competition.com/NN3/datasets.htm`

seasonal behavior in the data. In the proceeding section, such a methodology 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 the new methodology developed is 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. (1984) 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 methodology 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[4]. An important advantage of the STL procedure and the use of the Loess smoother, is the robustness of the returned trend and seasonal components to outliers in the data.

[4]A Loess smoother 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.

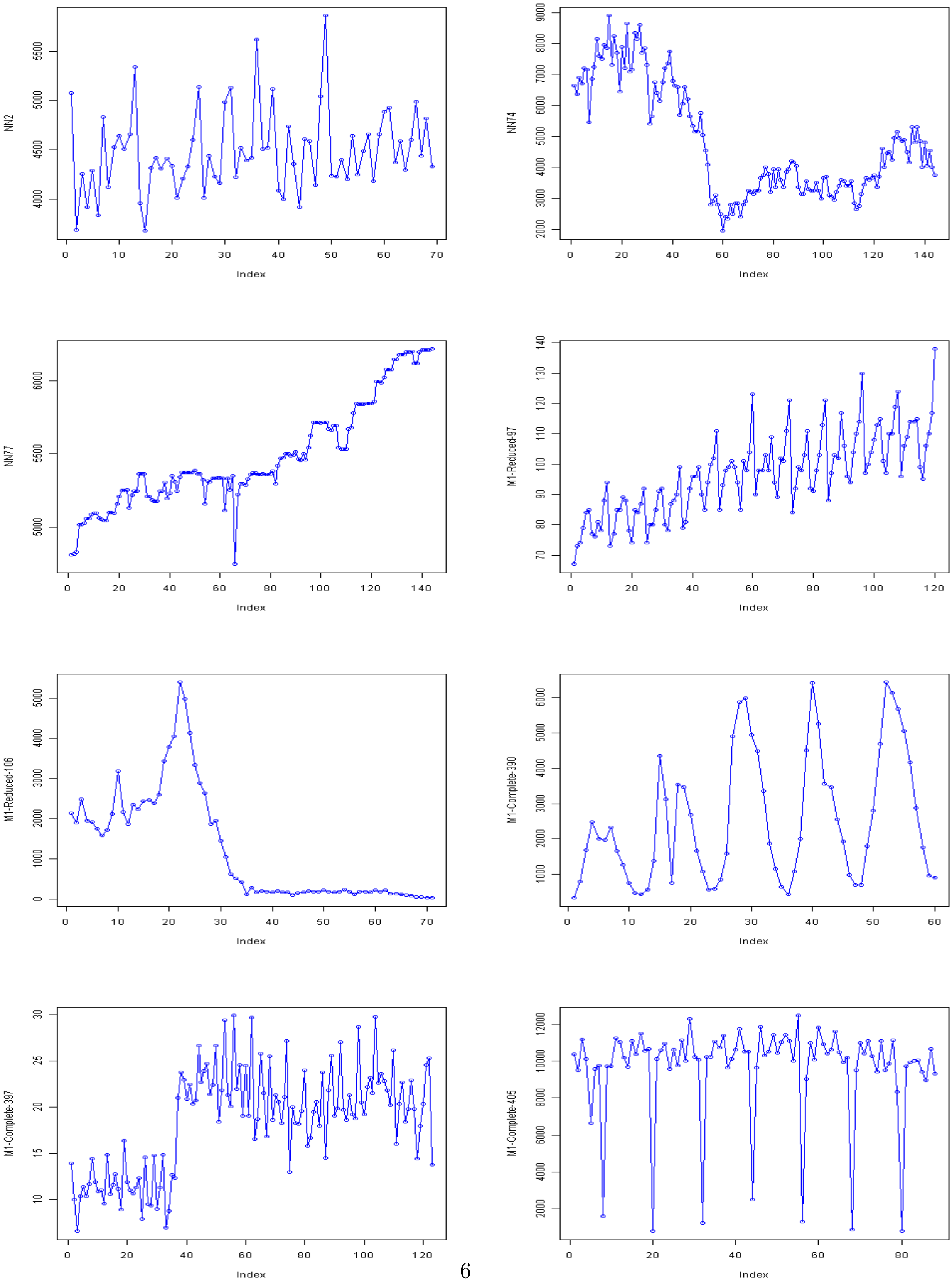

Figure 1: Time series plots for the 2^{nd}, 74^{th} and 77^{th} time series of the NN3 Competition, 97^{th} and 106^{th} time series of the M1 Competition reduced dataset, and 390^{nd}, 397^{th} and 405^{th} time series of the M1 Competition complete dataset.

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.

The iterated cycle is composed of two recursive procedures, the inner and the outer loop. The inner loop performs six basic steps:

1. *Detrending:* Let $s_t^{(k)}$ and $m_t^{(k)}$ be the seasonal and trend components obtained at the end of the k^{th} pass. At iteration $k+1$, the global series x_t is detrended by eliminating the estimated trend component $m_t^{(k)}$, i.e. $x_t - m_t^{(k)}$. At the start of the first iteration, $m_t^{(0)}$ is set to be zero.
2. *Seasonal Smoothing:* A Loess smoother is then applied to the sub-series obtained above $(x_t - m_t^{(k)})$ to form a preliminary seasonal component, $s_t^{\widetilde{(k+1)}}$.
3. *Filtering of Smoothed Seasonality:* A simple moving average is applied to the preliminary seasonal component of the second step, $s_t^{\widetilde{(k+1)}}$, followed by the application of a Loess smoother, to identify any remaining trend, $m_t^{\widetilde{(k+1)}}$.
4. *Detrending of Smoothed Seasonality:* The additive seasonal component is then estimated as the difference between the preliminary seasonal component of the second step, $s_t^{\widetilde{(k+1)}}$, and the preliminary trend component of the third step, $m_t^{\widetilde{(k+1)}}$, i.e. $s_t^{(k+1)} = s_t^{\widetilde{(k+1)}} - m_t^{\widetilde{(k+1)}}$.
5. *Deseasonalizing:* A seasonally adjusted series is computed by subtracting the result of the fourth step from the original data $(x_t - s_t^{(k+1)})$.
6. *Trend Smoothing* The seasonally adjusted series is then smoothed again by Loess to give an estimate of the trend component $m_t^{(k+1)}$.

