Method Description

General Information

Type of Entry (*Academic, Practitioner, Researcher, Student*)	**Academic**
First Name	**Evangelos**
Last Name	**Spiliotis**
Country	**Greece**
Type of Affiliation (*University, Company-Organization, Individual*)	**University**
Affiliation	**Forecasting & Strategy Unit – National Technical University of Athens**

Team Members (*if applicable*):

1st Member	
First Name	Vassilios
Last Name	Assimakopoulos
Country	Greece
Affiliation	Forecasting & Strategy Unit – National Technical University of Athens

Information about the method utilized

Name of Method	**4Theta**
Type of Method (*Statistical, Machine Learning, Combination, Other*)	**Statistical**
Short Description (up to 200 words)	**We examine modifications on the decomposition framework of Theta to boost its performance. This includes considering non-linear patterns of trend, adjusting trend intensity and introducing a multiplicative expression of the method. The extensions proposed transform Theta into a generalized forecasting algorithm for automatic extrapolation with enhanced flexibility and improved properties compared to its classical form.**

Extended Description:

4Theta: Generalizing the Theta method for automatic forecasting

Evangelos Spiliotis[a,*], Vassilios Assimakopoulos[a]

[a] *Forecasting and Strategy Unit, School of Electrical and Computer Engineering, National Technical University of Athens, 9 Iroon Polytechniou Str, 15773 Zografou Athens, Greece*

Abstract

4Theta examines some modifications on the decomposition framework of Theta to boost forecasting performance. This includes considering non-linear patterns of trend, adjusting trend intensity and introducing a multiplicative expression of the method. The extensions proposed transform Theta into a generalized forecasting algorithm for automatic extrapolation with enhanced flexibility and improved properties compared to its classical form.

1. The Theta method: Past, present and future

The Theta method, proposed by Assimakopoulos & Nikolopoulos (2000), is a univariate forecasting method which decomposes the original data into two or more lines, called the Theta lines, extrapolates them using appropriate forecasting models and then combines their predictions to obtain the final forecasts. The Theta lines derive by modifying the local curvature of the original time series through a coefficient θ applied to the second differences of the data. In practice, coefficient θ can be considered as a transformation parameter which adjusts 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 respect, the new lines maintain the mean and the slope of the original data, but not their curvatures.

The transformation can become very helpful in time series forecasting as a coefficient of $0 < \theta < 1$ will lead to a less fluctuated line, able to identify the long-term characteristics

*Corresponding author

Email addresses: `spiliotis@fsu.gr` (Evangelos Spiliotis), `vassim@fsu.gr` (Vassilios Assimakopoulos)

of the data like trend (Assimakopoulos, 1995), while a coefficient of $\theta > 1$ will create a highly curved line, emphasizing the short-term characteristics of the data like level. Thus, the method can indirectly distinguish hidden patterns in the data, use proper techniques and models to extrapolate each one of them and combine their findings to boost forecasting performance. The advantage of Theta, as a decomposition method, derives exactly from this simple "divide and conquer" property: There is no single extrapolative model capable of efficiently capturing all the available information hidden in a time series. Yet, if multiple series of reduced amount of information are used instead of the original one, i.e., Theta lines, even the most common models can achieve adequate forecasting performance.

Following the previous suggestions, Theta line θ at point t can be obtained by equation 1a, or 1b according to the simplified modifying formula provided by Nikolopoulos et al. (2012) as follows:

$$Y_t^{\theta} = \theta Y''_t = \theta(Y_t - 2Y_{t-1} + Y_{t+2}) \tag{1a}$$

$$= \theta Y_t + (1-\theta)Y_t^0 = \theta Y_t + (1-\theta)(b + at) \tag{1b}$$

where Y'' the second differences of the data and b and a the intercept and slope of the simple linear regression in time Y^0.

