Exploring the Theta Method

Gerwyn Ng *

23 April 2020

Abstract

Accuracy, robustness and reliability are the key objectives in forecasting performance. In this context, a new method, the Theta method, has gain significant attention from researchers and practicing forecasting due to its remarkable performance in the M3-Competition (2000). The Theta method (Assimakopoulos & Nikolopoulos, 2000) is a univariate forecasting method which decomposes the deseasonalized data into two or more components, referred to as "Theta-lines". The decomposition method is based on the concept of modifying the local curvature of the original time-series through a coefficient 'Theta', that is applied directly to the second differences of the data. The Theta-lines are then extrapolated, and recombined to produce the final forecasts. This paper reviews the methodology and empirical results of the classical framework (Assimakopoulos & Nikolopoulos, 2000) as well as the key extensions/modifications proposed, namely the Optimized Theta Method (Fioruci et al., 2015) and AutoTheta (Spiliotis et al., 2020).

1 Introduction

The primary purpose of forecasting is to produce accurate, robust and reliable predictions. As such, the value of both theoretical and applied forecasting methods are often evaluated based on their contributions to improving the accuracy of post-sample predictions. By comparing the post-sample forecasting accuracy of various methods, the empirical performance of these methods can be determined in an objective, measurable manner. The Makridakis Competitions (M-Competitions), led by Spryos Makridakis, are such empirical studies that have evaluated the performance of a larger number of major time series forecasting methods led by recognized experts. In particular, one competition series, the M3-Competition (Makridakis & Hibon, 2000), selected 3003 series on a quota basis to include various types of time series data (micro, macro, industry, finance, etc.) and different time intervals between successive observations (yearly, quarterly, etc). The M3-Competition concluded with a relative simple approach, the Theta method, proposed by Assimakopoulos & Nikolopoulos (2000) (hereafter A&N), outperforming other competitors methods, which included

*Masters of Science (Economics) Programme, Singapore Management University

classical methods and specialized software, using various evaluation metrics across different data frequency and data categories. Although the classical Theta method is straightforward to use and is not based on influential statistical theories, its remarkable forecasting performance has called researchers' and practising forecasters' attention.

The main contribution of the classical Theta method by A&N is a novel decomposition approach which exploits information in the original data that may be captured or modelled by a single, direct extrapolation on the original data. The idea behind this decomposition is a simple "divide and conquer" strategy where the original time-series is first decomposed into short-term features and a long-term trend by applying a coefficient θ on the second differences of the data, and then extrapolated separately before recombining to give the final forecasts. In its classical form, A&N considered the simplest form where the seasonally adjusted data is decomposed into two Theta-lines with coefficients 0 and 2. Later on, Fioruci et al. (2015) extended this method with an optimal selection of the coefficient θ that best describe the short-term features of the series while maintaining the long-term trend component. The empirical studies revealed that the choice of θ can, indeed, influence the forecasting performance of the classical method, indicating that the decomposition approach has value in improving forecasting accuracy. However, the Theta method appears to be suitable only very limited circumstances given its assumptions of a linear long-term trend and a additive relationship between the component forecasts. In this sense, Spiliotis et al. (2020) transformed the decomposition framework of Theta into a generalised forecasting algorithm capable of producing automated extrapolation with flexibility to deal with non-linear patterns of trend, trend intensity and a multiplicative expression of the decomposition approach.

In spite of a lack of strong statistical explanation, the Theta method has seen modifications and extensions which resulted in significant improvement in forecasting accuracy since its inception. Bearing in mind that both empirical and theoretical studies are equally important in advance the field of forecasting, and related domains (Makridakis & Hibon, 2000), I believe that the Theta method is worth examining to understand the interesting theoretical and empirical issues it has raise. In this aspect, consolidating the anomalies between theory and practice will offer us insights on the potential of the proposed decomposition approach and its value in predicting time-series.

The paper is organized as follows. Section 2 describes the classical Theta method of A&N. Section 3 describes the Optimized Theta Method proposed by Fioruci et al. (2015). Section 4 describes the AutoTheta algorithm proposed by Spiliotis et al. (2020).