Hence, 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 $(e_t^{(k+1)} = x_t - m_t^{(k+1)} - s_t^{(k+1)})$. 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[5]. In the current

[5]For a more detailed description of the STL decomposition procedure, the reader is referred to Cleveland et al., 1990.

application, an automated method was implemented within the algorithm for the implementation of the STL procedure, which tests for outliers in the data based on the equation.:

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

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[6] returns, m_t, s_t and e_t, as in equation (1). Figure 2 depicts the results from the application of the STL decomposition procedure on series NN52 from the NN3 Competition dataset.

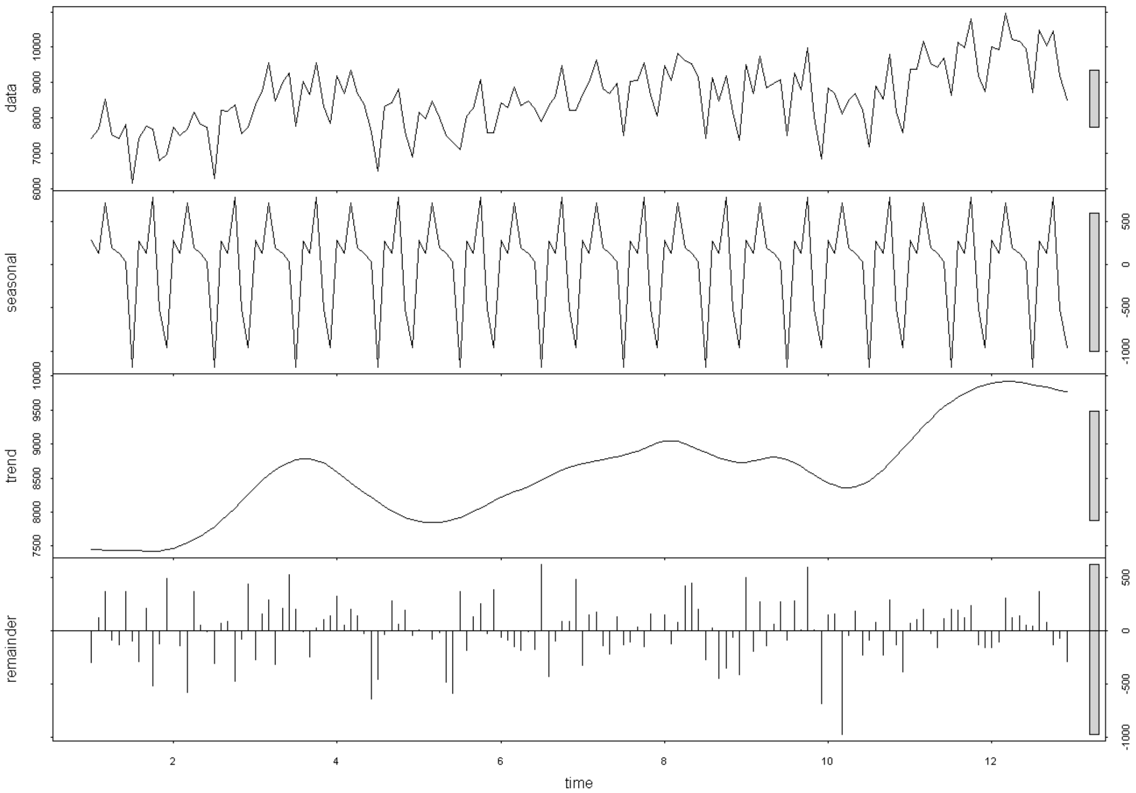

Figure 2: Results from the application of the STL decomposition procedure on time series NN52.

The success of the new methodology therefore relies on the successful interpolation of linear combinations of the additive components.

[6]The STL decomposition procedure can be readily implemented in **R-Language** using the function `stl()`.

4.2. Extrapolating the Disaggregated 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 best suited for the extrapolation of each individual component, four forecasting methodologies namely ARIMA (Box & Jenkins, 1976), Theta (Assimakopoulos & Nikololopoulos, 2000), HDT (Holt, 1957) and HW (Winters, 1960), were applied on early hold-out data and 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 & Timmermann, 1995, Hyndman *et al.*, 2000, Chatfield *et al.*, 2001, and Hyndman *et al.*, 2005). The Theta method which, as shown by Hyndman & Billah (2001) is simple exponential smoothing with drift, was the best performing method in the M3-Competition (Makridakis & Hibon, 2000), and was reported as the second best statistical method for the NN3 Competition after Wildi (Crone & 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 & 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 by the authors, were used to choose the optimal parameters for the implementation of the other three forecasting techniques. The prediction intervals for the Theta method were computed using the underlying state space model (Hyndman & Billah, 2001).

