



PRIFYSGOL  
**BANGOR**  
UNIVERSITY

## Forecasting with a hybrid method utilizing data smoothing, a variation of the Theta method and shrinkage of seasonal factors

Spiliotis, Evangelos; Assimakopoulos, Vassilios; Nikolopoulos, Konstantinos

### International Journal of Production Economics

DOI:  
[10.1016/j.ijpe.2018.01.020](https://doi.org/10.1016/j.ijpe.2018.01.020)

Published: 01/03/2019

Peer reviewed version

[Cyswllt i'r cyhoeddiad / Link to publication](#)

*Dyfyniad o'r fersiwn a gyhoeddwyd / Citation for published version (APA):*  
Spiliotis, E., Assimakopoulos, V., & Nikolopoulos, K. (2019). Forecasting with a hybrid method utilizing data smoothing, a variation of the Theta method and shrinkage of seasonal factors. *International Journal of Production Economics*, 209, 92-102.  
<https://doi.org/10.1016/j.ijpe.2018.01.020>

#### Hawliau Cyffredinol / General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

#### Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

# **Forecasting with a hybrid method utilizing data smoothing, a variation of the Theta method and shrinkage of seasonal factors**

## **Abstract**

In this paper, we discuss how extrapolation can be advanced by using some of the most successful elements and paradigms from the forecasting literature. We propose a new hybrid method that utilises: a) the decomposition approach of the Theta method, but instead of considering a linear trend we allow for nonlinear trends, b) rather than employing the extrapolation method on the raw data, we first apply smoothing to the data, and c) when seasonality is present, we employ the shrinkage approach to the derived indices instead of simply applying classical seasonal decomposition. We empirically evaluate the new proposition in the M3-Competition data with very promising results in terms of forecast accuracy.

## **1. Introduction**

Research in time series extrapolation methods is quite fragmented to the extent that any improvement in forecasting performance is very incremental and most benefits in the process either come from better information sharing or management judgment (*Goodwin et al., 2013*). In order for a newly proposed forecasting approach to have any credibility and claim that it can stand the test of time, it should compete favorably and beat the two best established benchmarks in the field: Damped Trend Exponential smoothing (*Gardner and McKenzie, 1985*) and the Theta method (*Assimakopoulos and Nikolopoulos, 2000*).

In this paper, we discuss how extrapolation could be advanced by using some of the most successful elements and paradigms from the past, and we propose such a new hybrid method. First, we exploit the decomposition approach of the Theta method, but instead of considering a linear trend we allow for more flexibility and experiment with a set of nonlinear trends. Furthermore, instead of employing the method on the raw data, we first apply smoothing to the data to obtain a much better starting point for our extrapolation. This has proven to be very beneficial for the Naive method in many empirical investigations (*Makridakis et al., 1983*). Finally, instead of just applying classical seasonal decomposition when seasonality is present, we employ the well celebrated shrinkage approach of *Miller and Williams (2003)* to the derived indices (from the aforementioned classical approach).

The rest of the paper is organized as follows: in the next section, we discuss the literature on the Theta method so far. In *Section 3* we describe the original Theta model and present some fertile areas for improvement. In *Section 4* we present and analyze the first two elements of the proposed method, demonstrating its possible advantages. *Section 5* presents the hybrid method's forecasting performance as a whole using the 3003 series of the M3-Competition. Finally, *Section 6* provides our conclusions and some thoughts on future research.

## **2. The Theta model: a literature review**

The Theta model, proposed by *Assimakopoulos and Nikolopoulos (2000)*, is a univariate forecasting method based on modifying the local curvature of the time series through a coefficient "Theta" ( $\theta$ ) applied to the second differences of the data. The modification leads to the creation of new lines which maintain the mean and the slope of the original data, but

not their curvature. The smaller the value of the Theta coefficient, the larger the degree of curve deflation, and vice-versa. Thus, fluctuated lines with  $0 < \theta < 1$  can be used for identifying long-term trends (*Assimakopoulos, 1995*), while highly curved ones with  $\theta > 1$  for emphasizing the short-term characteristics of the series, such as the running level. In practice,  $\theta$  can be considered as a transformation adjusting the curvatures of the series according to the distance of its points with the ones of a simple linear regression in time, obtained for  $\theta = 0$ . In this regard, two or more Theta lines can be created, extrapolated and combined accordingly to mime the short- and long-term behavior of the series.

In its original form, as applied to the monthly M3 data, the Theta model consists of two Theta lines with  $\theta$  values of 0 and 2 calculated on the seasonally adjusted data. Theta line (0) has zero curvature and equals to a simple linear regression line. On the other hand, Theta line (2) represents a line with double curvature of the original series. The first line is forecasted by extrapolating the regression line, while the second one using Simple Exponential Smoothing (SES) (*Gardner, 1985*). The forecasts are combined using equal weights and then reseasonalized. This form of Theta participated in the M3-Competition (*Makridakis and Hibon, 2000*) and became popular for outperforming the rest of its competitors, particularly for monthly series and microeconomic data. It is notable that the model, despite being quite simplistic, performed far better than the participating advanced methods and expert system, such as ForecastPro and ForecastX. Till today, it remains a difficult benchmark to beat.

Since the first appearance of the model, and considering its notable performance in the M-3 competition, a lot of work has been done both in the direction of implementing it in Forecasting Support Systems (FSSs), and testing its accuracy on different data sets.

*Nikolopoulos and Assimakopoulos (2003)* developed a system integrating Theta model forecasts with automated rule based and judgmental adjustments for supporting decisions. *Tavanidou et al. (2003)* included it among the forecasting techniques of a web-based FSS. Latter, *Pagourtzi et al. (2007)* evaluated the Theta model for forecasting quarterly housing prices and the total average dwelling prices in the UK. *Petropoulos et al. (2008)* also suggested some ideas for its implementation in Intermittent Demand forecasting (*Spithourakis et al., 2014*). Moreover, the model was assessed on a large dataset of non-demand forecasting series by forecasting the evolution of the S&P500 index (*Nikolopoulos et al., 2012a*) and indicating possible gains in supply chain management and planning (*Nikolopoulos et al., 2012b*). More recently, *Thomakos and Nikolopoulos (2015)* proposed the extension of the univariate Theta model for forecasting multivariate time series and assessed its performance in real macroeconomic and financial time series. The results of all these studies were very promising as the performance of Theta was either on par, or better than the benchmarks set.

The research on the Theta model has advanced further in the direction of optimizing its parameters and generalizing its use. *Constantinidou et al. (2012)* suggested a neural network approach for optimizing the combination weights of two Theta lines. *Petropoulos and Nikolopoulos (2013)* included multiple Theta lines to extract more information from the provided data and further boost the performance of the model. *Thomakos and Nikolopoulos (2014)* introduced an approach for selecting the optimal value of the Theta coefficient when forecasting with a single Theta line and provided a formula for defining the optimal weights when combining two lines. Finally, *Fioruci et al. (2016)* proposed a method for optimally

selecting the second Theta line based on validation schemes, when the first line is calculated for  $\theta = 0$ .

As the literature indicates, although the Theta method is based on a generic decomposition framework, the research is mainly focused on optimizing the weights of the combined Theta lines. In the usual case of combining one Theta line (0) with another, this process is equal to identifying the most appropriate value of the Theta coefficient, if the original data is to be reconstructed from the individual lines. Thus, the decomposition process is limited to combining a straight line. This is useful for identifying the long trend, with the “best” curved one being effective in identifying the short-term characteristics of the series. A question arising at this point is what happens if the trend of the model is not linear. As *Hyndman and Billah (2003)* indicated, the Theta model can be expressed as simple exponential smoothing with a drift relative to the slope of the linear trend fitted to the data. Given that a time series is very likely to follow a nonlinear pattern, it becomes crucial to further expand Theta to nonlinear trends, especially when referring to mid- or long-term forecasts.