In its original form, called classic Theta, the method consists of two Theta lines with θ coefficients of 0 and 2 calculated on the seasonally adjusted data. The deseasonalization is performed using the classical multiplicative decomposition by moving averages (Makridakis et al., 1998), provided that a significant seasonal pattern has been identified by an autocorrelation test. The first line, of zero fluctuations, is forecasted by extrapolating the linear regression line in time, while the second one, of double curvature, using SES. The forecasts are combined using equal weights and then reseasonalized. This form of Theta participated in the M3 competition (Makridakis & Hibon, 2000) and became popular for outperforming the rest of its competitors including sophisticated methods and expert systems, particularly for monthly series and microeconomic data. Nowadays, classic Theta is sometimes referred as "SES with drift" since the level of the series obtained by extrapolating Y^2 though SES is drifted by half of the slope estimated for Y^0 (Hyndman & Billah, 2003). Although this simplification has been proven to hold only under specific assumptions (Nikolopoulos et al., 2012), it is quite helpful in understanding the way the method works.

Of course, given that Theta is practically a decomposition framework, one can use different lines than these of the classic approach to extrapolate a series, e.g., a double-lined model of coefficients 0 and $\theta > 2$ to further stress the trend component of the series (Constantinidou et al., 2001), or a triple-lined model to extract more information from the provided data (Petropoulos & Nikolopoulos, 2013). Such approaches have been previously examined in the literature with encouraging results, indicating that the Theta model has a lot of potential if properly exploited. Yet, given its simplicity and ease of parameterization, as well as the limited differences observed in terms of accuracy in contrast to more complicated approaches, the two-lined version of the method has become the most popular one, expressed as follows:

$$Y_t = w_{\theta_1} Y_t^{\theta_1} + w_{\theta_2} Y_t^{\theta_2} \tag{2}$$

where w_{θ_1} and w_{θ_2} are the weights of the two Theta lines used, summing up to one.

As noted by Fioruci et al. (2016), in order for the original data to be properly decomposed and reconstructed by the individual Theta lines, for the case of the two-lined model the weights can be directly calculated as follows:

$$w_{\theta_1} = \frac{\theta_2 - 1}{\theta_2 - \theta_1} \tag{3a}$$

$$w_{\theta_2} = 1 - w_{\theta_1} \tag{3b}$$

with the limitations of $\theta_1 \leq 1$ and $\theta_2 \geq 1$.

Given that Y^0 is the only Theta line able of properly dealing with the long term characteristics of the data, it is strongly recommended to be used within the Theta method. In this case, the generic form of the two lined model can be further simplified as follows:

$$
\begin{aligned}
Y_t &= w_0 Y_t^0 + w_\theta Y_t^\theta \\
&= \frac{\theta - 1}{\theta} Y_t^0 + \frac{1}{\theta} Y_t^\theta \\
&= \frac{\theta - 1}{\theta}(b + at) + \frac{1}{\theta} Y_t^\theta \qquad (4)
\end{aligned}
$$

where θ has a value greater than 1 and used to further curve the original data so that the running level of the series, estimated by applying SES on Y^θ, to be drifted by $a\frac{\theta-1}{\theta}$ per period, calculated via linear regression Y^0.

Lots of research has been conducted to identify the optimal parameter of θ and utilize the performance of the method. The studies of Constantinidou et al. (2001), Thomakos & Nikolopoulos (2014) and Fioruci et al. (2016) provide useful insights in that direction and display promising results. Yet, even if the optimal coefficient of θ is identified, according to equation 4 the method will be unable to properly handle time series of non-linear patterns of trend, such as exponential ones. This is because in its classic form, Theta drifts SES forecasts only linearly according to the regression line estimated. The limitation may lead to poor performance, especially for long term forecasts where the component of trend becomes dominant. For instance, damped exponential smoothing is capable of damping the running trend of the series if needed, that way outperforming simpler linear models (e.g., Holt exponential smoothing). Thus, incorporating a relative property in the Theta method becomes promising.

Another limitation of Theta classic is the fact that the components of trend and level, simulated through lines Y^0 and Y^θ, are additively related. The same stands for the seasonal component which, if present, is calculated through a multiplicative decomposition process. Undoubtedly, time series characteristics are not always additively connected, nor is seasonality always multiplicatively expressed. Thus, alternative relations may need to be assumed by a model to effectively forecast any type of data examined. For instance, both additive and multiplicative models are available in ETS, enabling it to efficiently capture the patterns of more complex time series (Hyndman et al., 2002). Therefore, the idea of introducing a multiplicative expression for the Theta method and considering both additive and multiplicative seasonal patterns, becomes encouraging.

The extensions proposed to generalize the Theta method and enhance its flexibility are analytically examined in the next section. Their implementation leads to the creation of the Theta framework, called *4*Theta, which can be exploited for a more holistic modeling and automatic forecasting.