2 The Classical Theta Method

The classical Theta method decomposes the seasonally adjusted time-series into two components, referred to as "Theta-lines". The Theta-lines are constructed by modifying the local curvature of the time-series through a coefficient "Theta" (θ) that is applied directly to the second differences of the original time-series. Each Theta-line is then extrapolated separately with an appropriate method depending on its characteristic. Finally, the Theta-lines are recombined to produce the final forecasts.

2.1 Constructing the Theta-lines

Noting that the data can be written as:

$$X_i = X_1 + (i-1)(X_2 - X_1) + \Big(\sum_{t=2}^{i-1}(i-t)X''_{t+1}\Big),$$
$$\text{where } X''_t = (X_t - X_{t-1}) - (X_{t-1} - X_{t-2})$$

the local curvatures of the series are then modified by applying a coefficient Theta, $\theta \in \mathbb{R}$, to the second differences of the time series such that:

$$Y''(\theta) = \theta X''_{data} \tag{1}$$

Intuitively, the second differences can be interpreted as the discrete analogy of the second derivative. That is, if the local curvatures are gradually reduced ($\theta < 1$), then the time series is subdued, emphasising on long-term characteristics of the data. On the other hand, if the local curvature is increased ($\theta > 1$), then the time series is elaborated, magnifying short-term characteristics of the data. In the extreme case where $\theta = 0$, the time series is transformed to a linear regression line (e.g. a linear time trend).

Following this procedure, the Theta-lines is constructed as follows:

$$Y_i(\theta) = Y_1 + (i-1)(Y_2 - Y_1) + \theta\Big(\sum_{t=2}^{i-1}(i-t)X''_{t+1}\Big)$$

While the placement of these lines with respect to the original data can be done in many ways, A&N estimated Y_1 and $(\mathrm{Y}_2 - \mathrm{Y}_1)$ using the standard OLS estimation procedure. That is, for a fixed θ, Y_1 and $(\mathrm{Y}_2 - \mathrm{Y}_1)$ are the values which minimize the sum of squared errors:

$$\operatorname*{argmin}_{Y_1,(Y_2-Y_1)} \Big(\sum_i (Y_i - X_i)^2\Big) \tag{2}$$

The solutions were derived by A&N with intensive algebraic manipulations on the respective first order conditions. On the other hand, Hyndman & Billa (2001) replicated the results in relatively straightforward approach, which I find it to be more intuitive in understanding the method, by noting that (1) is a second-order difference equation and thus, has the solution:

$$Y_i(\theta) = a_\theta + b_\theta(i-1) + \theta X_i \tag{3}$$

where a_θ and b_θ are constants. That is, (3) can be simplified to a regression of $(1-\theta)X_i$ against time $(i-1)$. Thus, the solution is given by:

$$\hat{b_\theta} = \frac{6(1-\theta)}{n^2-1}\big(\frac{2}{n}\sum_{t=1}^{n} tX_t - (n-1)\bar{X_t}\big)$$
$$\text{and } \hat{a_\theta} = (1-\theta)\bar{X}_t - \hat{b_\theta}(n-1)/2$$

and the mean value of the Theta-line is given by:

$$\bar{Y}(\theta) = \hat{a_\theta} + \hat{b_\theta}(n-1)/2 + \theta\bar{X} = \bar{X}$$

These results reveals an important property of the Theta-lines. That is, the mean and the slope of the Theta-lines remain the same ($b_\theta/(1-\theta)$ is independent of θ) as compared to those of the original data, with the local curvatures of the data filtered out or magnified depending on the parameter θ.

An alternative formula of the Theta-line (Nikolopoulos et al., 2011) is also provided:

$$Y(\theta) = \hat{\alpha} + \hat{\beta}t + \theta\hat{e} \tag{4}$$

where $\hat{\alpha}$ and $\hat{\beta}$ are the standard least square estimates of time series $\mathrm{X_t}$ over time t. In this sense, a Theta-line is simply *the linear time trend regression adjusted by the residuals multiplied by the parameter* θ. Consistent with A&N's derivations, the original time series remains untouched when $\theta = 1$.

