**University of Bath**

**Alternative formats**
If you require this document in an alternative format, please contact:
openaccess@bath.ac.uk

Contents lists available at ScienceDirect

# International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

# Models for optimising the theta method and their relationship to state space models

Jose A. Fiorucci [a,b], Tiago R. Pellegrini [c], Francisco Louzada [d], Fotios Petropoulos [e,*], Anne B. Koehler [f]

[a] *Department of Statistics, Federal University of São Carlos, Brazil*
[b] *Department of Statistics, University of Connecticut, USA*
[c] *Department of Mathematics and Statistics, University of New Brunswick, Canada*
[d] *Department of Applied Mathematics and Statistics, University of São Paulo, Brazil*
[e] *Cardiff Business School, Cardiff University, UK*
[f] *Professor Emerita, Miami University, OH, USA*

## ARTICLE INFO

## ABSTRACT

Accurate and robust forecasting methods for univariate time series are very important when the objective is to produce estimates for large numbers of time series. In this context, the Theta method's performance in the M3-Competition caught researchers' attention. The Theta method, as implemented in the monthly subset of the M3-Competition, decomposes the seasonally adjusted data into two "theta lines". The first theta line removes the curvature of the data in order to estimate the long-term trend component. The second theta line doubles the local curvatures of the series so as to approximate the short-term behaviour. We provide generalisations of the Theta method. The proposed Dynamic Optimised Theta Model is a state space model that selects the best short-term theta line optimally and revises the long-term theta line dynamically. The superior performance of this model is demonstrated through an empirical application. We relate special cases of this model to state space models for simple exponential smoothing with a drift.

## 1. Introduction

The development of accurate, robust and reliable forecasting methods for univariate time series is very important when large numbers of time series are involved in the modelling and forecasting process. In industrial settings, it is very common to work with large lines of products; thus, efficient sales and operational planning (S&OP) depend heavily on accurate forecasting methods.

Despite the advantages of automatic model selection algorithms (Hyndman & Khandakar, 2008; Hyndman, Koehler, Snyder, & Grose, 2002; Poler & Mula, 2011), there is still a need for accurate extrapolation methods. Forecasting competitions have played an important role in moving toward the forecasting of large numbers of time series, with the objective of identifying high-performing methods. The Theta method attracted the attention of researchers by its simplicity and surprisingly good performance (Koning, Franses, Hibon, & Stekler, 2005; Makridakis & Hibon, 2000), and has been one of the benchmarks in more recent forecasting competitions (Athanasopoulos, Hyndman, Song, & Wu, 2011).

* Correspondence to: Logistics and Operations Management Section, Cardiff Business School, Cardiff University, Cardiff, CF10 3EU, UK.
  *E-mail address:* petropoulosf@cardiff.ac.uk (F. Petropoulos).

The Theta method (Assimakopoulos & Nikolopoulos, 2000, hereafter A&N) is applied to non-seasonal or deseasonalised time series, where the deseasonalisation is usually performed via the multiplicative classical decomposition. The method decomposes the original time series into two new lines through the so-called theta coefficients, denoted by $\theta_1$ and $\theta_2$ for $\theta_1, \theta_2 \in \mathbb{R}$, which are applied to the second difference of the data. The second differences are reduced when $\theta < 1$, resulting in a better approximation of the long-term behaviour of the series (Assimakopoulos, 1995). If $\theta$ is equal to zero, the new line is a straight line. When $\theta > 1$, the local curvatures are increased, magnifying the short-term movements of the time series (A&N). The new lines produced are called theta lines, denoted here by $Z(\theta_1)$ and $Z(\theta_2)$. These lines have the same mean value and slope as the original data, but the local curvatures are either filtered out or enhanced, depending on the value of the $\theta$ coefficient.

In other words, the decomposition process has the advantage of exploiting information in the data that usually cannot be captured and modelled completely through the extrapolation of the original time series. The theta lines can be regarded as new time series and are extrapolated separately using an appropriate forecasting method. Once the extrapolation of each theta line has been completed, recomposition takes place through a combination scheme in order to calculate the point forecasts of the original time series. Combining has long been considered as a useful practice in the forecasting literature (for example, Clemen, 1989; Makridakis & Winkler, 1983; Petropoulos, Makridakis, Assimakopoulos, & Nikolopoulos, 2014), and therefore its application to the Theta method is expected to result in more accurate and robust forecasts.

The Theta method is quite versatile in terms of choosing the number of theta lines, the theta coefficients and the extrapolation methods, and combining these to obtain robust forecasts. However, A&N proposed a simplified version involving the use of only two theta lines with prefixed $\theta$ coefficients that are extrapolated over time using a linear regression (LR) model for the theta line with $\theta_1 = 0$ and simple exponential smoothing (SES) for the theta line with $\theta_2 = 2$. The final forecasts are produced by combining the forecasts of the two theta lines with equal weights. In the M3-Competition, this simplified version of the Theta method was applied only to the monthly time series (Nikolopoulos, Assimakopoulos, Bougioukos, Litsa, & Petropoulos, 2011).

The performance of the Theta method has also been confirmed by other empirical studies (for example Nikolopoulos, Thomakos, Petropoulos, Litsa, & Assimakopoulos, 2012; Petropoulos & Nikolopoulos, 2013). Moreover, Hyndman and Billah (2003), hereafter H&B, showed that the simple exponential smoothing with drift model (SES-d) is a statistical model for the simplified version of the Theta method. More recently, Thomakos and Nikolopoulos (2014) provided additional theoretical insights, while Thomakos and Nikolopoulos (2015) derived new theoretical formulations for the application of the method to multivariate time series, and investigated the conditions under which the bivariate Theta method is expected to forecast better than the univariate one. Despite these advances,

we believe that the Theta method deserves more attention from the forecasting community, given its simplicity and superior forecasting performance.

One key aspect of the Theta method is that, by definition, it is dynamic. One can choose different theta lines and combine the produced forecasts using either equal or unequal weights. However, A&N limit this important property by fixing the theta coefficients to have predefined values. Thus, the Theta method, as implemented in the M3-Competition, is limited in the sense that it focuses only on specific information in the data. On the other hand, if the selection of the appropriate theta lines had been carried out through optimisation, the method could focus on the information that is actually important.

The contributions of this work are fourfold. First, we extend the A&N method by the optimal selection of the theta line that describes the short-term movements of the series best, maintaining the long-term component. The forecasts derived from the two theta lines are combined using appropriate weights, which ensures the recomposition of the original time series. Second, we provide theoretical and practical links between the newly proposed model, the original Theta method and the SES-d model. Third, we also perform a further extension of the model that allows the regression line (the long term component) to be revised at every time period. An empirical evaluation using the M3-Competition database is undertaken in order to obtain insights into the performances of the proposed models. The results reveal improvements in the forecasting accuracy when using the model with both extensions. This model outperforms several benchmarks as well as the A&N simplified version of the Theta method. Fourth, we reproduce the results for the Theta method, as applied to the monthly data in the M3-Competition, very closely.

The paper is organised as follows. Section 2 reviews the original Theta method of A&N and its relationship with the SES-d model. Section 3 presents different models for optimising the Theta method. Section 4 presents the forecasting performances of the proposed models, compared to a list of widely used benchmarks. The evaluation includes more than 3000 time series. Lastly, Section 5 presents our final comments and directions for future research.