2. The Theta framework (4Theta)

Given the limitations of Theta classic, we propose and implement some modifications to generalize its use and enhance its performance for more complex types of data. This includes mechanisms for considering both linear and non-linear patterns of trend, adjusting trend intensity, recognizing either additive or multiplicative seasonality, as well as additive and multiplicative connections between the components of trend and level.

The extensions proposed transform Theta classic into an integrated forecasting framework, *4*Theta, which generates various forecasting models, each one of different properties. In this respect, using simplified selection criteria, it becomes possible to identify the best model per time series examined and exploit it for extrapolation.

2.1. Adjusting trend pattern and intensity

The Theta method is somewhat limited in its original form as it is forced to generate forecasts which are linearly related. Additionally, the trend of the forecasts is half that of the fitted trend line ($a/2$). Thus, if the long-term components of the data persist in the future, the method will be unable to properly follow their pattern and result to rather pessimistic or optimistic forecasts, depending on the sign of the trend. The same stands for damped trend series, where slope a could be further shrunk to enable a better fit, as well as stationary series for which classic Theta will assume a pseudo-trend through Y^0, possibly leading to poor results. In such cases, an optimal θ coefficient could be used instead of $\theta = 2$ to properly adjust trend intensity, i.e., w_0, and extrapolate the series.

Such examples are presented in Figure 1 where classic Theta (green line) is used to predict a stationary time series (plot a), as well as data with damped (plot b) and linear (plot c) patterns of trend. As seen, in all cases the accuracy of the originally proposed method is far from optimal. Yet, by selecting a proper coefficient of θ the forecasts can be significantly improved. For instance, $\theta = 1$ (SES), $1.5 < \theta < 2$ and $2.5 < \theta < 3$ can be used for the stationary, damped and linear series, respectively (red line), and boost forecasting performance.

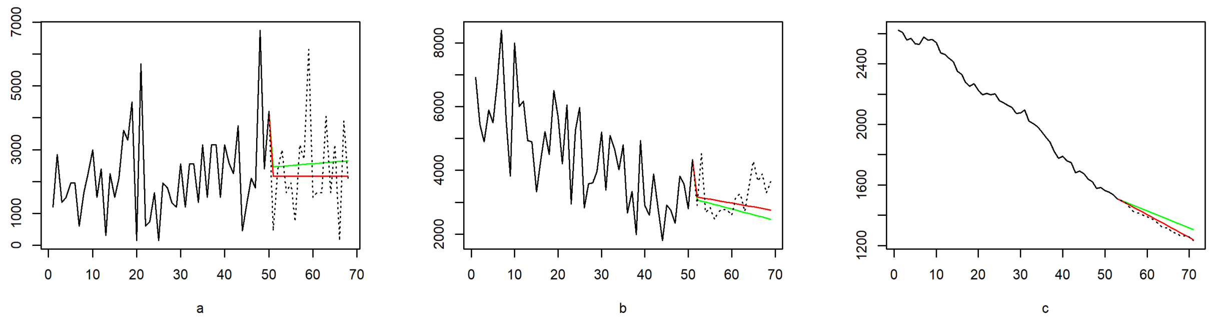

Figure 1: Forecasts generated by classic Theta (green) for a stationary (a), a damped trended (b) and a linear trended (c) time series. The dotted part of the series demonstrates the values observed in the future, while the red lines the forecasts generated for selecting the optimal θ coefficient.

There are many ways for selecting the optimal coefficient of θ (Constantinidou et al., 2001; Thomakos & Nikolopoulos, 2014; Fioruci et al., 2016). To do so efficiently, i.e., without significantly increasing the computational cost of the method, the parameter estimation in this study is based on minimizing the mean absolute errors (MAE) using the Brent method, as implemented in the optimize() function of the R statistical software (R Core Team, 2017). The function applies a combination of golden section search and successive parabolic interpolation, converging rapidly into a reliable solution. Thus, the model that best fits the training sample is accurately identified and used for extrapolating the series. Moreover, to further enhance the parameterization process, we limit the space of potential solutions into $1 \leq \theta \leq 3$. This enables Theta to properly fit the data without exaggerating.