These forecasting methods were applied to the raw data to predict 18 observations ahead, using only the first 36 observations (3 years) in the sample. Therefore, only the first 54 observations from each time series are used in this analysis. The 'best' method for each time series, in terms of mean absolute scaled error (MASE), was recorded and examined relative to the structural components in the time series[7]. Firstly, in order to determine the strength of each components in

[7]The choice of the forecasting horizon and the historical window was based on an *ad hoc* selection. However,

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:

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

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, formed group G_M. For each of the time series in the group, the coefficients of determination of the three components are then recorded. Therefore, each group, G_M was associated with a matrix of nx3 coefficients of determination, n being the number of time series for which method M had the smallest MASE, i.e.,

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

The purpose of this classification was to determine the relationship between the performance of each individual forecasting method on the raw data in respect to the features of the series. From the analysis some important conclusions were drawn:

- For time series with high levels of seasonality, the best forecasting methods were ARIMA and HW method.
- For time series with high levels of trend component, the best forecasting methods were HDT, Theta and ARIMA.
- For time series with high levels of error component, the best forecasting method was ARIMA.

experimental results not reported here, revealed that the relative performance of the various methods did not vary significantly when different historical windows and lead times were used.

4.3. Extrapolating the Error Component

The most important step in the application of the new methodology 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.

Although it is customary in the literature to assume that the error component is white noise, nevertheless in the current methodology, information can still be drawn from its subseries and therefore discarding it completely can affect negatively estimation accuracy. Information 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:

$$se_t = s_t + e_t \quad (5)$$

$$me_t = m_t + e_t \quad (6)$$

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

4.4. The New Forecasting Method

In this paper, the ARIMA method is used for the estimation of me_t, and the HW method for the estimation of se_t. The seasonality component was extrapolated using the ARIMA method and the trend component using the Theta method[8]. The choice of these methods is supported by the preliminary analysis carried out in the previous section. From the analysis, it was found that the

[8]Other permutations assumed by the best performing method for each component, found in section 4.2, were also investigated. A number of them returned comparable results with the one chosen.

aforementioned methods were selected as the most accurate forecasting methods for time series with very high levels of trend and seasonal components respectively. In addition, the ARIMA method can deal reasonably well with a high level of residual variability in the data.

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:

$$\hat{x}_{t+1} = (\hat{m}_{t+1}^{(Th)} + \hat{s}_{t+1}^{(AR)} + \hat{me}_{t+1}^{(AR)} + \hat{se}_{t+1}^{(HW)})/2 \tag{7}$$

The new methodology is therefore based on the linear combination of the extrapolated sub-series. Accordingly, there is an element of originality in the methodology 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 model 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 & Clements, 2002).

5. Application

5.1. Performance Evaluation of the New Forecasting Method

The performance evaluation of the new methodology is benchmarked against the four forecasting methods namely HW, HDT, Theta and ARIMA, and is carried out using the last 18 observations in the sample.

A set of measures were adopted to evaluate the performance of the forecasting methods. These can be categorized in scale-dependent, scaled, symmetric and relative. Table 1 gives the list of error measures examined under the four evaluation categories.

Y_t is the real observation and F_t the predicted observation at time t. Also,

$$\epsilon_t = Y_t - F_t \tag{8}$$

$$q_t = \frac{\epsilon_t}{\frac{1}{n-1}\sum_{i=2}^{n}|Y_i - Y_{i-1}|} \tag{9}$$

$$r_t = \frac{\epsilon_t}{\epsilon_t^*} \tag{10}$$

A. Scale-Dependent Measures		
MAE	Mean Absolute Error	$mean(\|\epsilon_t\|)$
MdAE	Median Absolute Error	$median(\|\epsilon_t\|)$
MSE	Mean Square Error	$mean(\epsilon_t^2)$
RMSE	Root Mean Square Error	$\sqrt{MSE}$
B. Scaled Errors		
MASE	Mean Absolute Scaled Error	$mean(\|q_t\|)$
MdASE	Median Absolute Scaled Error	$median(\|q_t\|)$
RMSSE	Root Mean Squared Scaled Error	$\sqrt{mean(q_t^2)}$
C. Symmetric Errors		
sMAPE	Sym. Mean Abs. Perc. Error	$mean\left(200\frac{\|Y_t-F_t\|}{(Y_t+F_t)}\right)$
sMdAPE	Sym. Median Abs. Perc. Error	$median\left(200\frac{\|Y_t-F_t\|}{(Y_t+F_t)}\right)$
D. Relative Error Measures		
MRAE	Mean Relative Absolute Error	$mean(\|r_t\|)$
MdRAE	Median Relative Absolute Error	$median(\|r_t\|)$
GMRAE	Geometric Mean Rel. Abs. Error	$gmean(\|r_t\|)$

Table 1: List of error measures employed for the performance evaluation of the new forecasting method

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 used is the random walk model where F_t is equal to the last observation, Y_{t-1}.

Scale-dependent measures are based on the variability of the predictions when compared to the real observations and are useful when comparing methods for the same data set, which is also the purpose of this analysis. Relative errors 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 & Collopy, 1992, Fildes, 1992, Thompson, 1990, 1992). However, in the case of equal consecutive observations, this error measure category returns infinite values. This was also observed in the application of Hyndman & Koehler (2006), where they admitted this to be a “serious deficiency” of the relative error measures. In the current application, the relative error measures were 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 & Koehler (2006). Specifically, they recommended MASE (Mean Absolute Scaled Error) "to become the standard measure for forecast accuracy" due to the fact that it is always defined and finite, unlike other measures in certain occasions. Finally, symmetric errors were the main error measures used in the NN3 competition to evaluate performance across each forecasting horizon and across all time series.