## 2. Theta method and SES-d

### 2.1. The original Theta method

Originally, A&N proposed the theta line as the solution of the equation

$$\nabla^2 Z_t(\theta) = \theta \, \nabla^2 Y_t, \quad t = 3, \ldots, n, \tag{1}$$

where $Y_1, \ldots, Y_n$ is the original time series (non-seasonal or deseasonalised) and $\nabla$ is the difference operator (i.e., $\nabla X_t = X_t - X_{t-1}$). The initial values of $Z_1$ and $Z_2$ are obtained by minimising $\sum_{t=1}^{n}[Y_t - Z_t(\theta)]^2$. However, an analytical solution to compute the $Z(\theta)$ was obtained by H&B, which is given by

$$Z_t(\theta) = \theta Y_t + (1 - \theta)(A_n + B_n t), \quad t = 1, \ldots, n, \tag{2}$$

where $A_n$ and $B_n$ are the minimum square coefficients of a simple linear regression over $Y_1, \ldots, Y_n$ against $1, \ldots, n$, given by

$$A_n = \frac{1}{n} \sum_{t=1}^{n} Y_t - \frac{n+1}{2} B_n;$$

$$B_n = \frac{6}{n^2 - 1} \left( \frac{2}{n} \sum_{t=1}^{n} t \, Y_t - \frac{1+n}{n} \sum_{t=1}^{n} Y_t \right). \tag{3}$$

From this point of view, the theta lines can be interpreted as functions of the linear regression model applied to the data directly. However, note that $A_n$ and $B_n$ are only functions of the original data, not parameters of the Theta method.

Finally, the forecasts produced by the Theta method for $h$ steps ahead of $n$ are an ad-hoc combination (50%–50%) of the extrapolations of $\mathbf{Z}(0)$ and $\mathbf{Z}(2)$ by the linear regression model and the simple exponential smoothing model respectively. We will refer to the above setup as the standard Theta method (STheta).

The steps for building the STheta method of A&N are as follows:

1. *Deseasonalisation*: The time series is tested for statistically significant seasonal behaviour. A time series is seasonal if

$$|r_m| > q_{1-a/2} \sqrt{\frac{1 + 2 \sum_{i=1}^{m-1} r_i^2}{n}},$$

   where $r_k$ denotes the lag $k$ autocorrelation function, $m$ is the number of the periods within a seasonal cycle (for example, 12 for monthly data), $n$ is the sample size, $q$ is the quantile function of the standard normal distribution, and $(1 - a)\%$ is the confidence level. A&N opted for a 90% confidence level. If the time series is identified as seasonal, then it is deseasonalised via the classical decomposition method, assuming the seasonal component to have a multiplicative relationship.[1]
2. *Decomposition*: The seasonally adjusted time series is decomposed into two theta lines, the linear regression line $\mathbf{Z}(0)$ and the theta line $\mathbf{Z}(2)$.
3. *Extrapolation*: $\mathbf{Z}(0)$ is extrapolated as a normal linear regression line, while $\mathbf{Z}(2)$ is extrapolated using SES.
4. *Combination*: The final forecast is a combination of the forecasts of the two theta lines using equal weights.
5. *Reseasonalisation*: If the series was identified as seasonal in step 1, then the final forecasts are multiplied by the respective seasonal indices.

This approach, based on two theta lines with ad-hoc values for the $\theta$ coefficients and equal weights for the recomposition of the final forecasts, resulted in the best performance in the largest forecasting competition to date, the M3-Competition (Makridakis & Hibon, 2000).

---

[1] Arguably, this seasonality test does not work well if a series has one or more unit roots with a slow rate of decay of the autocorrelation function.

## 2.2. SES with drift

Hyndman and Billah (2003) demonstrated that there is a relationship between the STheta method and the Simple Exponential Smoothing with drift model (SES-d) given by

$$Y_t = \ell_{t-1}^{**} + b + \varepsilon_t, \tag{4}$$

$$\ell_t^{**} = \ell_{t-1}^{**} + b + \alpha \varepsilon_t, \tag{5}$$

for $t = 1, \ldots, n$, where $\{\varepsilon_t\}$ is white noise and $(\alpha, b, \ell_0^{**})$ are the smoothing, growth (drift) and initial level parameters respectively.

For a non-seasonal time series, the forecasts produced by STheta and SES-d coincide if

$$b = 0.5 B_n \quad \text{and} \quad \ell_0^{**} = (\ell_0^* + A_n)/2, \tag{6}$$

where $\ell_0^*$ is the initial level parameter of the SES model applied to $\mathbf{Z}(2)$. The second part of Eq. (6) is more general than in H&B's derivation, since they used a simple initialisation for the SES model, i.e., $\ell_0^* = Z_1(2) = 2Y_1 - A_n - B_n$ (or equivalently $\ell_1^{**} = Y_1$).

We can deal with seasonal time series by considering the same prior seasonal test, prior seasonal adjustment and posterior reseasonalisation steps of STheta.

## 2.3. Other generalisations of Theta method

Very few generalisations of the univariate STheta method have been proposed in the literature. For example, Nikolopoulos and Assimakopoulos (2005) and Petropoulos and Nikolopoulos (2013) argue for the use of more theta lines, $\theta \in \{-1, 0, 1, 2, 3\}$, in order to extract even more information from the data. The empirical evidence suggests that the consideration of more/different theta lines can result in improvements relative to the original Theta method, but a formal procedure for the selection of appropriate theta lines is yet to be proposed.

Moreover, Constantinidou et al. (2012) and Petropoulos and Nikolopoulos (2013) suggested the use of unequal weights in the recomposition procedure for the final forecasts. This approach is appealing intuitively, as asymmetric weights, which are linked directly with the forecast horizon, are likely to offer better approximations of the short- and long-term components. However, by definition, the decomposition of the original series in $Z_t(0)$ and $Z_t(2)$ suggests the use of equal weights, if the aim is to reconstruct the original signal:

$$
\begin{aligned}
0.5 Z_t(0) + 0.5 Z_t(2) &= 0.5(A_n + B_n t) + 0.5 \\
&\quad \times [2Y_t - (A_n + B_n t)] \\
&= Y_t.
\end{aligned}
$$

In other words, the use of weights that are derived directly from the decomposition procedure (the corresponding $\theta$ coefficients) may provide a more valid model.

## 3. Models for optimising the Theta method

Assume that either the time series $Y_1, \ldots, Y_n$ is non-seasonal or it has been seasonally adjusted using the

multiplicative classical decomposition approach. Let $X_t$ be the linear combination of two theta lines,

$$X_t = \omega Z_t(\theta_1) + (1-\omega)Z_t(\theta_2), \qquad (7)$$

where $\omega \in [0, 1]$ is the weight parameter. Assuming that $\theta_1 < 1$ and $\theta_2 \geq 1$, the weight $\omega$ can be derived as

$$\omega := \omega(\theta_1, \theta_2) = \frac{\theta_2 - 1}{\theta_2 - \theta_1}. \qquad (8)$$

It is straightforward to see from Eqs. (7) and (8) that $X_t = Y_t$, $t = 1, \ldots, n$, i.e., the weights are calculated properly in such a way that Eq. (7) reproduces the original series. In Theorem 1 of Appendix A, we prove that the solution is unique and that the error from not choosing the optimal weights ($\omega$ and $1 - \omega$) is proportional to the error of a linear regression model. As a consequence, the STheta method is given simply by setting $\theta_1 = 0$ and $\theta_2 = 2$, while from Eq. (8) we get $\omega = 0.5$. Thus, Eqs. (7) and (8) allow us to construct a generalisation of the Theta model that maintains the re-composition propriety of the original time series for any theta lines $Z_t(\theta_1)$ and $Z_t(\theta_2)$.