In this paper we propose the replacement of the original Theta line (0) with simple nonlinear lines (e.g. exponential or logarithmic curves) and the construction of a second one, so that the original time series is reconstructed from their combination. Given the increased flexibility of the Theta line (0) and its improved fit, the second line will be curved only at the points which diverge from the trend pattern. Thus, it will be much more stable and effective in modeling level variations. In total, the final forecasts of the model will be the improved running level of the series, adjusted by the defined slope. An empirical investigation, performed on the M3-Competition data set, indicates improvements on forecasting accuracy

for the proposed approach. Forecasting performance can be further enhanced if smoothing is applied to the original data before extrapolation.

### 3. Insights on the Theta method

The Theta model is based on a decomposition process proposed by *Assimakopoulos and Nikolopoulos (2000)* for adjusting the curvatures of the time series. The adjustment is achieved by applying the coefficient Theta ( $\theta \in R$ ) to the second differences of the data leading to the creation of a Theta line  $Y^\theta$ , as shown in the following equation:

$$Y_t^\theta = \theta Y_t'', \text{ with } Y_t'' = Y_t - 2Y_{t-1} + Y_{t+2} \quad (1)$$

where  $Y_t$  is the original time series at time  $t$ . If the value of Theta is greater than one, then the process creates a line with stronger curvatures than the original series, obtained for  $\theta = 1$ . On the other hand, for  $0 < \theta < 1$  the created lines are fluctuated leading to a completely straight line for  $\theta = 0$ , which is equal to the simple linear regression in time. Given the simplified modifying formula of Theta provided by *Nikolopoulos et al. (2012b)*, every line created will be proportionally adjusted based on the distance between  $Y^0$  and  $Y^1$ . Thus, the new lines maintain the mean and the slope of the original data, but with adjusted curvatures as follows:

$$Y_t^\theta = \theta Y_t + (1 - \theta) Y_t^0 = \theta Y_t + (1 - \theta)(b + at) \quad (2)$$

where  $b$  and  $a$  the intercept and slope of the simple linear regression in time. Observations with a long distance from  $Y^0$  will face a strong modification, and vice versa. The Theta coefficient can also obtain negative values with similar effects. However, since the

created lines will be symmetric to the original series with respect to  $Y^0$ , they are of no interest in this study and are not discussed.

After the creation of the individual Theta lines, a forecasting method is used to extrapolate them. One, two or more lines can be considered. The forecasts are then combined using appropriate weights. Based on the previous equation, Theta lines with coefficients greater than 1 can be used for capturing short-term components of the series, while lines with coefficients close to zero for estimating long-term characteristics. At this point we also note that, in order the individual lines to reconstruct the original time series, specific limitations must be set regarding their weights. As noted by *Fioruci et al. (2016)*, for the case of two Theta lines the weights can be directly calculated as follows:

$$w_{\theta_1} = \frac{\theta_2 - 1}{\theta_2 - \theta_1} \text{ and } w_{\theta_2} = 1 - w_{\theta_1} \text{ with the limitations of } \theta_1 \leq 1, \theta_2 \geq 1 \quad (3)$$

where  $\theta_i$  and  $w_{\theta_i}$  are the coefficient of Theta and the weight chosen for line  $i$ , respectively.

Equation (3) can be further simplified when the first of the two lines is the  $Y^0$  as follows:

$$w_0 = \frac{\theta - 1}{\theta} \text{ and } w_{\theta} = \frac{1}{\theta} \quad (4)$$

Given its lower complexity, this form of the model is also the most used. Moreover, empirical results suggest that  $Y^0$  must be included in the model to obtain robust forecasts and capture the long trend of the data (*Assimakopoulos, 1995*).

However, models of multiple Theta lines can also be considered under relative limitations. Following the above mentioned suggestions, the generic form of the Theta model is given as follows:

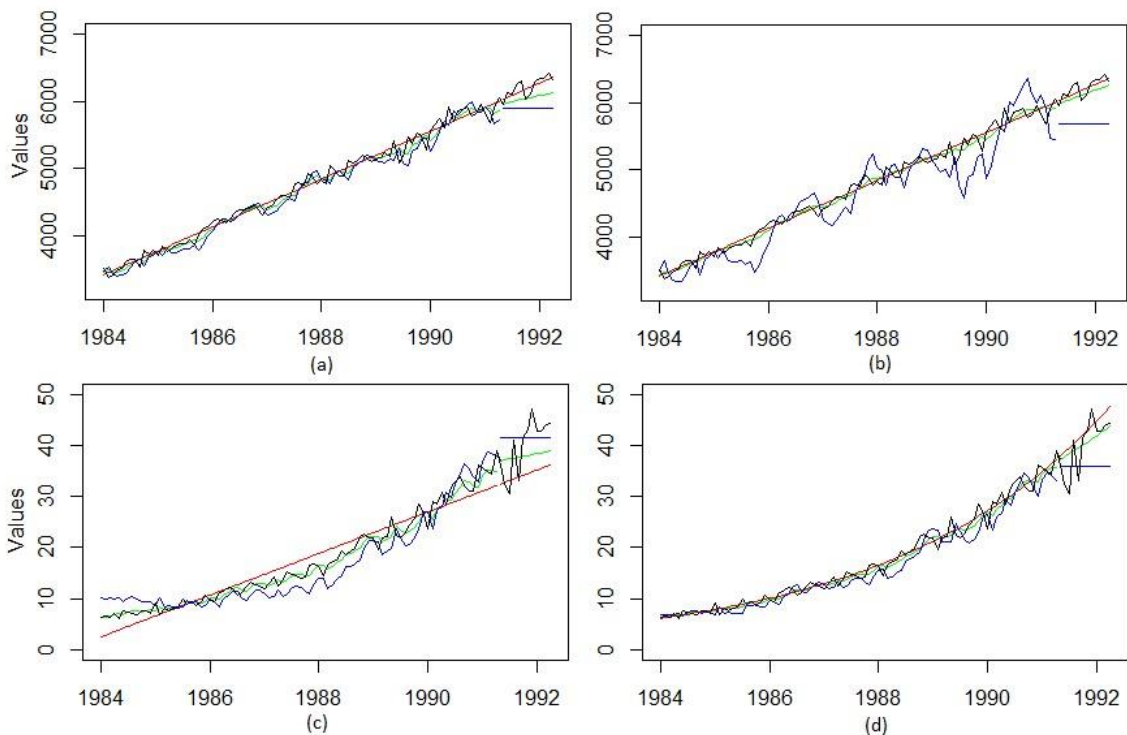


$$Y_t = \frac{\theta - 1}{\theta} Y_t^0 + \frac{1}{\theta} Y_t^\theta = \frac{\theta - 1}{\theta} (b + at) + \frac{1}{\theta} Y_t^\theta, \theta \geq 1 \quad (5)$$

The original form of the model consists of two Theta lines with Theta coefficients of 0 and 2 calculated for the seasonally adjusted data. More specifically, a seasonality test is first used to test the seasonal behavior of the series using an auto-correlation function. The data are considered seasonal if a 90% confidence is reported. In case of seasonal time series, the classical multiplicative decomposition by moving averages (*Makridakis et al., 1983*) is used for estimating the seasonal component. Then, the seasonally adjusted series is created by dividing the original one with the calculated seasonal factors and the  $Y^0$  and  $Y^2$  lines are calculated accordingly. Linear regression line and SES are used for extrapolating  $Y^0$  and  $Y^2$  respectively, and equal weights are selected for combining the individual forecasts, as suggested earlier. Finally, re-seasonalization is applied for the seasonally adjusted series.

Although this approach, has proven to be very effective (*Makridakis and Hibon, 2000*), it still has some drawbacks. *Hyndman and Billah (2003)* have proven that classic Theta is equal to SES with a drift equal to the half of the slope of the linear trend. This remark can be expanded for different coefficients of Theta when the first line is the  $Y^0$ . For instance, by increasing the coefficient of Theta, one can further drift the forecasts derived from SES, and vice versa. Taking this into consideration, two key issues arise: (1) identifying the trend type of the data and (2) selecting an appropriate  $\theta$  for optimally adjusting its intensity. The second point has been the subject of many studies and different formulas and methodologies have been proposed for identifying the optimal Theta coefficient. Yet, the first one remains unnoticed.