Nevertheless, there are still cases that the method will perform poorly in its original form, even if the optimal coefficient of θ is selected. Exponentially trended series are such examples: Their growth rate, which changes over time, significantly differs from slope a and therefore cannot be effectively extrapolated through a linear model. Figure 2 visualizes this issue, where classic Theta and linear regression, i.e., Theta of $w_0 = 1$, is used to extrapolate an exponentially trended series. As seen, the linear expression of the method always leads to unreasonable results.

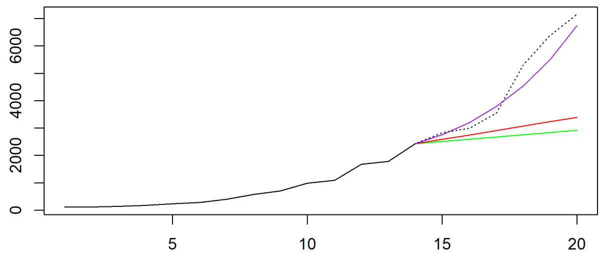

Figure 2: Forecasts generated by the original Theta method for $w_0 = 0.5$ (green) and $w_0 = 1$ (red line) for an exponentially trended series. The dotted part of the series demonstrates the values observed in the future. The purple line demonstrates the forecasts generated by exponentially trended Theta model.

Hopefully, the expansion of the Theta model for non-linear patterns of trend can easily be achieved by simply replacing Y^0 with alternative curves. Given that according to equation 1b Theta lines derive directly trough the selected Y^0, the rest of the equations will hold for any type of curve considered instead. This insight demonstrates the flexibility of Theta and its ability to effectively decompose any type of time series into multiple lines of different properties. Thus, by replacing the originally proposed linear curve with an exponential one, we end up with a non-linear Theta model able of effectively capturing exponential patterns of trend. In the example of Figure 2 the forecasts of such a model are visualized, demonstrating its superiority.

The formulas for estimating the above mentioned curves of Y^0 are given by equations 5 and can be directly incorporated to the generic form of the Theta method (equation 4) to modify the way the forecasts generated through SES are drifted.

$$\text{Linear trend:} \quad {Y_t}^0 = b + at \tag{5a}$$

$$\text{Exponential trend:} \quad {Y_t}^0 = be^{at} \quad \text{, or} \quad log({Y_t}^0) = log(b) + at \tag{5b}$$

where a and b are 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 mention that the parameters of the curves could also have been optimized for minimizing the one-period ahead SSE, as proposed by Fioruci et al. (2016). The former approach is used in a two-fold aim: (i) to maintain the properties of the originally proposed method and (ii) incorporate the long-term memory of the regression line to ensure robustness. If the latter approach was adopted, then the model would emphasize the most recent trend of the data (exponential smoothing logic) and not the long-term one (Theta logic), which is undesirable given that the Theta exploits this feature to mitigate the effect of temporal changes and randomness and perform well on long forecasting horizons.

To sum up, following the modifications presented above, the Theta method can presently mime the properties of more flexible models, namely Damped (damped trend), Holt (linear trend) and Simple (insignificant trend) exponential smoothing (Gardner, 2006). Moreover, it can further improve its forecastability by considering non-linearities and exponential growth rates, which for the case of exponential smoothing are limited captured.

2.2. Additive and multiplicative expression of the Theta method

Many classic forecasting models, such as exponential smoothing, are expressed both in a multiplicative and additive form, meaning that the components of the series (level, seasonality and trend) may be combined in two different ways and relatively interact (Hyndman et al., 2002). For instance, if we assume multiplicative seasonality, the seasonal effect in a trended exponential smoothing model will extend as the level of the series increases, while if we assume an additive one, it will remain relatively constant across time. Similar connections can be found for the rest of the time series components, enabling flexible models like ETS to effectively follow the latest patterns of the data through an iterative procedure.

In contrast, according to equation 4, the Theta method averages the forecasts of SES and regression in time to extrapolate the data. Thus, the level and trend component are independent from each other and additively connected. The same stands for the seasonal component which, given that the data are externally seasonally adjusted (if needed), is not directly handled by the method itself. Moreover, since adjustment is performed through a multiplicative decomposition process, seasonal factors multiplicatively interact with the forecasts. In other words, Theta is far more deterministic than iterative, letting regression on Y^0 deal with trend, SES on Y^θ with level, and the deseasonalization process on Y^1 with seasonality. Thus, classic Theta becomes quite static and can be considered as a deterministic equivalent of the Holt-Winters multiplicative model.