In order to maintain the modelling of the long-term component and retain a fair comparison with the STheta method, in this work we fix $\theta_1 = 0$ and focus on the optimisation of the short-term component, $\theta_2 = \theta$ with $\theta \geq 1$. Thus, $\theta$ is the only parameter that requires estimation so far. The theta decomposition is now given by

$$Y_t = \left(1 - \frac{1}{\theta}\right)(A_n + B_n t) + \frac{1}{\theta}Z_t(\theta), \quad t = 1, \ldots, n.$$

The $h$-step-ahead forecasts calculated at origin $n$ are given by

$$\widehat{Y}_{n+h|n} = \left(1 - \frac{1}{\theta}\right)[A_n + B_n(n+h)] + \frac{1}{\theta}\widetilde{Z}_{n+h|n}(\theta), \qquad (9)$$

where $\widetilde{Z}_{n+h|n}(\theta) = \widetilde{Z}_{n+1|n}(\theta) = \alpha \sum_{i=0}^{n-1}(1-\alpha)^i Z_{n-i}(\theta) + (1-\alpha)^n \ell_0^*$ is the extrapolation of $Z_t(\theta)$ by an SES model with $\ell_0^* \in \mathbb{R}$ as the initial level parameter and $\alpha \in (0, 1)$ as the smoothing parameter. Note that for $\theta = 2$, Eq. (9) corresponds to Step 4 of the STheta algorithm. After some algebra, we can write

$$\widetilde{Z}_{n+1|n}(\theta) = \theta\,\ell_n + (1-\theta)\left\{A_n\left[1 - (1-\alpha)^n\right]\right.$$
$$\left. + B_n\left[n + \left(1 - \frac{1}{\alpha}\right)\left[1 - (1-\alpha)^n\right]\right]\right\}, (10)$$

where $\ell_t = \alpha Y_t + (1-\alpha)\ell_{t-1}$ for $t = 1, \ldots, n$ and $\ell_0 = \ell_0^*/\theta$.

In the light of Eqs. (9) and (10), we suggest four stochastic approaches. These approaches differ due to the parameter $\theta$, which may be either fixed at two or optimised, and the coefficients $A_n$ and $B_n$, which can be either fixed or dynamic functions. To formulate the state space models, it is helpful to adopt $\mu_t$ as the one-step-ahead forecast at origin $t - 1$ and $\varepsilon_t$ as the respective additive error, i.e., $\varepsilon_t = Y_t - \mu_t$ if $\mu_t = \widehat{Y}_{t|t-1}$. We assume $\{\varepsilon_t\}$ to be a Gaussian white noise process with mean zero and variance $\sigma^2$.

### 3.1. Optimised and standard Theta models

Let $A_n$ and $B_n$ be fixed coefficients for all $t = 1, \ldots, n$, so that Eqs. (9) and (10) configure the state space model

given by

$$Y_t = \mu_t + \varepsilon_t, \qquad (11)$$

$$\mu_t = \ell_{t-1} + \left(1 - \frac{1}{\theta}\right)$$
$$\times \left\{(1-\alpha)^{t-1}A_n + \left[\frac{1 - (1-\alpha)^t}{\alpha}\right]B_n\right\}, \qquad (12)$$

$$\ell_t = \alpha Y_t + (1-\alpha)\ell_{t-1}, \qquad (13)$$

with parameters $\ell_0 \in \mathbb{R}$, $\alpha \in (0, 1)$ and $\theta \in [1, \infty)$. The parameter $\theta$ is to be estimated along with $\alpha$ and $\ell_0$. We call this the optimised Theta model (OTM).

The $h$-step-ahead forecast at origin $n$ is given by

$$\widehat{Y}_{n+h|n} = E[Y_{n+h}|Y_1, \ldots, Y_n]$$
$$= \ell_n + \left(1 - \frac{1}{\theta}\right)\left\{(1-\alpha)^n A_n\right.$$
$$\left. + \left[(h-1) + \frac{1 - (1-\alpha)^{n+1}}{\alpha}\right]B_n\right\},$$

which is equivalent to Eq. (9). The conditional variance $Var[Y_{n+h}|Y_1, \ldots, Y_n] = [1 + (h-1)\alpha^2]\sigma^2$ can be computed easily from the state space model. Thus, the $(1 - a)\%$ prediction interval for $Y_{n+h}$ is given by

$$\widehat{Y}_{n+h|n} \pm q_{1-a/2}\sqrt{[1 + (h-1)\alpha^2]\sigma^2}.$$

For $\theta = 2$, OTM reproduces the forecasts of the STheta method; hereafter, we will refer to this particular case as the standard Theta model (STM). In Theorem 2 of Appendix A, we show that OTM is mathematically equivalent to the SES-d model. As a corollary of Theorem 2, STM is mathematically equivalent to SES-d with $b = \frac{1}{2}B_n$. Therefore, for $\theta = 2$, the corollary also re-confirms the H&B result on the relationship between STheta and the SES-d model.

### 3.2. Dynamic optimised and dynamic standard Theta models

So far, we have set $A_n$ and $B_n$ as fixed coefficients for all $t$. We will now consider these coefficients as dynamic functions; i.e., for updating the state $t$ to $t + 1$ we will only consider the prior information $Y_1, \ldots, Y_t$ when computing $A_t$ and $B_t$. Hence, we replace $A_n$ and $B_n$ in Eqs. (9) and (10) with $A_t$ and $B_t$. Then, after applying the new Eq. (10) to the new Eq. (9) and rewriting the result at time $t$ with $h = 1$, we have

$$\widehat{Y}_{t+1|t} = \ell_t + \left(1 - \frac{1}{\theta}\right)$$
$$\times \left\{(1-\alpha)^t A_t + \left[\frac{1 - (1-\alpha)^{t+1}}{\alpha}\right]B_t\right\}. \qquad (14)$$

Then, assuming additive one-step-ahead errors and rewriting Eqs. (3) and (14), we obtain

$$Y_t = \mu_t + \varepsilon_t \qquad (15)$$

$$\mu_t = \ell_{t-1} + \left(1 - \frac{1}{\theta}\right)$$
$$\times \left[(1-\alpha)^{t-1}A_{t-1} + \left(\frac{1 - (1-\alpha)^t}{\alpha}\right)B_{t-1}\right] \qquad (16)$$

$$\ell_t = \alpha Y_t + (1 - \alpha)\ell_{t-1} \tag{17}$$

$$A_t = \bar{Y}_t - \frac{t+1}{2}B_t \tag{18}$$

$$B_t = \frac{1}{t+1}\left[(t-2)B_{t-1} + \frac{6}{t}(Y_t - \bar{Y}_{t-1})\right], \tag{19}$$

$$\bar{Y}_t = \frac{1}{t}[(t-1)\bar{Y}_{t-1} + Y_t] \tag{20}$$

for $t = 1, \ldots, n$. Eqs. (15)–(20) configure a state space model with parameters $\ell_0 \in \mathbb{R}, \alpha \in (0, 1)$ and $\theta \in [1, \infty)$. The initialisation of the states is performed assuming $A_0 = B_0 = B_1 = \bar{Y}_0 = 0$. From here on, we will refer to this model as the dynamic optimised Theta model (DOTM).

An important property of the DOTM is that when $\theta = 1$, which implies that $Z_t(1) = Y_t$, the forecasting vector given by Eq. (9) will be equal to $\widehat{Y}_{t+h|t} = \widetilde{Z}_{t+h|t}(1)$. Thus, when $\theta = 1$, the DOTM is the SES method. When $\theta > 1$, DOTM (as SES-d) acts as a extension of SES, by adding a long-term component. Also, for $\theta = 2$, we have a stochastic approach of STheta, which is referred to hereafter as the dynamic standard Theta model (DSTM).