To better illustrate this problem, a characteristic example is presented in *Figure 1*. Consider two different time series, one with a linear and one with an exponential pattern of trend. 88 observations are used for training the model and 12 for evaluating its forecasting performance. Using the classic form of Theta, the linear series is adequately modeled (*a*) and higher coefficients of Theta can be used to further stress trend intensity and improve forecasting accuracy (*b*). However, in the second case, the exponential pattern of the data can barely be captured by classic Theta, regardless of the value of  $\theta$  used (*c*). By considering an exponential curve as  $\gamma^0$ , the results can be significantly improved (*d*). In this respect, in the next section we suggest and discuss the expansion of classic Theta into a flexible decomposition method considering both linear and nonlinear patterns of trend.



*Figure 1. Performance of the Theta method for linear and nonlinear time series: (a) and (c) display the forecasts of classic Theta for linear and exponential trended series, (b) Theta*

*classic for  $\theta = 7$  and (d) Theta model for  $Y^0 = be^{at}$ . The red line represents  $Y^0$ , the blue one  $Y^\theta$  and the green one their combination (Theta model)*

#### **4. Smoothing and forecasting with a variation of the Theta method tailored for nonlinear time series**

##### **4.1. Introducing the nonlinear trend component**

As discussed in *Section 3*, Theta model can capture linear trend and adjust slope intensity through the Theta coefficient. In this regard, high coefficient values can be chosen for significantly trended time series, and lower ones for stable or noisy data to ensure robustness. However poor results are expected in cases when nonlinear trends are identified.

Given its simplified expression (equations 2 and 5), the two-lined Theta model has the advantage of reconstructing the original series from the individual Theta lines, even if the line is used as  $Y^0$ . Since originally  $Y^\theta$  is forecasted with SES,  $\frac{1}{\theta}Y^\theta$  will mainly specify the level of the forecasts while  $\frac{\theta-1}{\theta}Y^0$  will drift them to capture the trend of the data. Thus, by replacing linear regression as  $Y^0$  with another curve, it is possible to automatically specify which type of trend will be considered.

In this paper, four additional curves are examined: exponential, logarithmic, inverse and power. Undoubtedly, any other type of curve, such as a polynomial one could be considered (keep in mind that linear regression is a polynomial curve of the first degree). Our choices were made given that the performance of the proposed approach must be directly comparable with that of Theta classic, i.e., simple curves of the same complexity (number of

estimated parameters  $a$  and  $b$ ) must be examined. Moreover, the trend obtained from more complex and flexible curves is difficult to be qualitatively specified, while over fitting can also become a great issue. Finally, the five curves provided in total manage to simulate almost any basic pattern of trend.

The formulas for obtaining the above-mentioned curves  $Y_t^0$  are given as follows:

Linear regression: 
$$Y_t^0 = b + at \tag{6.1}$$

Exponential curve: 
$$Y_t^0 = be^{at}, \text{ or } \log(Y_t^0) = \log(b) + at \tag{6.2}$$

Logarithmic curve: 
$$Y_t^0 = b + a \log(t) \tag{6.3}$$

Inverse curve: 
$$Y_t^0 = b + a \frac{1}{t} \tag{6.4}$$

Power curve: 
$$Y_t^0 = bt^a, \text{ or } \log(Y_t^0) = \log(b) + a \log(t) \tag{6.5}$$

where parameters  $a$  and  $b$  can be easily estimated by applying the least squares method for minimizing the sum of the squared errors (SSE) produced by the linear form of the equations, with the estimate being computed at the same period.

At this point we note that the parameters could have also been computed for minimizing the one-period ahead SSE, as proposed by Fioruci et al. (2016), called the “Dynamic Optimized Theta Model”. The former approach is preferred as it is the one originally proposed for the Theta method and we are interested in maintaining its classic properties. Moreover, the aim of this study is to expand the Theta method by introducing more types of trend and not by optimizing the parameterization process itself. Last but not least, the latter approach would

force the model to emphasize the most recent trend of the data (exponential smoothing logic) and not the long-term one (Theta logic). This is undesirable since Theta has been proven to perform well on long-term forecasts, exactly because of this feature. Yet, this is indeed a fertile area for future research and should be considered as a powerful alternative.

In Figure 2 the suggested curves are used for modeling the trend pattern of two different time series. As seen, the proposed approach is likely to lead to significant improvements since, in cases like the one presented on the right diagram, linear line produces unreasonably optimistic forecasts. As reported, this phenomenon could also be mitigated by selecting a better  $\theta$  coefficient for the classic double-lined Theta model (Thomakos and Nikolopoulos, 2014). Yet, it becomes clear that our suggestion offers a far more flexible solution and optimization of  $\theta$  can be supplementary used after selecting  $Y_t^0$ . This secondary problem, partially examined in the literature, is part of our future research and is not further examined here due to space limitations.

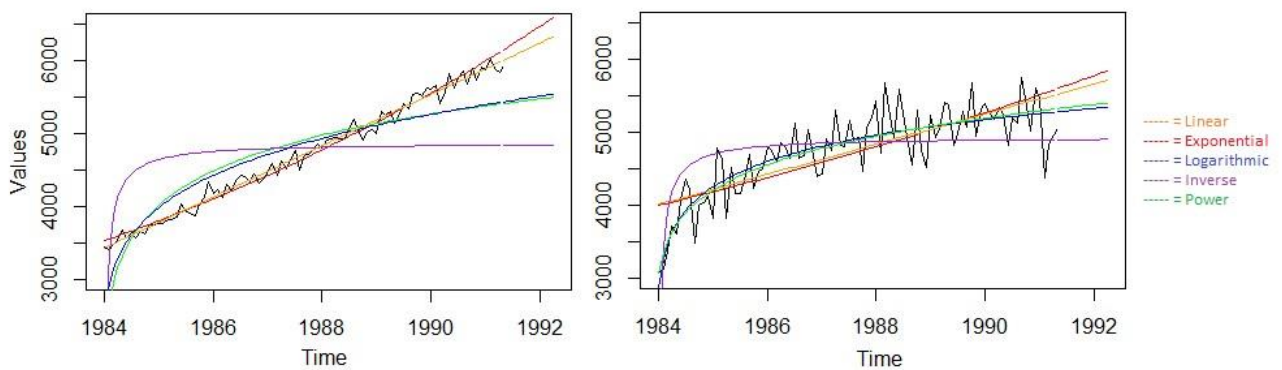


Figure 2. Fitting various  $Y_t^0$  into series with different trend patterns. According to their fit, linear line seems to be the most appropriate choice for the left series, while for the right one the power and logarithmic curve.

## 4.2. Simplifying the nonlinear form of Theta method using smoothed time series

As demonstrated in the previous sections, the good performance of the Theta model is basically due to its ability to capture both the long- (trend) and short-term (level) components of the series. Researches also come into a general agreement that these components are easier to model for a smoothed time series. For instance, *Kourentzes et al. (2014)* argue that temporal aggregated series of low frequencies can be used for identifying long-term characteristics of the data, while *Proietti and Lütkepohl (2013)* suggest the use of power transformations for rescaling the historical data, therefore simplifying their patterns and making them more consistent across the whole data set. Similar results can also be obtained through outlier (extreme values, level shifts etc.) detection techniques which eliminate the carry-over effect of the abnormalities on the forecast and the bias in the estimates of the model parameters (*Ledolter, 1989*).

Although all the approaches listed above for strengthening the valuable components of the series seem quite promising, in our case they would significantly increase the complexity of the forecasting process. This is undesirable given that simplicity is one of the strongest characteristics of the method. For instance, temporal aggregation is a time intensive process, while handling of outliers requires a lot of parameterization for appropriately detecting and removing possible abnormalities. Finally, transformations are only applicable to positive observations. In this respect, a smoothing process is considered as the most effective alternative for eliminating the variance of the data and highlighting its characteristics.