In addition, by redefining the residuals and rearranging terms, the Theta-lines can also be expressed as:

$$\begin{aligned} Y(\theta) &= \hat{\alpha} + \hat{\beta}t + \theta\hat{e} \\ &= \hat{\alpha} + \hat{\beta}t + \theta(X_t - \hat{\alpha} - \hat{\beta}t) \\ &= \theta X_t + (1-\theta)(\hat{\alpha} - \hat{\beta}t) \end{aligned} \tag{5}$$

From this perspective, the Theta-lines are simply *a weighted average of the original time series and the linear time trend regression.*

All three definition of the Theta-lines, (3), (4) and (5), accentuates the desirable property of decomposing the original data with Theta-lines. That is, the resulting series created maintain the mean and slope of the original data but not the curvatures. In other words, the Theta-lines provide flexibility to modify the characteristic of the original series (via the coefficient θ) without losing important information. As a result, better approximations of the long-term trends or the short-term features of the series can be made.

2.2 General Formulation of the Theta Method

In its original form, A&N decomposed the seasonally adjusted data (classical multiplicative decomposition by moving-averages after series was tested for statistical significant seasonal behavior using an auto-correlation test) into two Theta-lines with θ coefficients chosen such that they are symmetrical to one i.e $\theta_1 = 0$ and $\theta_2 = 2$:

$$Data = \frac{1}{2}\left[Y_1(\theta_1 = 0) + Y_2(\theta_2 = 2)\right] \tag{6}$$

$\mathrm{Y}_1(\theta_1)$ is the linear regression line of the data and $\mathrm{Y}_2(\theta_2)$ has second order differences exactly twice the original time series. Here, $\mathrm{Y}_1(\theta_1)$ describe the long-term linear time trend while $\mathrm{Y}_2(\theta_2)$ magnifies the short-term behaviour using doubled local curvatures. $\mathrm{Y}_1(\theta_1)$ is extrapolated using the regression line while $\mathrm{Y}_2(\theta_2)$ is extrapolated via Simple Exponential Smoothing (SES). The forecasts are combined using a simple average and then reseasonalized to produce the final forecasts.

This form was employed by A&N to produce forecasts for M3-Competition and became popular for outperforming other competing methods, which included both classical methods and specialized software, particularly for monthly series and microeconomics data.

2.3 Performance

The 3003 series of the M3-Competition were selected on a quota basis to include various types of time series data (micro, industry, macro, etc.) and different time intervals between successive observations (yearly, quartyly etc.). Each series was also decided to have a minimum number of observations to ensure enough data is available to develop an adequate forecasting method. Refer to Table 1 for more details about the 3003 series.

Frequency	Micro	Industry	Macro	Finance	Demog	Other	Total
Yearly	146	102	83	58	245	11	645
Quarterly	204	83	336	76	57	0	756
Monthly	474	334	312	145	111	52	1428
Other	4	0	0	29	0	141	174
Total	828	519	731	308	413	204	3003

Table 1: The M3-Competition 3003 series data

The 24 methods used to produce forecasts are listed in Table 2. A deseasonalised naive model (random walk) and the Dampen Trend Exponential Smoothing were set as benchmark models.

Method	Description
Simple	
1. Naive2	Deseasonalized Random Walk
2. Single	SES
Explicit Trend Methods	
3. Holt	Automated Holt's Linear Exponential Smoothing (two parameters)
4. Robust-Trend	Non-parametric version of Holt with median estimates of trend
5. Winter	Holt-Winter's linear and seasonal exponential smoothing
6. Dampen	Dampen Trend Exponential Smoothing
7. PP-autocast	Commercially available forecasting package of Dampen
8. Theta-sm	Successive smoothing plus a set of rules for dampening the trend
9. Comb S-H-B	Combination of Single, Holt and Dampen
Decomposition	
10. Theta	Decomposition approach, extrapolation and combination of individual forecasts
ARIMA/ARARMA	
11. B-J automatic	Box-Jenkins methodology of 'Business Forecast System'
12. Autobox1	Robust ARIMA univariate Box-Jenkins with/without Intervention Detection
13. Autobox2	
14. Autobox3	
15. AAM1	Automatic ARIMA with/without intervention analysis
16. AAM2	
17. ARARMA	Automated Parzen's methodoloy with auto regressive filter
Expert Systems	
18. ForecastPro	Exponential Smoothing/Box Jenkins/Poisson and negative binomial models/Croston's Method/Simple Moving Average
19. SmartFfcs	Conducts a forecasting tournament among four exponential smoothing and two moving average methods
20. RBF	Rule based forecasting using random walk, linear regression and Holt's
21. Flores/Pearce1	Expert system
22. Flores/Pearce1	
23. ForecastX	Runs tests for seasonality and outliers before selecting from several methods
Neural Networks	
24. Automat ANN	Automated Artifical Neural Network