The out-of-sample one-step-ahead forecasts produced by DOTM at origin $n$ are given by

$$\widehat{Y}_{n+1|n} = E[Y_{n+1}|Y_1, \ldots, Y_n]$$

$$= \ell_n + \left(1 - \frac{1}{\theta}\right)$$

$$\times \left\{(1-\alpha)^n A_n + \left[\frac{1-(1-\alpha)^{n+1}}{\alpha}\right]B_n\right\}; \tag{21}$$

for a horizon $h \geq 2$, the forecasts $\widehat{Y}_{n+2|n}, \ldots, \widehat{Y}_{n+h|n}$ are computed recursively using Eqs. (16)–(21) by replacing the non-observed values $Y_{n+1}, \ldots, Y_{n+h-1}$ with their expected values $\widehat{Y}_{n+1|n}, \ldots, \widehat{Y}_{n+h-1|n}$. The conditional variance $Var[Y_{n+h}|Y_1, \ldots, Y_n]$ is hard to write analytically. However, the variance and prediction intervals for $Y_{n+h}$ can be estimated using the bootstrapping technique, where a (usually large) sample of possible values of $Y_{n+h}$ is simulated from the estimated model.

Note that, in contrast to STheta, STM and OTM, the forecasts produced by DSTM and DOTM are not necessary linear. This is also a fundamental difference between DSTM/DOTM and SES-d: while the long-term trend ($b$) in SES-d is constant, this is not the case for DSTM/DOTM, for either the in-sample fit or the out-of-sample predictions.

### 3.3. Parameter estimation

The parameters are estimated by minimising the sum of squared errors (SSE),

$$(\widehat{\ell_0}, \widehat{\alpha}, \widehat{\theta}) = \arg\min_{\ell_0, \alpha, \theta} \sum_{t=1}^{n} \varepsilon_t^2 = \arg\min_{\ell_0, \alpha, \theta} \sum_{t=1}^{n}(Y_t - \mu_t)^2.$$

Of course, the SSE does not necessarily need to start at $t = 1$. We suggest starting at $t = 3$ for DSTM/DOTM, since $A_t$ and $B_t$ are linear regression coefficients and need at least two points in order to be well defined.

The SSE estimator is equivalent to a maximum likelihood estimator. This result follows from the supposition of Gaussian distributed errors, since, after replacing $\sigma^2$ with its estimator $\widehat{\sigma^2} = SSE/n$, the log-likelihood is given by

$$l(\ell_0, \alpha, \theta) = -\frac{n}{2}\log\widehat{\sigma^2} - \frac{n}{2}(1 + \log 2\pi).$$

In this section, we have proposed the STM, OTM, DSTM, and DOTM, four very simple and easy to implement models. The latter three models expand the robust Theta method of A&N, and all four build on the state space approach (Hyndman et al., 2002). These models use just two theta lines, with OTM and DOTM optimising the amplification of the local curvature. The forecasts derived from these theta lines are combined optimally so as to retain the re-composition of the original signal. In the next section, we evaluate the performances of the proposed models on the M3-Competition data set.

## 4. Empirical evaluation

### 4.1. Design

In order to obtain insights into the performances of the proposed models, STM, OTM, DSTM and DOTM, we present their accuracies compared to each other and to the STheta and SES-d approaches. A full list of the methods and models considered is presented in Table 1, along with the starting values for optimising the various parameters. Note that, to mimic what might be done in practice, the starting values are based on the model being used, and do not correspond mathematically to the mathematically equivalent models/method.

We consider two variants of the SES-d model. The first considers a fixed value for $b$ (equal to $B_n/2$). Assuming perfect optimisers, we should expect this version to produce the same forecasts as STheta and STM. The second version optimises the value of $b$ and is mathematically equivalent to OTM. Perfect optimisers should also produce the same forecasts for these two models; that is, choices such as the starting values for the parameters should not matter. However, we know that different starting values may affect the optimal value of a parameter even for the same model. The parameter estimation is based on minimising the sum of squared errors (SSE) using the Nelder–Mead algorithm, as implemented in the `optim()` function of the R statistical software.

Moreover, we consider five benchmarks that have been used widely in the forecasting literature. A full list and details of the benchmark methods considered are presented in Table 2, including automatic algorithms implemented in the *forecast* package by Hyndman and Khandakar (2008).

The various Theta methods and models listed in Table 1 are applied to the seasonally adjusted data, with the final forecasts being reseasonalised following the procedure described in Section 2 (steps 1 and 5). The five benchmark methods (Naive, SES, Damped, ETS and ARIMA) are applied to both the original data and the seasonally adjusted data, where the same deseasonalisation/reseasonalisation procedure has been followed. In all cases, the seasonally adjusted data and the seasonal indices to be used for the reseasonalisation are calculated by considering only

**Table 1**
The different Theta methods and models considered in the empirical evaluation.

| Method/model | Section | Equations | Starting values for parameter optimisation |
|---|---|---|---|
| STheta | 2.1 | | $\ell_0^* = A_{10}, \alpha = 0.5$ |
| SES-d ($b = B_n/2$) | 2.2 | (4)–(5), where $b = B_n/2$ | $\ell_0^{**} = A_{10}, \alpha = 0.5$ |
| SES-d ($b$ optimised) | 2.2 | (4)–(5) | $\ell_0^{**} = A_{10}, \alpha = 0.5, b = B_{10}$ |
| STM | 3.1 | (11)–(13), where $\theta = 2$ | $\ell_0 = y_1/2, \alpha = 0.5$ |
| OTM | 3.1 | (11)–(13) | $\ell_0 = y_1/2, \alpha = 0.5, \theta = 2$ |
| DSTM | 3.2 | (15)–(20), where $\theta = 2$ | $\ell_0 = y_1/2, \alpha = 0.5$ |
| DOTM | 3.2 | (15)–(20) | $\ell_0 = y_1/2, \alpha = 0.5, \theta = 2$ |

**Table 2**
The benchmark methods used in the current study.

| Method | Reference | Description |
|---|---|---|
| Naive | | ARIMA(0,1,0) |
| SES | Brown (1956) | ETS(A,N,N) |
| Damped | Gardner and McKenzie (1985) | ETS(A,Ad,N) |
| ETS | Hyndman and Khandakar (2008) | ETS automatic algorithm based on AICc |
| ARIMA | Hyndman and Khandakar (2008) | Automatic ARIMA based on AICc |

the in-sample data points (training set) and setting the confidence level to 90%. Adjusting the data for seasonality prior to forecasting has also been the practice in other forecasting studies, such as the M3-Competition (Makridakis & Hibon, 2000). However, we believe that a higher confidence level (95%) has been used for identifying series as seasonal.

The evaluation is performed by considering real data from the M3-Competition (Makridakis & Hibon, 2000), giving a total of 3003 time series of different frequencies. Table 3 presents the distribution of the series across the different frequencies. The forecast horizon used in this study matched that of the original M3-Competition. The empirical evaluation was implemented using the open-source statistical software provided by R Core Team (2015) (version 3.2.1), and the packages *forecast 6.1* and *Mcomp 0.10-34*. The computer used for this task was equipped with an Intel i5-4200U processor with 8 GB of RAM, and was running Windows 10.

We measured the out-of-sample performances of the different methods using two widely-used accuracy metrics, namely the symmetric mean absolute percentage error (sMAPE) and the mean absolute scaled error (MASE). The sMAPE is selected in order to enable us make comparisons with the original M3-Competition results, even though it penalises positive forecast errors more heavily than negative ones, with the discrepancy increasing at an increasing rate (see Figure 1 of Goodwin & Lawton, 1999). The calculation of the sMAPE is based on the symmetric absolute percentage error (sAPE), defined as

$$sAPE = 200 \frac{|Y_{n+i} - \widehat{Y}_{n+i|n}|}{|Y_{n+i}| + |\widehat{Y}_{n+i|n}|}.$$

The MASE metric was proposed by Hyndman and Koehler (2006), and is calculated based on the absolute scaled error (ASE), which is the absolute error divided by the mean of the absolute of the first seasonal difference of the time series, i.e.,

$$ASE = \frac{|Y_{n+i} - \widehat{Y}_{n+i|n}|}{(n-m)^{-1} \sum_{t=m+1}^{n} |Y_t - Y_{t-m}|},$$

**Table 3**
M3-Competition dataset.

| Frequency | Forecasting horizon ($h$) | Number of time series |
|---|---|---|
| Yearly | 6 | 645 |
| Quarterly | 8 | 756 |
| Monthly | 18 | 1428 |
| Other | 8 | 174 |
| Total | | 3003 |

where *m* is the number of periods in a year (one for yearly, four for quarterly, 12 for monthly and one for other data). In both sMAPE and MASE, the mean is calculated across both horizons and time series at once. This means that we assign equal weights to the point forecast errors of yearly and monthly series.