Extrapolating the smoothed time series instead of the original one offers an additional advantage to the method: the decrease of its complexity in terms of estimated parameters. Originally, SES is used for forecasting  $Y^{\theta}$  and linear regression line for extrapolating  $Y^0$ .

Since in the present study  $Y^0$  is modeled through two parameters  $(a, b)$ , the state values of SES increase the complexity of the method into four (smoothing parameter and initial level). However, given a completed smoothed time series (straight line), the coefficient of SES will be always forced to one. This is because, if no level variances are observable by the model, the last observation available will also be the best forecast for the period to come. This process is equal to the Naive method which has a complexity of zero. Therefore, after implementing a smoothing process, the total complexity of the Theta model can be further decreased into two. The last observation of the smoothed time series will specify an improved running level for the data, while the slope  $a$  its trend.

Using the simplified form of the Theta model (equation 5), and by replacing the forecasts of SES for  $Y^\theta$  with the ones of the Naive method (last known observation), the forecasts of the generic Theta model are given by the following equation:

$$\begin{aligned}
 Y_{t+n} &= \frac{\theta - 1}{\theta} Y_{t+n}^0 + \frac{1}{\theta} Y_{t+n}^\theta = \frac{1}{\theta} [(\theta - 1)Y_{t+n}^0 + [\theta Y_t + (1 - \theta)Y_t^0]] \quad (7) \\
 &= Y_t + \left(1 - \frac{1}{\theta}\right) (Y_{t+n}^0 - Y_t^0), \theta \geq 1
 \end{aligned}$$

where  $t$  is the forecast origin and  $Y_{t+n}$  is the forecast produced by the model  $n$  periods after the origin. The model can then be expressed for each of the trend types considered earlier by simply replacing  $Y_t^0$  and  $Y_{t+n}^0$  of equation 7 according to equations 6.1-6.5 as follows:

Linear regression: 
$$Y_{t+n} = Y_t + \left(1 - \frac{1}{\theta}\right) an \quad (8.1)$$

Exponential curve: 
$$Y_{t+n} = Y_t + \left(1 - \frac{1}{\theta}\right) b e^{at} (e^{an} - 1) \quad (8.2)$$

Logarithmic curve: 
$$Y_{t+n} = Y_t + \left(1 - \frac{1}{\theta}\right) a \log\left(\frac{t+n}{t}\right) \quad (8.3)$$

Inverse curve: 
$$Y_{t+n} = Y_t - \left(1 - \frac{1}{\theta}\right) a \frac{n}{t(t+n)} \quad (8.4)$$

Power curve: 
$$Y_{t+n} = Y_t + \left(1 - \frac{1}{\theta}\right) b [(t+n)^a - t^a] \quad (8.5)$$

As shown, if the series is completely smoothed (so there is no need of applying SES to  $Y_t^\theta$  to define the level of the series) the classic model is equivalent to a Naïve forecast with drift (equation 8.1), as proposed by *Hyndman and Billah (2003)*. However, this is not true when a nonlinear type of trend is examined as the data cannot be drifted in a constant way like before (add to the level  $n$  times the slope of the series). Therefore, the Theta decomposition framework is exploited to determine both the changeable slope at each data point and the way the component of trend is combined with that of level (see eq. 8.2-8.5). Moreover, for cases of non-smoothed data (where SES is originally used instead of Naïve for extrapolating  $Y_t^\theta$ ), the Theta method is obligatory for defying the running level of the data by determining the state values of SES (equation 5). Thus, we conclude that Theta framework is mandatory for effectively applying the proposed hybrid approach and that the process described cannot be substituted by simply drifting a level forecast.

#### 4.3. Proposing a technique for effectively smoothing time series

In the present study, a nonlinear mechanism is adopted for effectively smoothing time series. The mechanism is based on the theta transformation introduced by *Assimakopoulos*



(1995) for identifying long-term trends. The main idea of the approach is that the long-term trend of a time series can be estimated by aggregating the individual local trends, defined every three observations of the original series. In this respect, if the linear local trends calculated across the series are the same, the original series will be a completely straight line with no variances and a constant slope. On the other hand, if the individual slopes differ across the series, its long-term trend will be their average and strongly dependent on the variances observed. In order to better estimate the slope of the time series and mitigate the effect of local variances and outliers, a transformation for shrinking the local curvatures of the series and smoothing them can be applied.

In this regard, the second differences of the data are calculated for every observation and strong local variances are detected using the following formula:

$$LV_t = 300 \frac{|Y_{t-1} - 2Y_t + Y_{t+1}|}{Y_{t-1} + Y_t + Y_{t+1}} \quad (9)$$

where  $LV_t$  indicates the local variance at point  $t$ , as a percentage difference of its two neighbor observations.

After estimating the local variance for all the data points of the series, the strongest one is identified and limited based on the values of the neighboring observations using the formula given:

$$\hat{Y}_t = Y_t + s \left( \frac{Y_{t-1} + Y_{t+1}}{2} - Y_t \right) \quad (10)$$

where  $\hat{Y}_t$  is the new value of point  $t$ , leading to a smoother time series, and  $s \in [0,1]$  is the shrinkage parameter. If  $s$  is equal to zero, then no shrinkage is applied, and local variance remains the same. If  $s$  is equal to 1, then the new value of point  $t$  is on the straight line

connecting points  $t+1$  and  $t-1$  and local variance becomes equal to zero. In the rest of the cases the local variance is shrunk to some extent between its original value and zero.

This process is repeated by setting as an input the new time series created after applying a transformation, until no significant local variances occur. In the present study,  $s$  is set equal to 0.5 so that smoothing is performed effectively without dramatically changing the local curvatures of the series in each repetition. The process ends when none of the data points display a variance higher than 1%.

At this point we mention that, theoretically, in order for a time series to be considered as completely smoothed,  $LVs$  must all be equal to zero. The parameterization proposed, which loses the threshold to 1%, is just a suggestion provided to the reader for efficiently applying the smoothing process. This recommendation is based on empirical studies which conclude that stricter alternatives lead to equivalent results but exponentially increase computation time. In fact, having used multiple ad hoc optimization options, we find that changing the value of  $s$  results to insignificant differences in forecasting accuracy. Moreover, we learn that any limit lower than 1% leads to minor improvements. Yet, optimally parameterizing the smoothing process proposed is an interesting topic for future research. It is also mentioned that, if the threshold is higher than the  $LVs$ , it is possible for the smoothing parameter of SES to deviate from 1. In such cases, equations 7 and 8.1-8.5, which exploit this assumption, will approximately hold. However, this detail does not make any difference in practice as the level variance of such series is negligible compared to their mean. Thus, the forecasts generated are identical in both cases (Naïve and SES with  $a \neq 1$ ). We note that such cases are rare and were not observed in the present case study.

A limitation of the suggested approach in its original form, is that smoothing is only possible for data points 2 to  $n-1$ , where  $n$  is the length of the series: Calculating  $LV$  at points 1 and  $n$  requires two additional observations which, initially, are not available. Thus, the first and last observation of the series remain constant. Given that the rest of the observations are smoothed exploiting these unchangeable values, this may lead to poor forecasting performance in cases of them being outliers. For instance, the estimated trend and level of the smoothed series created may be misleading.

To overcome this problem, we propose the one-step-ahead extrapolation of the original series before smoothing using the Damped Trend Exponential Smoothing method (*Gardner and McKenzie, 1985*); both fore- and back-casting. Given the two data points created at the start and the end of the original time series, changes are now allowed for all of its observations making the whole process more flexible. Moreover, given that Damped is based on the running level and trend of the series, its forecasts are expected to be far more representative starting points for implementing the smoothing process than the original observations of the series. This leads to a potentially better forecasting origin for our extrapolation. After all, Damped is quite reasonably considered as a simple, yet accurate forecasting model which many more complex methods fail to compete with.

Once again, Damped can be replaced by any other forecasting model of our choice, as long as it effectively captures the running level of the series. For example, in our empirical evaluation we achieved similar results using Theta classic and other exponential smoothing models for extrapolating the series. However, Naïve and linear regression line lead to significantly worse results, mainly due to their inability to capture the trend and level of the series.