Table 2 reports the percentage of times that one method was found to be more accurate than another method across the four error measures examined, namely MAE, MASE, sMAPE[9] and MdRAE. The results for the other error measures were very similar and are not reported here to save space[10]. Therefore, every entry, $a_{i,j}$, in the table shows the percentage of times across the 111 time series, that method i had a smaller error than method j. It is evident from the results in table 2 that the new forecasting method outperforms the benchmark methods for all the error measures considered. It returned a smaller error for a larger percentage of time series than any of the other four forecasting methods considered. ARIMA was the second best method, outperforming the other three statistical techniques in more than 60% of the time series. However, this was outperformed by the new methodology in more than 50% of the series. HDT appears to be the weakest method investigated.

Under the MAE, MASE and sMAPE the new methodology returned the smallest error in 41 out of 111 time series. The result was 31 out of 111 for the MdRAE. Furthermore, the level of improvement in the predictions from the implementation of the new methodology was above 10% in all three evaluations. There was a **10.72%** average improvement in both the MAE and MASE evaluation, for the sMAPE, this was **11.01%** , and **18.14%** for the MdRAE.

[9] The sMAPE error measure can return negative values in the denominator. In order to avoid this effect, in the current application, the absolute value of the denominator was considered instead, i.e. $mean\left(200\frac{|Y_t-F_t|}{(|Y_t|+|F_t|)}\right)$.

[10] The results from the implementation of other error measures are available from the author upon request.

	HW	HDT	Theta	ARIMA	New
MAE					
HW	-	49.55	48.65	36.94	24.32
HDT	50.45	-	50.45	36.04	35.14
Theta	51.35	49.55	-	37.84	35.14
ARIMA	63.06	63.96	62.16	-	46.85
New	75.68	64.86	64.86	53.15	-
MASE					
HW	-	49.55	48.65	36.94	24.32
HDT	50.45	-	50.45	36.04	35.14
Theta	51.35	49.55	-	37.84	35.14
ARIMA	63.06	63.96	62.16	-	46.85
New	75.68	64.86	64.86	53.15	-
sMAPE					
HW	-	45.05	45.95	35.14	23.42
HDT	54.95	-	50.45	36.04	34.23
Theta	54.05	49.55	-	38.74	35.14
ARIMA	64.86	63.96	61.26	-	46.85
New	76.58	65.77	64.86	53.15	-
MdRAE					
HW	-	49.55	45.05	36.94	35.14
HDT	50.45	-	44.14	33.33	38.74
Theta	54.95	55.86	-	42.34	42.34
ARIMA	63.06	66.67	57.66	-	50.45
New	64.86	61.26	57.66	49.55	-

Table 2: Percentage of times method A (row) was more accurate than method B (column), across the 111 time series, for 18 step-ahead forecasts.

	MAE	MASE	sMAPE	MdRAE
HW	922.26	1.30	18.76	1.46
HDT	1004.42	1.35	18.73	1.52
Theta	1009.74	1.33	18.62	1.48
ARIMA	807.52	1.18	16.19	1.30
New	**797.68**	**1.16**	**15.65**	**1.28**

Table 3: The average error obtained calculated across the 111 time series.

Table 3 presents the average error obtained across the 111 time series for the MAE, MASE, sMAPE and MdRAE error measures, for each of the four statistical methods and the new forecasting methodology and table 4 shows the average ranking for each method across the four error measures. The smallest error and ranking across the five methods examined are shown in bold. It is clear from both tables that the new methodology results in more accurate and robust predictions, returning the smallest error in all four error measure categories examined, and having an average ranking of 2.41 for MAE and MASE, 2.40 for sMAPE and 2.67 for MdRAE.

	MAE	MASE	sMAPE	MdRAE
HW	3.41	3.41	3.50	3.33
HDT	3.28	3.28	3.24	3.33
Theta	3.26	3.26	3.23	3.05
ARIMA	2.64	2.64	2.63	**2.62**
New	**2.41**	**2.41**	**2.40**	2.67

Table 4: The average rank of each method obtained across the 111 time series.

Figure 3 presents a graphical depiction of the performance of the new methodology, compared to the four statistical methods, evaluated using the MASE measure. It is evident from figure 3 that the main advantage of the new forecasting methodology lies in its ability to perform robustly well, irrespective of the characteristics of the time series. Unlike the other statistical forecasting methods examined, which perform relatively well for a particular set of time series (e.g HW for highly seasonal time series) whilst they returns poor predictions for others, the new methodology performs consistently well across the whole set of time series examined. This is also shown in figure 4, where the prediction lines returned by the new forecasting methodology are depicted for some of the time series with a wide range of structural characteristics.