### 4.2. Results

The results regarding the forecasting performances of the various methods are presented in Table 4. The best result in each frequency (column) is marked in bold. We highlight the results for the four models proposed in this study in grey shading.

Focusing on the non-shaded panels, the STheta method had the best performance across all benchmarks, as expected, according to the sMAPE measure. Any numerical differences from the results published in the M3-Competition (Makridakis & Hibon, 2000) can be attributed to the use of different pre-fixed theta coefficients and extrapolation methods for each frequency of the data (see Nikolopoulos et al., 2011, for more details). The small differences in the monthly data are due to the use of different software and estimation procedures for the smoothing parameters and the initial level when extrapolating the $\mathbf{Z}(2)$.

This is a close reproduction of the Theta method as it was applied to the monthly data in the M3-Competition. The sMAPE of STheta for the monthly time series in this study is 13.83%, versus the value of 13.85% published by Makridakis and Hibon (2000). Moreover, we managed to obtain the exact same populations of seasonal quarterly

**Table 4**
Empirical results for all methods using the sMAPE and the MASE.

| Data | Methods | Yearly | Quarterly | Monthly | Other | All | Time (min) |
|---|---|---|---|---|---|---|---|
| sMAPE metric (%) | | | | | | | |
| Original data | Naive | 17.88 | 11.32 | 18.18 | 6.30 | 16.58 | 0.31 |
| | SES | 17.78 | 10.83 | 16.14 | 6.30 | 15.07 | 0.22 |
| | Damped | 17.07 | 10.96 | 16.25 | **4.30** | 15.02 | 1.05 |
| | ETS | 16.89 | 9.69 | 14.07 | 4.34 | 13.28 | 38.36 |
| | ARIMA | 17.62 | 9.99 | 15.30 | 4.54 | 14.27 | 23.48 |
| Seasonally adjusted data | Naive | 17.88 | 10.02 | 16.76 | 6.30 | 15.38 | 0.62 |
| | SES | 17.78 | 9.77 | 14.17 | 6.30 | 13.53 | 0.55 |
| | Damped | 17.07 | 9.79 | 13.96 | **4.30** | 13.24 | 1.52 |
| | ETS | 16.89 | 9.80 | 14.04 | 4.34 | 13.28 | 32.87 |
| | ARIMA | 17.62 | 9.74 | 15.68 | 4.54 | 14.49 | 21.40 |
| | STheta | 16.74 | 9.23 | 13.83 | 4.93 | 13.05 | 7.37 |
| | SES-d ($b = B_n/2$) | 16.72 | 9.23 | 13.87 | 4.93 | 13.08 | 10.15 |
| | SES-d ($b$ optimised) | 17.19 | 9.40 | 15.06 | 5.00 | 13.98 | 12.60 |
| | STM | 16.73 | 9.24 | 13.85 | 4.93 | 13.06 | 8.34 |
| | OTM | 16.60 | **9.14** | 14.11 | 4.85 | 13.21 | 10.05 |
| | DSTM | 16.69 | 9.24 | 13.82 | 4.92 | 13.04 | 10.86 |
| | DOTM | **15.94** | 9.28 | **13.74** | 4.58 | **12.90** | 13.59 |
| MASE metric | | | | | | | |
| Original data | Naive | 3.17 | 1.46 | 1.17 | 3.09 | 1.50 | 0.31 |
| | SES | 3.17 | 1.41 | 1.09 | 3.10 | 1.43 | 0.22 |
| | Damped | 2.92 | 1.37 | 1.10 | 1.81 | 1.36 | 1.05 |
| | ETS | 2.83 | 1.18 | 0.86 | **1.79** | 1.15 | 38.36 |
| | ARIMA | 2.99 | 1.17 | 0.88 | 1.87 | 1.19 | 23.48 |
| Seasonally adjusted data | Naive | 3.17 | 1.25 | 1.04 | 3.09 | 1.37 | 0.62 |
| | SES | 3.17 | 1.24 | 0.93 | 3.10 | 1.29 | 0.55 |
| | Damped | 2.92 | 1.17 | 0.88 | 1.81 | 1.17 | 1.52 |
| | ETS | 2.83 | 1.18 | 0.87 | **1.79** | 1.16 | 32.87 |
| | ARIMA | 2.99 | 1.15 | 0.89 | 1.87 | 1.19 | 21.40 |
| | STheta | 2.77 | 1.12 | 0.86 | 2.28 | 1.16 | 7.37 |
| | SES-d ($b = B_n/2$) | 2.77 | 1.12 | 0.86 | 2.27 | 1.16 | 10.15 |
| | SES-d ($b$ optimised) | 2.69 | 1.11 | 0.86 | 2.07 | 1.14 | 12.60 |
| | STM | 2.77 | 1.12 | 0.86 | 2.27 | 1.16 | 8.34 |
| | OTM | 2.71 | **1.10** | 0.86 | 2.23 | 1.14 | 10.05 |
| | DSTM | 2.76 | 1.12 | 0.86 | 2.27 | 1.16 | 10.86 |
| | DOTM | **2.59** | 1.12 | **0.85** | 1.94 | **1.12** | 13.59 |

and monthly series (555 and 780 respectively) as those reported by Nikolopoulos and Assimakopoulos (2005), by rounding the critical value of the $t$-statistic for identifying a series as seasonal to two decimal places (1.64). Thus, this paper also contributes to the replicability/reproducibility agenda (Boylan, Goodwin, Mohammadipour, & Syntetos, 2015).

This is also the first study to demonstrate the practical equivalence of the STheta method and the SES-d model, when the value of $b$ is fixed to $B_n/2$. The very minor numerical differences are attributed to the use of non-perfect optimisers, resulting in the selection of different optimal values for the initial level, $\ell^*$ and $\ell^{**}$, and the smoothing parameter, $\alpha$.

Focusing on the sMAPE measure for the benchmark methods (non-shaded panel), the superior performances of the STheta method and the SES-d ($b = B_n/2$) model are followed by that of Damped (on the seasonally adjusted data) and ETS. Examining the results across the various frequencies, STheta and SES-d ($b = B_n/2$) perform especially well for monthly and yearly data, while underperforming relative to Damped and ETS for the other data.

Considering the MASE metric, STheta and SES-d ($b = B_n/2$) once again perform very similarly. However, the best performer of the benchmark methods (non-shaded panel) is the SES-d ($b$ optimised) model, which performed rather poorly according to the sMAPE. This discrepancy between the two measures is due to the properties of the sMAPE, and the heavy penalty that it places on large positive errors. To demonstrate this, Appendix B (Table 7) also reports the median value of the symmetric absolute percentage errors (sMdAPE) across horizons and time series. We observe that, apart from SES-d ($b$ optimised), the relative rankings of various other methods are also improved (namely ETS, ARIMA and Damped applied to the seasonally adjusted data). Thus, we conclude that the results reported by the sMAPE are influenced heavily by the presence of positive symmetric percentage errors that are outliers.