At this point we also note that fore- and back-casting is applied just once at the beginning of the smoothing process and not every time a new data point is transformed. Thus, the forecasting model used nor directly or indirectly influences the original trend pattern of the data.

Undoubtedly, similar results can also be obtained through other smoothing mechanism, such as moving averages and kernel smoothers. The main difference between our approach and other smoothing techniques is that a perfectly smoothed series is created (i) without introducing any missing data points and (ii) by utilizing a nonlinear and selective function. For example, when using a moving average of 12, six missing points are introduced at the beginning and the end of the series. To fill in this data, an appropriate technique must be applied. Given that this process determines the forecasting origin, it becomes clear that its performance strongly affects forecasting accuracy. Moreover, the more the order of the moving average is increased for obtaining a smoother time series, the more the missing data which are introduced, expanding the uncertainty of the whole process. In the suggested approach, no such issues arise. Moreover, moving averages homogeneously smooth the data, without emphasizing on outliers and local extrema. In contrast, the suggested approach is selectively applied based on the local variances of the points. Thus, the approach leads to a much smoother and forecastable series. A comparison between the two approaches is presented in *Figure 3*.

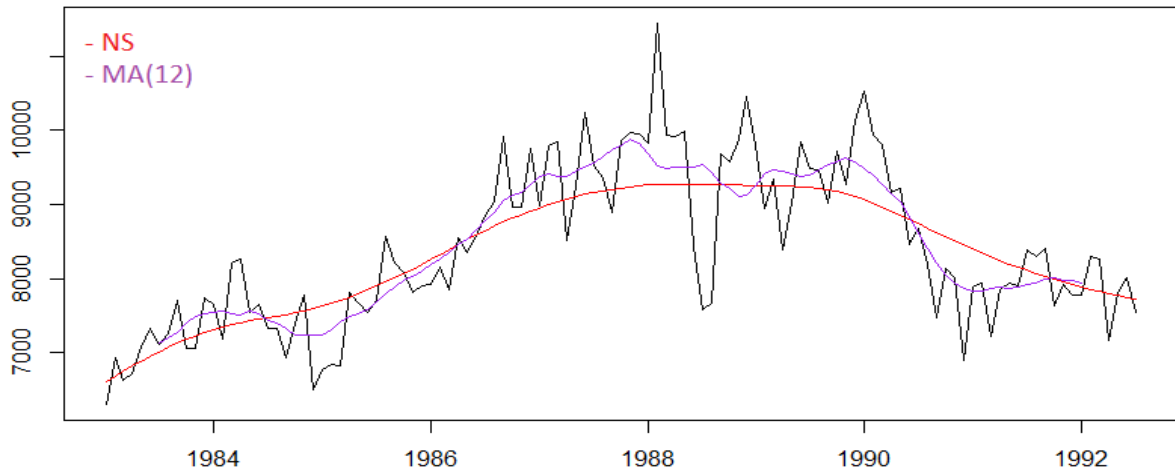


Figure 3. The smoothed time series obtained after applying the Nonlinear (NL) approach suggested and a Moving Average (MA) of order 12 to monthly data. As seen, the NL approach leads to a much smoother series without introducing missing data points.

## 5. Empirical evaluation and an improved seasonal adjustment

In order to evaluate the performance of our approach and provide some empirical evidence regarding its contribution, the data set of the M3-Competition is used as a case-study. The data set includes 3003 time series which can be distinguished in respect to their frequency (Yearly, Quarterly, Monthly, Other). The M3 remains as the largest forecasting competition ever conducted within which the original Theta model outperformed the rest of its competitors. For this reason we believe that it is the most appropriate choice for objectively assessing the forecasting accuracy of the nonlinear form of the model and obtaining some insights. Moreover, since the data are publicly available, they meet the conditions for reproducibility (Boylan, 2016).

The empirical evaluation is performed according to the principles of the original competition. A sample of historical data is provided for training the models (in-sample) and an additional number of previously unknown observations is used for evaluating their

performance (out-of-sample). The number of the out-of-sample observations depends on the frequency of the time series and is equal to the forecasting horizon set, that is 6, 8, 18 and 8 periods for Yearly, Quarterly, Monthly and Other series, respectively. The assessment in terms of forecasting accuracy is then performed using the symmetric Mean Absolute Error metric, defined as:

$$sMAPE = \frac{200}{n} \sum_{t=1}^n \frac{|Y_t - \hat{Y}_t|}{|Y_t| + |\hat{Y}_t|} \quad (11)$$

where  $n$  is the length of the forecasting horizon tested, and  $Y_t$  and  $\hat{Y}_t$  are the real observations and the forecasts produced by the model at point  $t$ , respectively.

The Mean Absolute Scaled Error (MASE) proposed by Hyndman and Koehler (2006) was also used for comparing the accuracy of forecasting methods such as sMAPE, which has been proven to penalize large positive errors more than negative ones and to increase the difference of their absolute size (Goodwin and Lawton, 1999). Yet, due to space limitations and similarity to the conclusions made, results are only displayed for the case of sMAPE. We just report that, according to MASE, smoothing seems to be more appropriate for the case of noisy data (e.g. monthly and quarterly data) leading to minor or no improvements when data are already smoothed (e.g. yearly and other data).

The results are provided in total as well as for each frequency individually. In detail, the data set used includes 645 Yearly, 756 Quarterly, 1428 Monthly and 174 Other time series. *Table 1* presents the performance of each model per frequency sub-set by applying five different trended models to the original time series of the competition, as described in *Subsection 3.2*. LRL refers to the linear form of the method, which is equivalent to the classic

Theta model. For reasons of comparison the original results of the Theta model, as provided at the Original Publication of the competition (*Makridakis and Hibon, 2003*), are also displayed (Theta OP). Accordingly, EXP, LOG, INV and POW refer to replacing linear regression with an EXponential, LOGarithmic, INVerse or POWer curve. An equal weighted COMbination of the five models is also considered to evaluate potential benefits of the method when multiple patterns of trend are considered simultaneously. Finally, two basic exponential smoothing models, SES and Holt are included in the results to provide evidence regarding whether combining existing good practices (smoothing and shrinked seasonal factors) is always beneficial and independent of the forecasting model used. In *Table 2* the same results are presented for the case of the smoothed time series, as described in *Subsection 3.3*.

To obtain more clear evidences, we estimate the percentage of the time series in which each model had the best performance across the rest of its kind, as well as its mean rank, from 1 to 6. The ranks are created for the original and smoothed series individually. We note that for reasons of simplicity and direct comparison, the Theta coefficient is set equal to 2 for all the models examined, although as reported earlier, one could possibly produce improved forecasts through its optimization (*Constantinidou et al., 2012*).