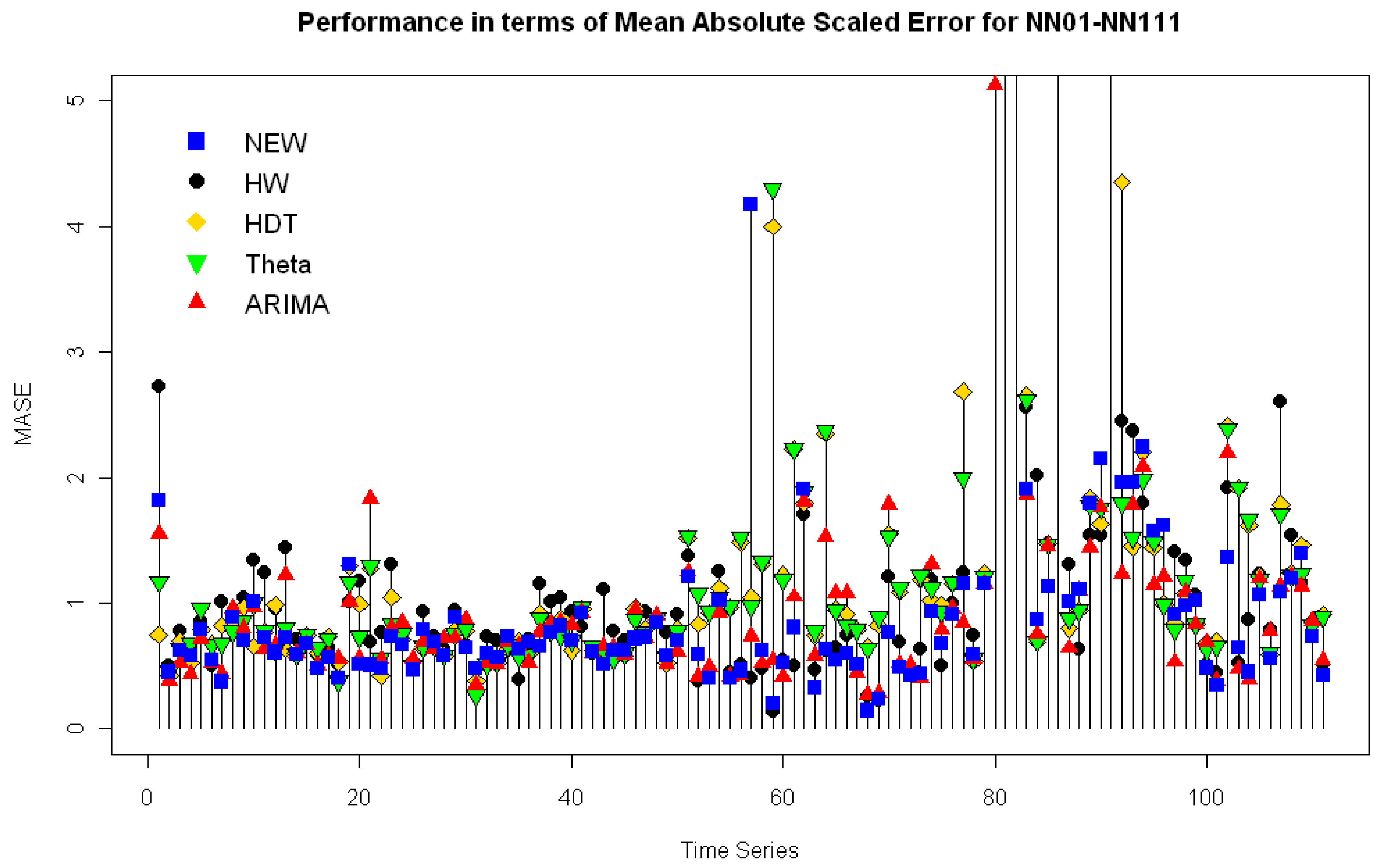

Figure 3: Performance of the five forecasting methods in terms of Mean Absolute Scaled Error (MASE), for 18 steps-ahead forecast.

5.2. Performance Evaluation on the M1 Competition Data

In order to investigate the robustness of the new methodology on a new dataset, this was implemented on the complete and reduced datasets from the M1 Competition. Like the NN3 Competition dataset, the M1 Competition datasets are characterized by a large range of time series with different structural characteristics. However, unlike the NN3 Competition dataset, the trend component is the most salient features in the majority of the time series of the M1 Competition.

The analysis of the new forecasting method on the M1 Competition datasets is limited to the quarterly and monthly time series. Annual data was excluded from the analysis as the STL decomposition method requires a time series frequency greater than two. Series with less than 36 observations were 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 datasets consisted of 76 and 729 time series for the complete and reduced sample, respectively.

The results from the evaluation of the new forecasting method on the M1 Competition datasets, using the four error measures (MAE, MASE, sMAPE and MdRAE) are reported in tables 5 to 10. The results indicate that the new methodology still performs relatively well, when compared to the other four methods, for both the reduced and complete datasets, and appears to be consistently superior for a larger range of time series than the other four statistical methodologies.

The best statistical forecasting methods for the M1 reduced dataset in terms of average error, for MAE, sMAPE and MdRAE error measures, is ARIMA followed by the new methodology (table 7), which outperforms the other four statistical techniques under the MASE error measure. However, in terms of average ranking across the four error measures, the new methodology outperforms the other techniques in all but the MdRAE error measures, for which HW appears to be the best (table 8). The results indicate that the new methodology is more robust than the other four statistical techniques thus always returning good prediction accuracies relative to the other methods across a wide range of time series.

For the M1 complete dataset, the new methodology returned the smallest average error across the 726 time series and across all four error measures examined (table 9). The same results were obtained for the average ranking across the four error measures, indicating that the new methodology outperforms the four statistical techniques in terms of accuracy. Hence, the new