As expected, the STM generates results that are equivalent to those of both the STheta method and the SES-d ($b = B_n/2$) model. Similarly, the performance of OTM is similar to that of SES-d ($b$ optimised) for the MASE metric, apart from the other data. At the same time, the OTM is less susceptible to resulting in outlying positive errors than

*J.A. Fiorucci et al. / International Journal of Forecasting 32 (2016) 1151–1161*

**Table 5**
Empirical results for SES-d (*b* optimised) and OTM for the trimmed sMAPE.

| Methods | Yearly | Quarterly | Monthly | Other |
|---|---|---|---|---|
| SES-d (*b* optimised) | 13.62 | 7.04 | 11.79 | 3.30 |
| OTM | 13.53 | 6.83 | 11.56 | 3.43 |

SES-d (*b* optimised), and thus performs better according to the sMAPE measure. This is apparent from the median and trimmed values of sMAPE reported in Appendix B (Table 7) and Table 5, respectively. The trimmed values of sMAPE are calculated as the arithmetic mean of the point forecast errors, after excluding the 10% of the series that produced the highest sMAPE values and the 10% that produced the lowest.

There are several reasons for the divergent results between OTM and SES-d (*b* optimised). First, the parameter space of *b* in SES-d does not correspond exactly to the $\theta$ parameter space in OTM. Second, the starting values in the optimiser (see Table 1) are natural ones for each model, and do not correspond mathematically (see Theorem 2 of Appendix A). These different starting values will contribute to both the differences in the comparison metrics that are caused by suboptimal solutions, and the differences in computational times (see Table 4). Third, the same increments for the parameters in the optimiser do not correspond mathematically in the two models.

Focusing on the two dynamic models (DSTM and DOTM), the DOTM produces the most accurate forecasts, outperforming the other methods and models in this study overall. This is true for all three error measures considered (sMAPE, sMdAPE and MASE). The DOTM outperforms all other Theta variants significantly for the yearly and other frequencies. This is a very interesting result, as the series classified as other was the single data category for which STheta did not perform well compared to the other benchmarks. However, DOTM is still outperformed in this category by ETS, Damped, and ARIMA. The only data frequency for which the DOTM does not improve on STheta is the quarterly time series, but the performances of the two are very similar.

We also considered the multiple comparisons with the best (MCB) test for all frequencies in order to compare DSTM, OTM and DOTM with STheta and SES-d (*b* optimised) statistically. In this test, a rank interval is constructed for each method (see Koning et al., 2005, for more details) using the mean absolute error (which will give ranks equivalent to those of the MASE). When the rank intervals considering pairs of methods do not overlap, the null hypothesis of the same performances is rejected in favour of the alternative hypothesis of significantly different performances. The average ranks and rank intervals of each method are presented in Fig. 1, which also presents a comparison of the average rank of each method with the best average rank, adopting a 10% significance level.

In line with our insights when examining the summarised results of the sMAPE and MASE measures, the DOTM provides better performances than the other four approaches, being ranked significantly higher. The models SES-d (*b* optimised) and OTM are slightly better (not significantly different) than either STheta or DSTM.

**Table 6**
Percentage improvements of DOTM over DSTM in terms of MASE (the numbers in brackets refer to sample sizes).

| Frequency | All (*n*) | Trended (*n*) | Non-trended (*n*) |
|---|---|---|---|
| Yearly | 6.42% (645) | 10.21% (380) | −0.64% (265) |
| Quarterly | 0.22% (756) | 1.83% (442) | −1.81% (314) |
| Monthly | 1.38% (1428) | 3.63% (698) | −0.77% (730) |
| Other | 14.44% (174) | 19.85% (113) | 3.58% (61) |

### 4.3. Discussion

Previous studies have shown that the Theta method is particularly efficient for trended data (Thomakos & Nikolopoulos, 2014). In order to obtain a better understanding of where the improvements are derived from, we split the data into non-trended and trended series, then calculate the percentage decrease in the value of the MASE when comparing DSTM and DOTM for each type of data. The categorisation of a series as trended or not is based directly on the model form chosen by the ETS algorithm applied to the original data. Table 6 presents the percentage drops in the value of MASE (increase in accuracy) of DOTM over DSTM. We observe that the performance improvements are driven mainly by the trended series. In such cases, DOTM outperforms DSTM by 10.21% and 19.85% for yearly and other data respectively. Also, larger than average improvements are recorded for the monthly and quarterly trended series. Similar insights were obtained when contrasting the performances of the two methods using the sMAPE measure.

However, the question remains as to why the DOTM performs better than the DSTM. The optimisation of the $\theta$ value for the second theta line is linked directly to the amplification of the local curvatures of the series (A&N). The quite arbitrary selection of $\theta = 2$ suggests that the long-term deterministic trend is just as important as the short-term behaviour of the series, which might not be the case for all time series. Therefore, DOTM selects the degree of amplification of the short-term behaviour of the series optimally. An analysis of the optimally-selected $\theta$ values shows that when $\theta \leq 2$ (58% of the series), the average performance improvement is −0.12% (percentage difference between the DOTM and DSTM performances in terms of MASE). However, the improvement increases to 7.57% and 9.94% for the series where $\theta > 2$ (42%) and $\theta > 3$ (31%) respectively. Similar insights are obtained for the sMAPE metric. Thus, we observe that DOTM works particularly well when there is a need to consider higher theta values that capture and model the short-term behaviour of the series effectively.

Of course, one could argue that the OTM also considers optimal values of the $\theta$ parameter. However, OTM falls short of DOTM with regard to the stochasticity aspect of the linear regression part of the model, and its effect on the selection of an optimal $\theta$ value. To test this, we calculate the absolute percentage differences of the optimal $\theta$ values for OTM and DOTM, and divide the series into two equally sized groups, corresponding to small and large differences. The performance improvement of DOTM over OTM, measured via the MASE, increases from 0.39% for small differences in the optimised $\theta$ values to 3.44%
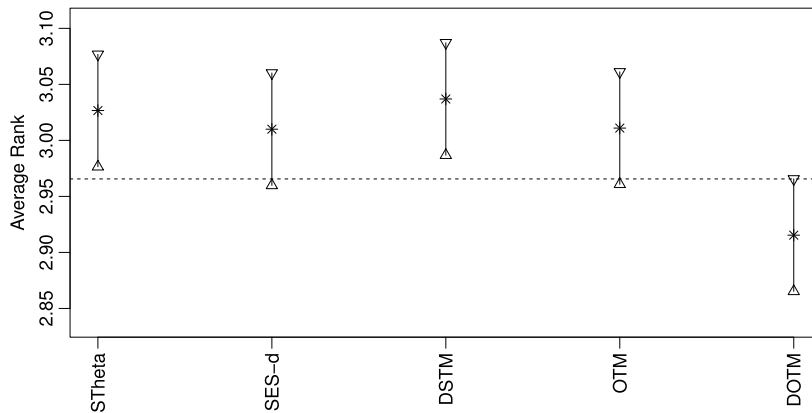
**Fig. 1.** MCB intervals for selected forecasting methods.

for large differences. As a result, the dynamic updating of $A_t$ and $B_t$ has a positive impact on the optimisation of the $\theta$ values, especially when a significantly different value is selected.

In terms of the computational times achieved, DOTM and DSTM are more computational intensive than the original Theta model, as expected. It could be argued that the $\times 1.5$ additional computational cost of DSTM over STheta is not worthwhile, given the marginal gains in forecasting performance. However, the robust performance of DOTM clearly pays off. In any case, the calculation times of both models are significantly lower than those of the two automatic model selection algorithms (`ets()` and `auto.arima()`) implemented using the *forecast* package (Hyndman & Khandakar, 2008).