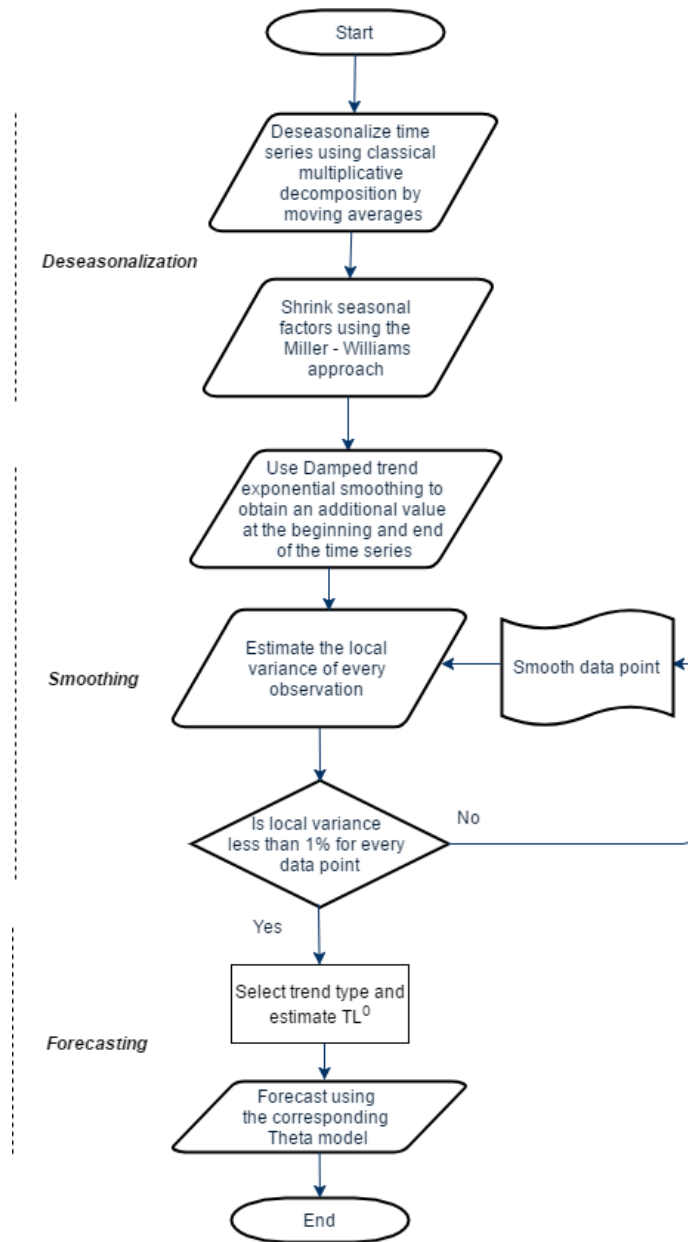
Finally, to further improve the performance of each model, the shrinkage estimators of time series seasonal factors of *Miller and Williams (2003)* are considered when seasonally adjusting the time series. This is done given that the literature strongly indicates the beneficial effect of shrinkage estimators on forecasting accuracy over classical decomposition. In brief, the seasonal factors of classical decomposition method are estimated for each time series and then the James-Stein (1961) or the Lemon-Krutchkoff

(1969) shrinkage estimators are applied based on the skewness of the seasonals and the value of the James-Stein shrinkage parameter. We also mention that, since classical multiplicative decomposition uses simple moving averages (e.g. MA(12) for monthly data) to estimate the seasonal indexes, no assumptions are made at this point regarding the trend pattern of the series, which is therefore exclusively determined through the Theta model chosen.

To summarize the methodological framework proposed and help the reader better understand its step, we visualize the whole process in the chart flow presented in *Figure 4*.

A summary of the results is presented in *Tables 3* and *4* for the original and the smoothed time series, respectively. As seen, in total, the exponential form of the method outperforms the rest of the approaches by improving the accuracy of classic Theta by 1.01% and 1.50% for the original and the smoothed series, respectively. For the case of the shrunked seasonal factors the benefits of considering an exponential trend are quite similar (0.97% and 1.47%). The data with an Other frequency are best forecasted by LRL, especially when the original data are used. However, for the rest of the frequencies, the exponential model is by far the most accurate one, especially among the yearly and monthly data with an average improvement of 2.08% and 1.49%, respectively.





*Figure 4. Chart flow of the proposed methodological framework including three individual procedures: (i) deseasonalization using shrunk seasonal factors, (ii) smoothing and (iii) forecasting through a generalized and simplified Theta model of defined trend type.*

It is also notable that, in general, smoothing seems to be far more effective for the case of quarterly and monthly data, leading to minor improvements or even decreasing the accuracy

of the models when dealing with yearly and other data. For instance, although the average performance of the models is increased by 3.03% and 1.88% for quarterly and monthly data, respectively, it is decreased by 0.87% and 2.51% for the case of yearly and other time series. This is an interesting conclusion mainly related with the noise and the fluctuations of the data which are more likely to be observed in higher frequencies. Thus, smoothing should be mainly used among noisy time series to help us better identify the signal of the data (level and trend), and be omitted when the time series is already smoothed.

Another encouraging finding is the fact that LRL and EXP are the best performing models in only 19% and 41% percent of the series. This means that, in 40% of cases the data are best modeled by another curve, such as the inverse one, which holds the best performance across 30% of the series. Moreover, the relative performance of the models is not affected when shranked seasonal estimators are applied. In this respect, it becomes quite clear that, although in total the rest of the models increase the forecast error, if a selective model was applied for defining the most appropriate trend type, significant gains would have been achieved.

*Table 2: Performance of the linear and nonlinear forms of Theta method according to sMAPE per time series frequency – Classical Decomposition applied.*

Trend Type	Yearly	Quarterly	Monthly	Other	Total	Mean Rank	1 <sup>st</sup> Rank (%)
Original Time Series							

Theta - OP	16.90	8.96	13.85	4.41	12.73	-	-
LRL	16.73	9.29	13.87	<b>4.92</b>	12.81	3.07	18.68
EXP	<b>16.44</b>	<b>9.28</b>	<b>13.68</b>	5.31	<b>12.68</b>	3.05	41.23
LOG	17.20	9.60	14.00	5.87	13.11	3.77	3.96
INV	17.73	9.81	14.25	6.27	13.42	4.25	28.37
POW	16.86	9.56	14.05	5.96	13.05	3.63	3.93
COM	16.46	9.40	13.82	5.62	12.80	3.23	3.83
SES	17.77	9.79	14.26	6.28	13.43	-	-
HOLT	19.15	11.23	15.82	4.67	14.73	-	-
Smoothed Time Series							
LRL	16.80	9.04	13.71	<b>5.07</b>	12.70	3.10	18.68
EXP	<b>16.39</b>	<b>8.98</b>	<b>13.48</b>	5.46	<b>12.51</b>	3.07	40.76
LOG	17.44	9.31	13.73	6.01	12.97	3.78	2.00
INV	17.97	9.51	13.88	6.40	13.23	4.25	31.10
POW	17.12	9.25	13.72	6.10	12.88	3.57	3.46
COM	16.70	9.12	13.61	5.77	12.69	3.23	4.00
SES	18.06	9.53	13.89	6.42	13.26	-	-
HOLT	17.59	11.32	17.01	4.86	15.00	-	-

Table 3: Performance of the linear and nonlinear forms of Theta method according to sMAPE per time series frequency – Shrinkage estimators applied.

Trend Type	Yearly	Quarterly	Monthly	Other	Total	Mean Rank	1 <sup>st</sup> Rank (%)
Original Time Series							
LRL	16.73	9.28	13.79	<b>4.92</b>	12.77	3.06	18.85
EXP	<b>16.44</b>	<b>9.27</b>	<b>13.62</b>	5.31	<b>12.65</b>	3.05	41.06
LOG	17.20	9.59	13.94	5.87	13.08	3.77	4.03
INV	17.73	9.81	14.20	6.27	13.39	4.25	28.31
POW	16.86	9.56	14.00	5.96	13.03	3.64	3.86
COM	16.46	9.39	13.76	5.62	12.77	3.23	3.90
SES	17.77	9.78	14.21	6.28	13.40	-	-
HOLT	19.15	11.23	15.74	4.67	14.69	-	-
Smoothed Time Series							
LRL	16.80	9.04	13.65	<b>5.07</b>	12.67	3.09	18.68
EXP	<b>16.39</b>	<b>8.98</b>	<b>13.43</b>	5.46	<b>12.48</b>	3.06	41.03
LOG	17.44	9.31	13.68	6.01	12.94	3.78	2.33
INV	17.97	9.50	13.84	6.40	13.20	4.25	30.87
POW	17.12	9.25	13.67	6.10	12.86	3.58	3.33

COM	16.70	9.11	13.56	5.77	12.66	3.24	3.76
SES	18.06	9.52	13.85	6.42	13.23	-	-
HOLT	17.59	11.30	17.02	4.86	15.00	-	-

Additionally, as the results indicate, smoothing has proven to be quite beneficial for all of the models considered, leading to an improvement from 0.88% to 1.43%, depending on the trend type considered (an average of 1.2%). This leads even some of the generally misperforming models to outperform the classic model of Theta, such as the logarithmic, the power and the inverse model for the monthly data. We also note that, although the accuracy is improved in total numbers when shrinkage of seasonals is considered, the benefits of smoothing the series remain the same. Thus, we conclude that the gains of applying shrinkage factors and smoothing the data are independent. Once again the ranks for the smoothed time series are very close with that of the original data. In this regard, identifying the type of trend remains a crucial issue and smoothing can be used beforehand to simplify the selection of  $Y_0$  and better estimate the short- and long-term components of the series.