Given the limitations above, to enable both an additive and a multiplicative connection between the components of the series, a Theta line can be alternatively estimated using the following formulas:

$$\text{Additive } \theta \text{ line:} \quad {Y_{\boldsymbol{at}}}^{\theta} = \theta Y_t + (1-\theta){Y_t}^0 \tag{6a}$$

$$\text{Multiplicative } \theta \text{ line:} \quad {Y_{\boldsymbol{mt}}}^{\theta} = \frac{Y_t^{\boldsymbol{\theta}}}{{Y_t^0}^{\boldsymbol{\theta-1}}} \tag{6b}$$

where $Y^{\boldsymbol{\theta}}$ denotes the θ_{th} power of the original or seasonally adjusted data.

The generic expression of the model is then given as follows per case:

$$\text{Additive } \theta \text{ model:} \quad Y_{\boldsymbol{at}} = \frac{\theta-1}{\theta}{Y_t}^0 + \frac{1}{\theta}{Y_{\boldsymbol{at}}}^{\theta} \tag{7a}$$

$$\text{Multiplicative } \theta \text{ model:} \quad Y_{\boldsymbol{mt}} = \sqrt[\boldsymbol{\theta}]{{Y_{\boldsymbol{mt}}}^{\theta}}\ \sqrt[\frac{\boldsymbol{\theta}}{\boldsymbol{\theta-1}}]{{Y_t}^0} \tag{7b}$$

where Y^0,${Y_{\boldsymbol{a}}}^{\theta}$ and ${Y_{\boldsymbol{m}}}^{\theta}$ are extrapolated independently, as previously described.

As seen, according to equation 7b, the multiplicative expression of the Theta method can now be exploited to handle data displaying multiplicative relations of level and trend. Moreover, multiplicative deseasonalization can be replaced by an additive one to introduce additive components of seasonality in both cases. In total, after the expansion suggested, the Theta method can generate up to six models of different properties but similar principles:

- (N,A): Negligible seasonality - Additive relations
- (N,M): Negligible seasonality - Multiplicative relations
- (A,A): Additive seasonality - Additive relations
- (A,M): Additive seasonality - Multiplicative relations
- (M,A): Multiplicative seasonality - Additive relations
- (M,M): Multiplicative seasonality - Multiplicative relations

which philosophy is close to that of the state space exponential smoothing methods.

At this point we note that, since multiplicative models can only handle data and generate forecasts of positive values, they become suitable for applications where negative forecasts are meaningless (e.g., demand, price or sales). On the other hand, the constraint of Y^0 and ${Y_{\boldsymbol{m}}}^{\theta}$ being both greater than zero is a limitation which should be taken into consideration before adopted. Yet, this constraint is applied to all multiplicative forecasting models found in the literature and should not question the usefulness of the approach.

Finally, given the suggestions of Section 2.1, two different patterns of trend can also be examined per case, leading in total to 12 Theta models of respective properties. Thus, model *4*Theta(M,A,L) will denote classic Theta and model *4*Theta(A,M,E) an exponentially trended model that multiplicatively connects the components of trend-level assuming additive seasonal adjustments.

2.3. Model selection

The advantage of substituting Theta classic with a set of models is quite blatant: The flexibility of the method is enhanced and more complex patterns of time series can be adequately captured. Moreover, given a selection criteria, one can directly detect the best performing model available in the Theta framework and use it to automatically extrapolate a set of diverse time series.

The selection criteria used in the literature vary, mainly depending on the computation time available. Given that the aim of *4*Theta is to efficiently provide automatic generated