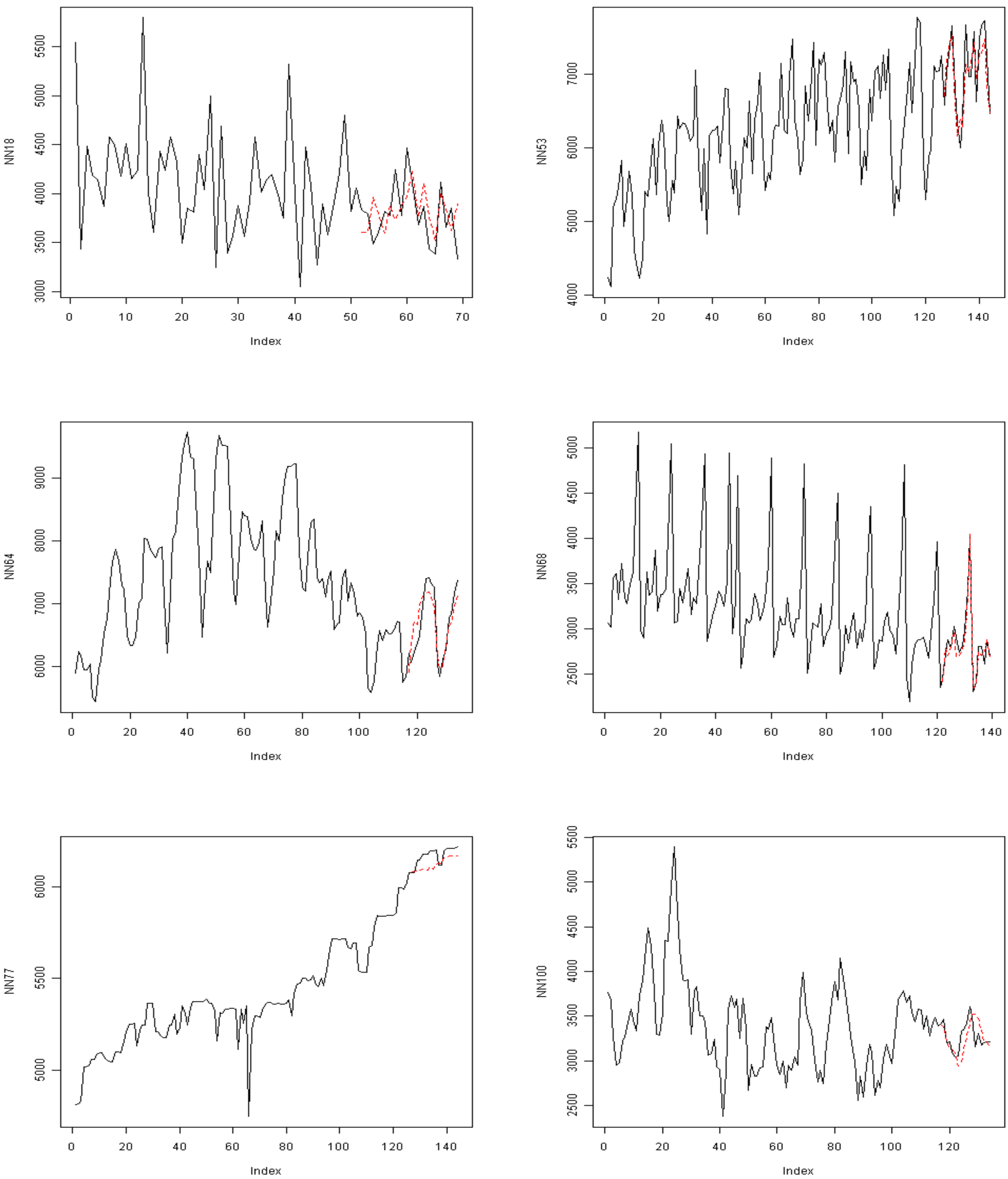

Figure 4: Time series plots of the raw data (solid line) and 18 steps ahead predictions (dashed line) for the 18^{th}, 53^{rd}, 64^{th}, 68^{th}, 77^{th} and 100^{th} time series of the NN3 Competition.

forecasting methodology consistently outperforms the HDT and Theta methods, and performs comparatively well to the ARIMA and HW methods for the two datasets. These results highlight further the superiority and robustness of the new forecasting method in predicting a large range of time series with very different characteristics.

	HW	HDT	Theta	ARIMA	New
MAE					
HW	-	64.47	65.79	53.95	48.68
HDT	35.53	-	48.68	36.84	30.26
Theta	34.21	51.32	-	36.84	36.84
ARIMA	46.05	63.16	63.16	-	46.05
New	51.32	69.74	63.16	53.95	-
MASE					
HW	-	64.47	65.79	53.95	48.68
HDT	35.53	-	48.68	36.84	30.26
Theta	34.21	51.32	-	36.84	36.84
ARIMA	46.05	63.16	63.16	-	46.05
New	51.32	69.74	63.16	53.95	-
sMAPE					
HW	-	64.47	68.42	52.63	50.00
HDT	35.53	-	48.68	36.84	28.95
Theta	31.58	51.32	-	36.84	35.53
ARIMA	47.37	63.16	63.16	-	47.37
New	50.00	71.05	64.47	52.63	-
MdRAE					
HW	-	63.16	60.53	50.00	57.89
HDT	36.84	-	46.05	35.53	34.21
Theta	39.47	53.95	-	44.74	40.79
ARIMA	50.00	64.47	55.26	-	48.68
New	42.11	65.79	59.21	51.32	-

Table 5: **M1 Reduced:** Percentage of times method A (row) was more accurate than method B (column), across the 76 time series, for 18 step-ahead forecasts.