Regarding the shrinkage of the seasonal factors, the benefits in forecasting accuracy seem to depend on the frequency of the data. The improvement is higher for the monthly time series (0.41%) than the quarterly data (0.06%) leading in total to an average improvement of 0.22% across the data set. The benefits also seem to be independent of the trend type considered and the appliance of the smoothing process.

We also mention that, although in the literature the combination of individual models usually leads to improved forecasting accuracy, in this study the combination of the five models seems to be an inadequate choice. This further enhances the point made earlier

regarding the importance of selecting the best model per series, since a simple combination of all of them leads to decreased forecasting performance overall. The poor performance of the combination is partially explained through the ranks of the individual models. For instance, LOG and POW are the best choice only in 6% of the data set and significantly worse than the rest of the models in terms of mean rank.

Another important finding of the study is that, although smoothing greatly improves the performance of SES (by 1.27%), it simultaneously decreases the forecasting accuracy of Holt exponential smoothing. This can be explained as follows: when we eliminate the noise of the data, on the one hand we end up with a more constant and representative running level for the time series considered – good for determining the forecasting origin-, and on the other hand we emphasize the local curvatures of the data (see for example *Figure 3*). Thus, if the forecasting model used considers among others the running trend of the data, estimated slope will be way too strong leading to significantly optimistic or pessimistic forecasts. For example, Holt on smoothed time series will work as a local regression model. However, this is not true for the case of the Theta method, as  $Y^0$  is estimated using the whole sample (long-term trend). Therefore, Theta is not influenced by the most recent curvatures of the series (short-term trend). Consequently, in contrast to trended exponential smoothing, the long-term trend pattern identified through the Theta decomposition framework, combined with the improved forecasting origin provided through smoothing, will lead to improved forecasting performance. Additionally, shrinkage of seasonal indexes seems to be model-independent, systematically improving the accuracy of all the forecasting models used.

To conclude, in case both shrunked seasonal factors and smoothed series are considered, an average improvement of 1.42% is reported across the models (from 1.11% to 1.62%

depending on the type of trend selected). The benefits of this approach are greater for the monthly and quarterly data, mainly due to the shrunk seasonal factors applied to these frequencies. The fact that smoothing eliminates the noise and defines a better forecasting origin, may also partially explain why the approach performs better for higher data frequencies. The best performance is achieved when the exponentially trended Theta model is applied to the smoothed time series using the shrunk seasonal factors instead of the ones produced by the classical decomposition method.

At this point it is worth mentioning that the hybrid method proposed can easily be extended for local models where data are read and used over a moving window. To do so, we used the first  $3*f$  observations of each time series to train the Theta models, with  $f$  being the frequency of the time-series. We repeated the process over a rolling window (shift one point at a time) until no more observations were available. The sMAPE was estimated across the windows for each time series according to the original experimental setup, and then averaged to evaluate the models.

The size of the training sample was selected given that, to estimate the seasonal factors, at least three periods of data are required. As seen in *Table 4*, the relative results are pretty much the same with the ones of the previous study for the case of the Quarterly and Monthly data, where a significant amount of data is available for training and testing the local models. Yet, the absolute error was increased over the series, mainly because fewer data were used to estimate long-term forecasts. Thus, we conclude that local models can benefit from the proposed methodology, as smoothing and damped seasonal factors tend to improve forecasting accuracy. Additionally, a simple combination of the individual models would be a robust choice for extrapolation.

Table 4: Performance of the linear and nonlinear forms of Theta method according to sMAPE per time series frequency when applied local y- Original approach vs. proposed one.

Trend Type	Quarterly	Monthly	Total	Quarterly	Monthly	Total
	Original Time Series			Smoothed Time Series and Seasonal Estimators Applied		
LRL	10.78	15.35	13.77	10.84	15.31	13.77
EXP	<b>10.67</b>	15.17	<b>13.61</b>	<b>10.71</b>	15.06	13.56
LOG	11.07	15.25	13.81	10.97	14.95	13.57
INV	11.29	15.32	13.93	11.15	15.00	13.67
POW	11.02	15.24	13.78	10.92	14.93	13.54
COM	10.81	<b>15.13</b>	13.64	10.81	<b>14.95</b>	<b>13.52</b>

## 6. Conclusions and future extensions

In this paper, we aspire to pave the way for researchers, when in quest of new hybrid methods, where they utilize and combine past successful approaches: the elements being, primarily the Theta method, but also good-old-traditional time series smoothing and shrinkage through the seasonal adjustment process.

First, we employ the decomposition approach of the Theta method, but instead of considering a linear trend we allow for more flexibility and experiment with a set of nonlinear trends. Then, instead of applying the method on the raw data, we apply smoothing first. Finally, instead of just applying classical seasonal decomposition, we employ the well celebrated shrinkage approach of *Miller and Williams (2003)* to the derived indices. The



results are very promising and we believe they should stimulate academics, practitioners and most importantly FSS software designers.

The Theta method has long been considered as a robust and accurate method for forecasting, especially in the cases of monthly and microeconomic time series. Since its first appearance in the literature, many suggestions have been made in the direction of generalizing its use either by involving more Theta lines or by optimizing the  $\theta$  coefficient, which is responsible for adjusting the linear slope used for drifting the forecasts. In the present paper, we suggest that to further generalize its use, apart from the  $\theta$  coefficient, nonlinear forms of trend should be considered. In this respect, we identify cases where nonlinear trend can be proven beneficial and provide a generic formula for modifying its drift accordingly. Five different types of trends are considered leading to an equal number of models with the same complexity of the original one. These can be used for adequately modeling any basic pattern of data.

The overall performance of our hybrid approach is validated using the data set of the M3-Competition taking into consideration the frequency of the time series. The results show that exponential trended model outperforms in total the classic form of the Theta method. However, for specific sub-sets the linear and other nonlinear forms of it are more accurate. This remark is very promising since it demonstrates that in practice there is no dominant model and that many types of trend should be considered when parameterizing its general form, based on the patterns of the data. This becomes more clear given that the exponential-trended model displays the top performance for 40% of cases, with the linear- and inverse-trended models following with 20% and 30%, respectively.

Since the Theta method is based on decomposing the data for highlighting their short- and long-term components, we also suggest that a smoothing process must be applied before forecasting. In this regard, the components of the series, and mainly the level of the data, are better captured. This technique will not only enhance the performance of the individual models, but also lead to less complex ones, by simplifying their parameterization. The results show that the forecasting accuracy of all the Theta models is boosted when smoothing is applied, improving their performance by more than 1.2%. The benefits are greater for high frequency data. This can be explained if we consider that noise and fluctuations are more likely to be observed on such type of time series. We also note that the relative performance of the models is maintained after smoothing, which means that smoothing should be used a-priori for improving the estimations of the models. Yet, selection of the most appropriate one remains an issue.

The ability of smoothing to enhance forecasting accuracy through the identification of a better level, is proven among others though the improved performance of SES. However, smoothing should not be applied in the case of forecasting methods which consider the running trend of the data (e.g. Holt exponential smoothing). This is because smoothing tends to emphasize local curvatures, leading to strong slope estimates and, therefore, to over-optimistic or pessimistic forecasts. In contrast, it becomes more than helpful for models considering the long-term trend of the data (e.g. Theta).

Another important finding of our study is that by using shrunk seasonal factors instead of the ones produced by the classical decomposition method, additional improvements can be achieved, especially for monthly time series. Moreover, the fact that the improvement is independent of the time series (smoothed time series or not) and the forecasting model used,

indicates that decomposing is a prerequisite when trying to improve the accuracy of a forecasting model in a simple, yet robust way.