## 5. Concluding remarks

In this paper, we have proposed a generalisation of the Theta method, namely the dynamic optimised Theta model. The DOTM selects the theta line to be used for the extrapolation of the short-term component of the series optimally, and also revises the $A_t$ and $B_t$ in the long-term component at each time period $t$. In addition, the proposed model is provided under a state space approach, which allows already consolidated statistical tools to be used for parameter estimation. The newly proposed model was contrasted with the original Theta method and other variants such as the SES-d model, both theoretically and empirically.

In terms of empirical forecasting performances, DOTM demonstrated improvements over the Theta method for all frequencies and error measures considered. At the same time, it was also the top performing approach across all benchmarks overall. Moreover, DOTM produced the best average ranking, which is statistically different from that of the original Theta method. Apart from the very promising empirical results for the DOTM, this study replicates the Theta method for the monthly time series in the M3-Competition. We also proved the mathematical equivalence of special cases of the DOTM, and compared their empirical forecasts in order to examine how much the optimiser affects the forecasts.

We believe that this study has significant managerial implications. We show that the new, optimised version of the Theta method improves on the forecasting performance of the original approach. Keeping in mind that the original Theta model was already a very good and robust estimator for fast-moving demand time series, the new DOTM achieves even higher levels of forecasting accuracy, which can be translated directly into profits. Thus, the DOTM is able to provide better statistical estimates, which can then be combined with judgemental overrides (Franses & Legerstee, 2011) in order to produce the final (operational) forecasts.

DOTM could be extended further by considering the appropriate selection of extrapolation methods for the theta lines, rather than using pre-fixed estimators, such as a linear regression line for $\mathbf{Z}(0)$ and SES for $\mathbf{Z}(\theta)$. Another path for future research should include the application of the proposed DOTM to a data set that is dominated by stationary data (Thomakos & Nikolopoulos, 2014). In addition, the current seasonality test should be revisited to allow it to distinguish between additive and multiplicative seasonality, and to work well if the series has one or more unit roots. Another interesting study would have the goal of obtaining an understanding of the differences between DOTM and the state space model for exponential smoothing with a stochastic trend (ETS(A,A,N)). Since DOTM outperforms the damped trend model ETS(A,Ad,N), it would be expected to outperform the ETS(A,A,N) model. Discovering the reasons why could be enlightening.

## Appendix A

**Theorem 1.** *Let $\theta_1 < 1$ and $\theta_2 \geq 1$. We will prove that*

(a) *the linear system given by $X_t = Y_t$ for all $t = 1, \ldots, n$, where $X_t$ is given by Eq. (7), has the single solution*

$$\omega = (\theta_2 - 1)/(\theta_2 - \theta_1);$$

(b) *the error of choosing a non-optimal weight $\omega_\delta = \omega + \delta$ is proportional to the error for a simple linear regression model.*

**Proof (a).** First, note that Eq. (2) can be rewritten as

$$Z_t(\theta) = \theta(Y_t - A_n - B_n t) + A_n + B_n t.$$

Thus, the linear system given by

$$X_t = Y_t, \quad \text{for all } t \in \{1, \ldots, n\}$$

implies

$$
\begin{aligned}
0 &= X_t - Y_t \\
&= \omega[\theta_1(Y_t - A_n - B_n t) + A_n + B_n t] + (1 - \omega) \\
&\quad \times [\theta_2(Y_t - A_n - B_n t) + A_n + B_n t] - Y_t \\
&= \omega\theta_1(Y_t - A_n - B_n t) + (1 - \omega)\theta_2(Y_t - A_n - B_n t) \\
&\quad - (Y_t - A_n - B_n t) \\
&= [\omega\theta_1 + (1 - \omega)\theta_2 - 1](Y_t - A_n - B_n t), \\
&\quad \text{for all } t \in \{1, \ldots, n\}.
\end{aligned}
$$

Note that $Y_t - A_n - B_n t$ is the residual of the linear regression model, and cannot be zero for all $t \in \{1, \ldots, n\}$. Hence, the upper equations are true if, and only if, $\omega\theta_1 + (1 - \omega)\theta_2 - 1 = 0$, which implies $\omega = (\theta_2 - 1)/(\theta_2 - \theta_1)$, as we want to show. □

**Proof (b).** The combination of theta lines associated with the weights $(\omega_\delta, 1 - \omega_\delta)$ is given by $X_t(\omega_\delta) = \omega_\delta Z_t(\theta_1) + (1 - \omega_\delta)Z_t(\theta_2)$. Note that, using (a), we have $X_t(\omega_0) = X_t(\omega) = Y_t$ for all $t = 1, \ldots, n$.

Let $\varepsilon_t(\delta) = Y_t - X_t(\omega_\delta)$ be the error associated with choosing $\omega_\delta$ rather than $\omega_0$. Then

$$
\begin{aligned}
\varepsilon_t(\delta) &= Y_t - X_t(\omega_\delta) \\
&= Y_t - (\omega + \delta)Z_t(\theta_1) - (1 - \omega - \delta)Z_t(\theta_2) \\
&= Y_t - X_t(\omega) + \delta[Z_t(\theta_2) - Z_t(\theta_1)] \\
&= \delta(\theta_2 - \theta_1)e_t,
\end{aligned}
$$

where $e_t = Y_t - (A_n + B_n t)$ is the residual of the linear regression at time $t$. □

**Theorem 2.** *The SES-d($\ell_0^{**}, \alpha, b$) model, where $\ell_0^{**} \in \mathbb{R}$, $\alpha \in (0, 1)$ and $b \in \mathbb{R}$, is equivalent to OTM($\ell_0, \alpha, \theta$), where $\ell_0 \in \mathbb{R}$ and $\theta \geq 1$, if*

$$\ell_0^{**} = \ell_0 + \left(1 - \frac{1}{\theta}\right)A_n \quad \text{and} \quad b = \left(1 - \frac{1}{\theta}\right)B_n.$$

**Proof.** Let $J_{t-1} = \left(1 - \frac{1}{\theta}\right)\left\{(1 - \alpha)^{t-1}A_n + \left[\frac{1 - (1-\alpha)^t}{\alpha}\right]B_n\right\}$. The OTM can be rewritten as

$$Y_t = \ell_{t-1} + J_{t-1} + \varepsilon_t$$

$$\ell_t = \alpha Y_t + (1 - \alpha)\ell_{t-1}$$

$$J_t = (1 - \alpha)J_{t-1} + \left(1 - \frac{1}{\theta}\right)B_n,$$

where $J_0 = \left(1 - \frac{1}{\theta}\right)(A_n + B_n)$. Then, just taking $\ell_t^{**} = \ell_t + J_t - \left(1 - \frac{1}{\theta}\right)B_n$, we can rewrite the last equations as

$$Y_t = \ell_{t-1}^{**} + \left(1 - \frac{1}{\theta}\right)B_n + \varepsilon_t;$$

$$\ell_t^{**} = \ell_{t-1}^{**} + \left(1 - \frac{1}{\theta}\right)B_n + \alpha\varepsilon_t,$$

which is the SES-d model with $\ell_0^{**} = \ell_0 + \left(1 - \frac{1}{\theta}\right)A_n$ and $b = \left(1 - \frac{1}{\theta}\right)B_n$, as we want to show. □

## Appendix B

See Table 7.