Table 2: The 24 methods included in the M3-Competition

In addition, the five evaluation metrics used in the M3-Competition to analyze the performance of the various methods are as follow (Makridakis & Hibon, 2000):

- **Symmetric Mean Absolute Percentage Error (sMAPE)** defined as:

$$\sum \frac{X - F}{(X + F)/2} * 100 \tag{7}$$

where X is the actual value and F is the forecast. The sMAPE penalises both under and over forecasts equally and bounds the measures with limits.

- **Average Ranking:** the average ranking of symmetric absolute percentage error of each method across each series and forecasting horizon. It aggregates the relative position of a chosen method across each series and forecasting horizon with equal and normalized weights.
- **Percentage Better:** the percentage of time that a given method has a smaller forecasting error than another method. It measure how much improvement a chosen method has over the benchmark models.
- **Median Symmetric Absolute Percentage Error**. Similary to sMAPE but it reports the median instead. It is more robust to extreme values.
- **Median Relative Absolute Error (Median RAE):** Absolute error for the proposed model relative to the absolute error for a benchmark model (deseasonalized random walk). A RAE score of 0 implies a perfect forecast, a score of 1 implies identical forecasts to the benchmark model, and a score greater than one implies forecasts worse than the benchmark model. Also, the median is taken to provide robust comparisons across alternative models.

Overall, the Theta method performed best across all data for all forecasting horizon and average of forecasting horizon (grouped horizons) when evaluated using an average sMAPE (see Table 6, Makridakis & Hibon, 2000). For microeconomics data, the Theta method performed best across all evaluation metrics. However, mixed results were observed for other categories (see Table 7, Makridakis & Hibon, 2000). In addition, the Theta method also performed best for quarterly data across all categories (except industry data) (see Table 9, Makridakis & Hibon, 2000). Lastly, the Theta method gave the best results for monthly data (see Table 14, Makridakis & Hibon, 2000).

2.4 Evaluation

Based on its performance in the M3-Competition, it appears that the Theta method is somewhat generalised enough to produce robust forecasts as observed from a lowest average sMAPE across all data (relative to the other 23 methods). The results also highlights the strength of the Theta method in dealing with microeconomics and monthly data.

A&N attributed the success of the Theta method to the decomposition approach, specifically the choice of $\theta_1 = 0$ and $\theta_2 = 2$. The reasoning was that the linear Theta-line, $Y_1(\theta_1 = 0)$, captured long-term trend of the time-series which is often neglected when a method tries to adapt to more recent trends or fluctuations. In the case of monthly data, which are generally characterized by a relatively large amount of volatility, most methods are not able to keep the long-term trend in memory. In contrast, the use of two Theta-lines, $Y_1(\theta_1 = 0)$ extrapolates the long-term trend linearly, keeping the long-term trend in memory, while $Y_2(\theta_1 = 2)$ increases the roughness of the series to augment short-term features, exploits more information from the original data to make more accurate forecast. Furthermore, the combination of the component forecasts to produce the final forecast also contributed to improving accuracy performance. In fact, the accuracy of combination of various methods outperforms, on average, the specific methods being combined and does well in comparison with other methods (Makridakis & Hibon, 2000).

However, under certain circumstances, the classical Theta method is equivalent to the SES with drift (Hyndman & Billah, 2003). Specifically, a SES with an added trend plus a constant, where the slope of the trend is half that of the fitted trend line through the original series (since equal weights were assigned to the components, as in (6)). Thus, the choice of Theta in the classical method tells us little about the potential of the decomposition approach proposed by A&N. In context, the difference in performance between the classical Theta method and its competitors could simply due to Theta's way of deseasonalizing the data, the concept of combining the individuals component forecasts or a combination of both. It also do not tell us the reason that makes the Theta method well suited for microeconomics and monthly data. For instance, is it Theta's ability to deal with extreme values?