	HW	HDT	Theta	ARIMA	New
MAE					
HW	-	59.53	56.93	52.67	49.52
HDT	40.47	-	45.82	39.92	34.43
Theta	43.07	54.18	-	41.43	33.61
ARIMA	47.33	60.08	58.57	-	43.76
New	50.48	65.57	66.39	56.24	-
MASE					
HW	-	59.53	56.93	52.67	49.52
HDT	40.47	-	45.82	39.92	34.43
Theta	43.07	54.18	-	41.43	33.61
ARIMA	47.33	60.08	58.57	-	43.76
New	50.48	65.57	66.39	56.24	-
sMAPE					
HW	-	59.67	57.61	52.40	50.75
HDT	40.33	-	45.54	40.19	33.61
Theta	42.39	54.46	-	41.29	33.20
ARIMA	47.60	59.81	58.71	-	44.31
New	49.25	66.39	66.80	55.69	-
MdRAE					
HW	-	57.06	54.46	51.17	48.29
HDT	42.94	-	44.58	43.76	33.61
Theta	45.54	55.42	-	45.68	37.45
ARIMA	48.83	56.24	54.32	-	44.03
New	51.71	66.39	62.55	55.97	-

Table 6: **M1 Complete:**Percentage of times method A (row) was more accurate than method B (column), across the 729 time series, for 18 step-ahead forecasts.

	MAE	MASE	sMAPE	MdRAE
HW	1475.94	2.87	16.53	2.57
HDT	1753.07	3.01	19.64	2.49
Theta	1569.82	2.91	18.85	2.39
ARIMA	**1203.24**	2.83	**15.17**	**2.21**
New	1389.20	**2.61**	16.51	2.24

Table 7: **M1 Reduced:** The average error obtained calculated across the 76 time series.

	MAE	MASE	sMAPE	MdRAE
HW	2.67	2.67	2.64	**2.68**
HDT	3.49	3.49	3.50	3.47
Theta	3.41	3.41	3.45	3.21
ARIMA	2.82	2.82	2.79	2.82
New	**2.62**	**2.62**	**2.62**	2.82

Table 8: **M1 Reduced:** The average rank of each method obtained across the 76 time series.

	MAE	MASE	sMAPE	MdRAE
HW	1989.14	2.94	17.67	2.45
HDT	2274.07	3.01	18.95	2.38
Theta	2156.25	2.91	18.21	2.22
ARIMA	1786.80	2.86	16.44	2.22
New	**1628.34**	**2.64**	**15.93**	**2.00**

Table 9: **M1 Complete:** The average error obtained calculated across the 729 time series.

	MAE	MASE	sMAPE	MdRAE
HW	2.81	2.81	2.80	2.89
HDT	3.39	3.39	3.40	3.35
Theta	3.28	3.28	3.29	3.16
ARIMA	2.90	2.90	2.90	2.97
New	**2.61**	**2.61**	**2.62**	**2.63**

Table 10: **M1 Complete:** The average rank of each method obtained across the 729 time series.

6. Conclusions

A new decomposition methodology was developed and applied to 111 time series from the NN3 competition. It constitutes an original attempt in the literature to extrapolate the target data through the individual extrapolation of the auxiliary sub-series returned from the application of a decomposition procedure, including the irregular component. The performance evaluation results, obtained from the implementation of four different error measures, showed the new method outperforming all competing statistical techniques in the literature for the NN3 Competition dataset, and performs comparatively well with the best forecasting methods for M1 Competition datasets. Furthermore, it performs persistently well across all time series, irrespective of their characteristics, underlying structure and level of noise in the data. This is an important development in the area of forecasting, since no method has ever being documented to perform consistently well for the majority of time series in previous forecasting competitions and large empirical studies.

The employment of different methodologies for the extrapolation of each of the disaggregated sub-series, together with the differentiability that characterizes the structure of each underlying series were the main factors for the success of the methodology. The increase in prediction accuracy obtained from the application of the new forecasting technique, the stability of the results across the three datasets examined, and the simplicity of the underlying methodology are some of the strengths underlying this novel approach to forecasting.

[1] Armnstrong, J.S (1989). Combining Forecasts: The End of the Beginning or the Beginning of the End? *International Journal of Forecasting, 5*, 585-588.

[2] Armstrong, J.S., Lusk, E.J., Gardner, E.S. Jr., Geurts, M.D., Lopes, L.L., Markland, R.E., McLaughlin, R.L., Newbold, P., Pack, D.J., Andersen, A., Carbone, R., Fildes, R., Newton, H.J., Parzen, E., Winkler, R.L. and Makridakis, S. (1983). Commentary on the Makridakis Time Series Competition (M-Competition). *Journal of Forecasting, 2*, 259-311.

[3] Armstrong, J. S. & Collopy, Fred, (1992). Error Measures for Generalizing About Forecasting Methods: Empirical Comparisons. *International Journal of Forecasting, 8*, 69-80.

[4] Assimakopoulos, V. & Nikolopoulos, K. (2000). The Theta Model: A Decomposition Approach to Forecasting. *International Journal of Forecasting, 16*, 521530.

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

[6] Chatfield, C., Koehler, A.B, Ord, J.K. & Snyder, R.D. (2001). A New Look at Models for Exponential Smoothing. *The Statistician, 50*, 147159.

[7] Chen, C. (1997). Robustness properties of some forecasting methods for seasonal time series: A Monte Carlo Study. *International Journal of Forecasting, 13*, 269-280.

[8] Clemen, R.T. (1989). Combining Forecasts: A Review and Annotated Bibliography. *International Journal of Forecasting, 5*, 559-583.

[9] Cleveland, R.B., Cleveland, W.S, McRae, J.E & Terpenning, I. (1990). STL: A Seasonal-Trend Decomposition Procedure Based on Loess. *Journal of Office Statistics, 6*, 3-73.

[10] Crone, S. F & Nikolopoulos, K. (2007). *NN3 Artificial Neural Networks and Computational Intelligence Forecasting Competition.* (`www.neuralforecastingcompetition.com/NN3/results.htm`).

[11] Damrongkulkamjorn, P. & Churueang, P. (2005). Monthly Energy Forecasting Using Decomposition Method with Application of Seasonal ARIMA. The 7th international Power Energy Conference.