**Table 7**
Empirical results for all methods using the sMdAPE.

| sMdAPE metric (%) | | | | | | | |
|---|---|---|---|---|---|---|---|
| Data | Methods | Yearly | Quarterly | Monthly | Other | All | Time (min) |
| Original data | Naive | 10.92 | 5.50 | 8.84 | 4.04 | 8.09 | 0.31 |
| | SES | 11.07 | 5.63 | 8.28 | 4.01 | 7.77 | 0.22 |
| | Damped | 9.23 | 5.27 | 7.83 | **1.86** | 7.10 | 1.05 |
| | ETS | 9.21 | 4.50 | **6.65** | 1.91 | 6.12 | 38.36 |
| | ARIMA | 9.59 | 4.45 | 6.86 | 1.92 | 6.23 | 23.48 |
| Seasonally adjusted data | Naive | 10.92 | 5.01 | 8.13 | 4.04 | 7.48 | 0.62 |
| | SES | 11.07 | 5.13 | 7.31 | 4.01 | 7.04 | 0.55 |
| | Damped | 9.23 | 4.48 | 6.80 | **1.86** | 6.23 | 1.52 |
| | ETS | 9.21 | 4.54 | 6.67 | 1.91 | 6.14 | 32.87 |
| | ARIMA | 9.59 | 4.35 | 6.94 | 1.92 | 6.24 | 21.40 |
| | STheta | 9.62 | 4.41 | 6.77 | 2.53 | 6.30 | 7.37 |
| | SES-d ($b = B_n/2$) | 9.60 | 4.40 | 6.81 | 2.53 | 6.30 | 10.15 |
| | SES-d ($b$ optimised) | 9.39 | 4.32 | 6.82 | 2.14 | 6.21 | 12.60 |
| | STM | 9.60 | 4.39 | 6.79 | 2.53 | 6.30 | 8.34 |
| | OTM | 9.28 | **4.26** | 6.78 | 2.40 | 6.23 | 10.05 |
| | DSTM | 9.63 | 4.36 | 6.80 | 2.55 | 6.31 | 10.86 |
| | DOTM | **8.63** | 4.45 | 6.73 | 2.14 | **6.11** | 13.59 |

# References

Assimakopoulos, V. (1995). A sucessive filtering technique for identifying long-term trends. *Journal of Forecasting*, *14*, 35–43.

Assimakopoulos, V., & Nikolopoulos, K. (2000). The theta model: a decomposition approach to forecasting. *International Journal of Forecasting*, *16*(4), 521–530.

Athanasopoulos, G., Hyndman, R. J., Song, H., & Wu, D. C. (2011). The tourism forecasting competition. *International Journal of Forecasting*, *27*, 822–844.

Boylan, J., Goodwin, P., Mohammadipour, M., & Syntetos, A. (2015). Reproducibility in forecasting research. *International Journal of Forecasting*, *31*, 79–90.

Brown, R. G. (1956). *Exponential smoothing for predicting demand*. Cambridge, Massachusetts: Arthur D. Little Inc.

Clemen, R. T. (1989). Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting*, *5*, 559–583.

Constantinidou, C., Nikolopoulos, K., Bougioukos, N., Tsiafa, E., Petropoulos, F., & Assimakopoulos, V. (2012). A neural network approach for the theta model. In *Lecture notes in information technology*: *Vol. 25. Information engineering* (pp. 116–120).

Franses, P. H., & Legerstee, R. (2011). Combining SKU-level sales forecasts from models and experts. *Expert Systems with Applications*, *38*, 2365–2370.

Gardner, J. E. S., & McKenzie, E. (1985). Forecasting trends in time series. *Management Science*, *31*, 1237–1246.

Goodwin, P., & Lawton, R. (1999). On the asymmetry of the symmetric mape. *International Journal of Forecasting*, *15*, 405–408.

Hyndman, R. J., & Billah, B. (2003). Unmasking the theta method. *International Journal of Forecasting*, *19*, 287–290.

Hyndman, R., & Khandakar, Y. (2008). Automatic time series forecasting: the forecast package for R. *Journal of Statistical Software*, *27*, 1–22.

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

Hyndman, R. J., Koehler, A. B., Snyder, R. D., & Grose, S. (2002). A state space framework for automatic forecasting using exponential smoothing methods. *International Journal of Forecasting*, *18*, 439–454.

Koning, A. J., Franses, P. H., Hibon, M., & Stekler, H. O. (2005). The M3 competition: Statistical tests of the results. *International Journal of Forecasting*, *21*(3), 397–409.

Makridakis, S., & Hibon, M. (2000). The M3-competition: results, conclusions and implications. *International Journal of Forecasting*, *16*, 451–476.

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

Nikolopoulos, K., & Assimakopoulos, V. (2005). Fathoming the theta model. In *25th International symposium on forecasting, ISF, San Antonio, Texas, USA*. unknown.

Nikolopoulos, K., Assimakopoulos, V., Bougioukos, N., Litsa, A., & Petropoulos, F. (2011). The Theta model: An essential forecasting tool for supply chain planning. In G. Lee (Ed.), *Lecture notes in electrical engineering*: *Vol. 123. Advances in automation and robotics, Vol. 2* (pp. 431–437). Berlin, Heidelberg: Spinger-Verlag.

Nikolopoulos, K., Thomakos, D., Petropoulos, F., Litsa, A., & Assimakopoulos, V. (2012). Forecasting S&P 500 with the theta model. *International Journal of Financial Economics and Econometrics*, *4*, 73–78.

Petropoulos, F., Makridakis, S., Assimakopoulos, V., & Nikolopoulos, K. (2014). 'horses for courses' in demand forecasting. *European Journal of Operational Research*, *237*(1), 152–163.

Petropoulos, F., & Nikolopoulos, K. (2013). Optimizing Theta model for monthly data. In *Proceedings of the 5th International conference on agents and artificial intelligence*.

Poler, R., & Mula, J. (2011). Forecasting model selection through out-of-sample rolling horizon weighted errors. *Expert Systems with Applications*, *38*, 14778–14785.

R Core Team (2015). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing, ISBN: 3-900051-07-0, URL: http://www.R-project.org/.

Thomakos, D., & Nikolopoulos, K. (2014). Fathoming the theta method for a unit root process. *IMA Journal of Management Mathematics*, *25*, 105–124.

Thomakos, D., & Nikolopoulos, K. (2015). Forecasting multivariate time series with the theta method. *Journal of Forecasting*, *34*, 220–229.

**Jose A. Fiorucci** is a Ph.D. student in Statistics at Federal University of São Carlos, Brazil. He holds Bachelor's and Master's degree in Computational Mathematics from University of São Paulo. His current research interests focus on automatic time series forecasting.

**Tiago R. Pellegrini** obtained his first degree in Statistics from Federal University of São Carlos (UFSCar) in 2010, M.Sc. in Statistics from UFSCar in 2012 and is a Ph.D. student in the Department of Mathematics and Statistics at University of New Brunswick. Current research interests are related to spatial models, time series analysis and computational statistics.

**Francisco Louzada** is a Full Professor in the Institute of Mathematical Science and Computing, University of São Paulo (USP), Brazil; director for the Center for Risk Analysis, USP/UFSCar, Brazil; CNPq Productivity Researcher level 1B. He received his Ph.D. degree in Statistics from the University of Oxford, UK, his M.Sc. degree in Computational Mathematics from USP, and his B.Sc. degree from UFSCar, Brazil. His main interests are survival analysis, data mining and statistical inference. More information: http://www.mwstat.com/franciscolouzada.

**Fotios Petropoulos** is Assistant Professor at Cardiff Business School of Cardiff University and the Forecasting Support Systems Editor of Foresight. His research expertise lies in behavioural aspects of forecasting and improving the forecasting process, applied in the context of business and supply chain. He is the co-founder of the Forecasting Society (www.forsoc.net).

**Anne B. Koehler** is an Professor Emerita of Decision Sciences at Miami University, Oxford, Ohio. She has numerous publications, many of which are on forecasting models for seasonal time series and exponential smoothing methods. She is a fellow of the International Institute of Forecasters and the Decision Sciences Institute.