Regarding the limitations of the proposed method, one short coming is that interest in predicting short term movement of the time series cannot be important since this movement has been smoothed out in the adjusted time series. Another short coming is that without a model, there is no way to compute prediction intervals with a sound statistical basis and approximations must be used instead. To overcome the issue of analytically deriving prediction intervals, a simulation process can be applied instead to iteratively generate multiple future sample paths. In this respect, the forecast distribution of the hybrid method is empirically, yet adequately defined. This is a common practice applied e.g. for the case of nonlinear autoregressive models, such as neural network models and other machine learning methods.

A very fertile area for research is the multivariate expansion of this hybrid method. For example, this could be done by exploiting the bivariate version of the Theta method (*Thomakos and Nikolopoulos, 2015*) but applied first on smoothed data rather than the original one: if the data are seasonal one can also experiment by damping the seasonal factors after the bivariate extrapolation of the deseasonalized time series.

In our point of view, through the suggested modification, the method can be transformed into a completely dynamic decomposition method offering numerous modeling opportunities. Taking into account the results of our study and the conclusions of previous relative works, we believe that selecting the most appropriate trend type through e.g. a validation technique, could be very beneficial. The advantages could be even greater when combined with methods optimizing the coefficient of Theta. This is part of our future

research. Moreover, we are very interested in further investigating the impact of smoothing ~~to~~ on the forecasting performance of the Theta method and other forecasting models, as well as identifying approaches for optimally selecting the parameters of the process. This could offer additional insights regarding the way the Theta method responds and reveal new opportunities for optimization. Finally, the hybrid approach could be expanded for local models where data are read and used over a moving window and further test its performance when working with few and most recent data.

### **Acknowledgements**

The authors would like to thank the reviewers for their useful for and against comments and recommendations which significantly helped us improve the quality of the paper further.

### **Appendix**

Notations for the parameters:

$Y_t$ :	The data of the time series at point t
$\widehat{Y}_t$ :	The forecast of the model at point t
$\acute{Y}_t$ :	The data of the smoothed series at point t
$\theta$ :	The coefficient of the Theta model
$Y_t^\theta$ :	The Theta line calculated for a coefficient of $\theta$ at point t
$Y''$ :	The second differences of a time series
$a$ :	The intercept of the deterministic model considered
$b$ :	The slope of the deterministic model considered
$w_\theta$ :	The weight chosen for Theta line $\theta$
$LV_t$ :	The local variance estimated for the time series at point t
$s$ :	The shrinkage parameter used for smoothing the data

### **References**

Assimakopoulos, V., 1995. A successive filtering technique for identifying long- term trends. Journal of Forecasting, John Wiley 14 (1), 35 – 43.

- Assimakopoulos, V., Nikolopoulos, K., 2000. The theta model: a decomposition approach to forecasting. *International Journal of Forecasting* 16 (4), 521 – 530.
- Boylan, J. E., 2016. Reproducibility. *IMA Journal of Management Mathematics* 27 (2), 1 107–108.
- Constantinidou C., Nikolopoulos K., Bougioukos N., Tsiafa E., Petropoulos F. and Assimakopoulos V., 2012. "A neural network approach for the Theta model", in: Ming M. (ed.), *Information Engineering, Lecture Notes in Information Technology*, 25, 116-120.
- Fioruci, J. A., Pellegrini, T. R., Louzada, F., Petropoulos, F., Koehler, A. B., 2016. Models for optimising the theta method and their relationship to state space models. *International Journal of Forecasting* 32 (4), 1151–1161.
- Gardner, E. S., Jr., 1985. Exponential smoothing: the state of the art. *Journal of Forecasting* 4 (1), 1 – 28.
- Gardner, E.S., Jr., & McKenzie, E., 1985. Forecasting trends in time series, *Management Science*, 31, 1237-1246.
- Goodwin, P., Dyussekeneva, K., Meeran, S. 2013. The use of analogies in forecasting the annual sales of new electronics products. *IMA Journal of Management Mathematics* 24(4), 407–422.
- Goodwin, P., Lawton, R., 1999. On the asymmetry of the symmetric MAPE. *International Journal of Forecasting* 15 (4), 405-408.
- Hyndman, R. J., Billah, B., 2003. Unmasking the theta method. *International Journal of Forecasting* 19 (2), 287 – 290.

- Hyndman, R. J., Koehler, A. B., 2006. Another look at measures of forecast accuracy. *International Journal of Forecasting* 22 (4), 679-688.
- James, W., & Stein, C., 1961. Estimation with quadratic loss. In *Proceedings of the fourth Berkeley symposium on mathematical statistics and probability*, vol. 1. Berkeley, CA: University of California Press, 361–379.
- Kourentzes, N., Petropoulos, F., Trapero, J. R., 2014. Improving forecasting by estimating time series structural components across multiple frequencies. *International Journal of Forecasting* 30 (2), 291 – 302.
- Ledolter, J., 1989. The effect of additive outliers on the forecasts from arima models. *International Journal of Forecasting* 5 (2), 231 – 240.
- Lemon, G. H., & Krutchkoff, R. G., 1969. Smooth empirical Bayes estimators: with results for the binomial and normal situations. In *Proceedings of the symposium on empirical Bayes estimation and computing in statistics*, Texas Tech University mathematics series no. 6, 110–140.
- Makridakis, S., Hibon, M., 2000. The M3-Competition: results, conclusions and implications. *International Journal of Forecasting* 16 (4), 451 – 476.
- Makridakis, S., Wheelwright, S. C., McGee, V. E., 1983. *Forecasting: Methods and applications* (Second Edition). New York: Wiley.
- Miller, D. M., Williams, D., 2003. Shrinkage estimators of time series seasonal factors and their effect on forecasting accuracy. *International Journal of Forecasting*, 19 (3), 669-684.

- Nikolopoulos, K., Assimakopoulos, V., 2003. Theta intelligent forecasting information system. *Industrial Management and Data Systems* 103 (9), 711 – 726.
- Nikolopoulos, K., Thomakos, D., Petropoulos, F., Litsa, A., Assimakopoulos, V., 2012a. Forecasting S&P500 with the Theta Model. *International Journal of Financial Economics and Econometrics*, 4 (1), 73 – 780.
- Nikolopoulos, K., Assimakopoulos, V., Bougioukos, N., Litsa, A., Petropoulos, F., 2012b. The Theta model: An essential forecasting tool for Supply Chain planning. *Lecture Notes in Electrical Engineering* 123 (1), 431 – 437.
- Pagourtzi, E., Assimakopoulos, V., Litsa, A., 2007. Theta Model forecasts quarterly & monthly dwelling prices in UK. *Briefings in Real Estate Finance* 17 (1), 1 – 30.
- Petropoulos, F., Nikolopoulos, K., 2013. Optimizing the theta model for monthly time series. *Proceedings of the 5th International Conference of Agents and Artificial Intelligence*.
- Petropoulos, F., Nikolopoulos, K., Assimakopoulos, V., 2008. Theta-Croston and Croston-Theta: investigating new ideas for Intermittent Demand forecasting. 28th International Symposium on Forecasting ISF 2008, June 22-25 2008, Nice, France.
- Proietti, T., Lütkepohl, H., 2013. Does the boxcox transformation help in forecasting macroeconomic time series? *International Journal of Forecasting* 29 (1), 88 – 99.
- Spithourakis, G. P., Petropoulos, F., Nikolopoulos, K., Assimakopoulos, V., 2014. A systemic view of the ADIDA framework. *IMA Journal of Management Mathematics* 25 (2), 125–137.

- Tavanidou, E., Nikolopoulos, K., Metaxiotis, K., Assimakopoulos, V., 2003. eTIFIS: An innovative eForecasting Web application. *International Journal of Software Engineering and Knowledge Engineering* 13 (2), 215 – 236.
- Thomakos, D., Nikolopoulos, K., 2014. Fathoming the theta method for a unit root process. *IMA Journal of Management Mathematics* 25 (1), 105 – 124.
- Thomakos, D., Nikolopoulos, K., 2015. Forecasting multivariate time series with the theta method. *Journal of Forecasting* 34, 220-229.