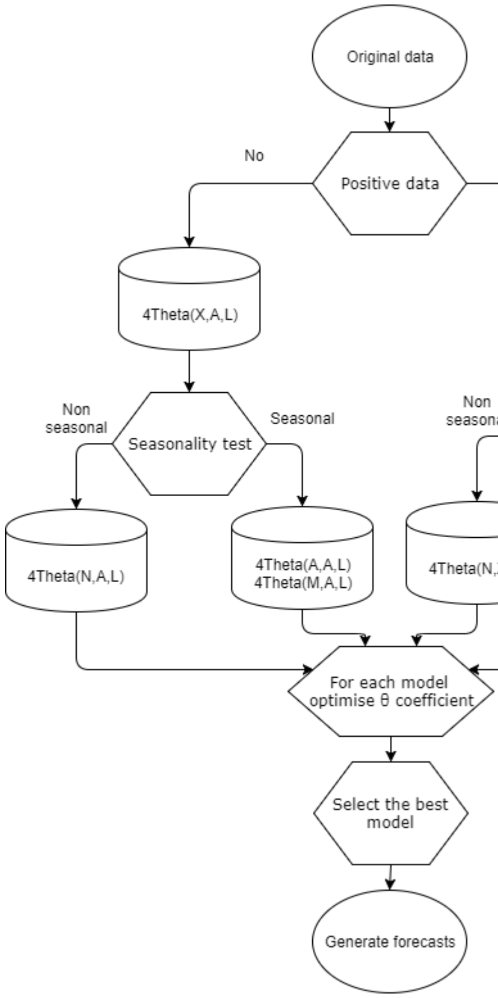

Figure 3: Flowchart of the 4Theta methodological approach. X stands for all options available per modeling parameter.

forecasts, the most effective options are examined. Using paradigms from the best practices found in the literature, such a criterion could be either a goodness of fit measure, like Mean Absolute or Squared Error (MAE or MSE), or an information criterion, like the AIC (Sakamoto et al., 1986). The latter is preferred in cases that complexity must be taken into consideration so that sophisticated models which perform similarly to less complex ones can be discarded and avoid over-fitting. Yet, for the case of 4Theta all models have identical complexity, i.e., number of parameters that need to be estimated. In this regard, an information criteria does not add any value over more simplistic solutions and a typical error metric, like MAE, can be used to identify the goodness of fit for each model in terms of forecasting accuracy, as follows

$$MAE = \frac{1}{n}\sum_{n=1}^{n}|Y_t - \hat{Y}_t| \tag{8}$$

where $\hat{Y}$ are the forecasts generated by the model and n the number of the available data points for training.

Therefore, for each time series examined, all possible models are estimated through *4*Theta and MAE is applied to determine which one should be selected for extrapolation. In case of non-positive data, the initial number of models is accordingly shrunk (exponential and multiplicative models are excluded) so that the computational time is further boosted. Respectively, a seasonality test is used to determine whether seasonal or non-seasonal models should be considered. The methodological approach of *4*Theta is summarized in Figure 3.

References

Assimakopoulos, V. (1995). A Successive Filtering Technique for Identifying Long- term Trends. *Journal of Forecasting, John Wiley*, *14*, 35–43.

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

Constantinidou, C., Nikolopoulos, K., Bougioukos, N., Tsiafa, E., Petropoulos, F., & Assimakopoulos, V. (2001). A neural network approach for the theta model. *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*, 11511161.

Gardner, E. S. (2006). Exponential smoothing: The state of the art-Part II. *International Journal of Forecasting*, *22*, 637–666.

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

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.

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

Makridakis, S. G., Wheelwright, S. C., & Hyndman, R. J. (1998). *Forecasting: Methods and applications (Third Edition)*. New York: Wiley.

Nikolopoulos, K., Assimakopoulos, V., Bougioukos, N., Litsa, A., & Petropoulos, F. (2012). The theta model: An essential forecasting tool for supply chain planning. In G. Lee (Ed.), *Advances in Automation and Robotics, Vol. 2: Selected Papers from the 2011 International Conference on Automation and Robotics (ICAR 2011), Dubai, December 1–2, 2011* (pp. 431–437). Berlin, Heidelberg: Springer Berlin Heidelberg.

Petropoulos, F., & Nikolopoulos, K. (2013). Optimizing the theta model for monthly time series. In *5th International Conference of Agents and Artificial Intelligence,15–18 February 2013. Barcelona, Spain*.

R Core Team (2017). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing Vienna, Austria.

Sakamoto, Y., Ishiguro, M., & Kitagawa, G. (1986). *Akaike Information Criterion Statistics*. D. Reidel Publishing Company.

Thomakos, D., & Nikolopoulos, K. (2014). Fathoming the theta method for a unit root process. *Oxford Journal of Management Mathematics*, *25*, 105124.