[12] De Gooijer, J.G. & Hyndman, R.J. (1998). 25 Years of IIF Time Series Forecasting: A Selective Review. TI 2005-068/4, Tinbergen Institute Discussion Paper.

[13] Diebold, F.X. & Lopez, J. (1996). Forecast Evaluation and Combination. *NBER Technical Working Papers, 0192, National Bureau of Economic Research, Inc.*

[14] Fildes, R. (1992). The Evaluation of Extrapolative Forecasting Methods. International Journal of Forecasting, 8, 81-98.

[15] Fildes, R., Hibon, M., Makridakis, S., & Meade, N. (1998). Generalizing About Univariate Forecasting Methods: Further Empirical Evidence. *International Journal of Forecasting 14*, 339-358.

[16] Fischer, B. (1995). *Decompositions of Time Series. Comparing Different Methods in Theory and Practice.* VIROS (Virtual Institute for Research in Official Statistics), Luxembourg.

[17] Hendry, D.F. & Clements, M.P (2002). Pooling of Forecasts. *Econometrics Journal, 5*, 1-26.

[18] Holt, C. C. (1957). Forecasting Trends and Seasonals by Exponentially Weighted Averages, *O.N.R. Memorandum 52/1957, Carnegie Institute of Technology. Reprinted in International Journal of Forecasting (2004), 20,*

5-10.

[19] Hyndman, R.J., Koehler, A.B., Snyder, R.D & Grose, S. (2000). A State Space Framework for Automatic Forecasting Using Exponential Smoothing Methods. Monash University, Department of Econometrics and Business Statistics, Working Papers, 9.

[20] Hyndman, R.J & Billah, B. (2001). Unmasking the Theta Method. *International Journal of Forecasting, 19*, 287-290.

[21] Hyndman, R.J & De Gooijer, J.G. (2006). 25 Years of Time Series Forecasting. *International Journal of Forecasting, 22*, 443-473.

[22] Hyndman, R.J. & Khandakar, Y. (2008). Automatic Time Series Forecasting: The Forecast Package for R", *Journal of Statistical Software, 26(3).*

[23] Hyndman, R.J., Koehler, A.B., Ord, J.K. & Snyder, R.D (2005). Prediction Intervals for Exponential Smoothing State Space Models. Journal of Forecasting, 24, 1737.

[24] Hyndman, Rob J. & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting, 22*, 679-688.

[25] Macaulay, F. C. (1938), The Movements of Interest Rates, Bond Yields, and Stock Prices in the United States since 1856, *National Bureau of Economic Research, New York.*

[26] Makridakis, S. (1989). Why Combining Works. International Journal of Forecasting, 5, 601-603.

[27] Makridakis, S., Andersen, A., Carbone, R., Fildes, R., Hibon, M., Lewandowski, R., et al. (1982). The Accuracy of extrapolative (time series) methods: Results of a forecasting competition. *Journal of Forecasting, 1(2), 111-153.*

[28] Makridakis, S., Andersen, A., Carbone, R., Fildes, R., Hibon, M., Lewandowski, R., Newton, J., Parzen, E., & Winkler, R. (1984). The Forecasting Accuracy of Major Time Series Methods, Wiley, Chichester.

[29] Makridakis, S. & Hibon, M. (2000). The M3-Competition: Results, Conclusions and Implications. *International Journal of Forecasting 16*, 451-476.

[30] Makridakis, S., Wheelwright, S. C., & Hyndman R. J. (1998). Forecasting: Methods and Applications, 3rd ed., Wiley: NY.

[31] Marcellino, M. (2004). Forecast Pooling for Short Time Series of Macroeconomic Variables. *Oxford Bulletin of Economic and Statistics, 66*, 91-112.

[32] 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.

[33] Persons, W.M. (1919). Indices of Business Conditions. *Review of Economics and Statistics, 1*, 5-107.

[34] Pesaran, M.H. & Timmermann, A. (2005). Selection of Estimation Window in the Presence of Breaks. Cambridge University and UCSD.

[35] Satchell, S. & Timmermann, A. (1995). On the Optimality of Adaptive Expectations: Muth revisited. *International Journal of Forecasting, 11*, 407416.

[36] Stock, J.H. & Watson, M. (1999). A Comparison of Linear and Nonlinear Univariate Models for Forecasting Macroeconomic Time Series. In: R.F Engle & H. White (Eds.), Festschrift in Honor of Clive Granger, 1-44.

[37] Stock, J.H. & Watson, M. (2004). Combination Forecasts of Output Growth in a Seven-Country Data Set.

Journal of Forecasting, 23, 405-430.

[38] Temraz, H. K, Salama, M.M.A & Quintana, V.H. (1996). Application of the Decomposition Technique for forecasting the load of a Large Electric Power Network, *IEE Proceedings, Generation, Transmission and Distribution, Vol. 143,* **1**, 13-18.

[39] Timmermann, A.G. (2006). Forecast Combinations. In: Elliott, Granger and Timmermann (Eds.), Handbook of Economic Forecasting: Elsevier, 135-196.

[40] Thompson, P. A. (1990). An MSE Statistic for Comparing Forecast Accuracy Across Series. *International Journal of Forecasting, 6*, 219-227.

[41] Thompson, P. A. (1992). A Commentary on Error Measures: A Statistician in Search of a Population. *International Journal of Forecasting, 8*, 103-104.

[42] Winters, P. (1960). Forecasting Sales by Exponentially Weighted Moving Averages. *Management Science* **6**, 324-342.