Evidently, the classical Theta method reveals little about the decomposition approach itself as there are too many dynamics that are at play during its implementation in the M3-Competition. As such, paying more attention to the decomposition approach itself is one way we can uncover insights that will make the Theta method more useful or relevant in the field of forecasting.

3 The Optimized Theta Method (OTM)

One promising characteristic of the model worth exploring is the parameter θ itself. Since the classical method only considers an abitrary choice of θ, this section considers the Optimized Theta Method (Fioruci et al., 2015) to understand the impact of the choice of θ have on the forecasting accuracy.

A generalisation of the Theta-lines combination in (6) can be expressed as follows:

$$Data = \omega Y_{1,t}(\theta_1) + (1-\omega)Y_{2,t}(\theta_2) \tag{8}$$

where $\omega \in [0, 1]$ is a weight parameter. Since our objective is to tune the parameter θ, the weight parameter must be a function of θ_1 and θ_2 to satisfy (7). To see this, we substitute (5) into (7):

$$\begin{aligned}
X_t &= \omega\big[\theta_1 X_t + (1-\theta_1)(\hat{\alpha}-\hat{\beta}t)\big] - (1-\omega)\big[\theta_2 X_t + (1-\theta_2)(\hat{\alpha}-\hat{\beta}t)\big] \\
&= \hat{\alpha} - \hat{\beta}t + \omega\theta_1(X_t - \hat{\alpha} + \hat{\beta}t) + (1-\omega)\theta_2(X_t - \hat{\alpha} + \hat{\beta}t) \\
(X_t - \hat{\alpha} + \hat{\beta}t) &= \omega\theta_1(X_t - \hat{\alpha} + \hat{\beta}t) + (1-\omega)\theta_2(X_t - \hat{\alpha} + \hat{\beta}t) \\
1 &= \omega\theta_1 + (1-\omega)\theta_2 \\
\omega &= \frac{\theta_2 - 1}{\theta_2 - \theta_1} \qquad (9)
\end{aligned}$$

with the restrictions $\theta_1 \leq 1$ and $\theta_2 \geq 1$. Given that the $Y_t(\theta = 0)$ is the only Theta-line capable of dealing with the long-term characteristic of the data, the OTM restricts $\theta_1 = 0$ and only considers optimizing $\theta_2 = \theta$. Thus the decomposition for the OTM is given by:

$$X_t = \big(1 - \frac{1}{\theta}\big)(\hat{\alpha} + \hat{\beta}t) + \frac{1}{\theta}Y_t(\theta) \qquad (10)$$

where $\theta \geq 1$. It follows that the forecasts for k steps ahead of t are given by:

$$\hat{X}_{t+k|t} = \big(1 - \frac{1}{\theta}\big)\big[\hat{\alpha} + \hat{\beta}(t+k)\big] + \frac{1}{\theta}\hat{Y}_{t+k|t}(\theta) \qquad (11)$$

where $\hat{Y}_{t+k|t}(\theta)$ is extrapolated using SES as in A&N (although in practice, any extrapolation method can be used). When $\theta = 1$, the forecast is $\hat{Y}_{t+k|t}(1)$ which reduces the OTM to the defined extrapolation method applied on X_t. When $\theta \geq 1$ with SES applied, the OTM acts as an extension to the SES with a drift of $\big(1 - \frac{1}{\theta}\big)\hat{\beta}$ per period t.

3.1 Estimation

The optimal θ is estimated by minimizing a symmetric loss function using a Generalised Rolling Origin Evaluation (GROE). The GROE is formalized by Fioruci et al. (2015) using ideas from a rolling evaluation scheme (Tashman, 2000). Essentially, the GROE is the specific validation scheme chosen to optimized θ.

The idea of a validation scheme is to choose θ such that the out-of-sample performance is maximized. In practice, however, out-of-sample performance is unobserved. As such, one potential strategy ("leave p out validation") is to arbitrarily split the N in-sample data into two : $N - p$ samples for fitting Theta (training sample) and p samples for evaluating the forecasting performance associated with the fitted θ (validation sample). This way, the forecasting performance on the validation sample acts as a proxy for the out-of-sample performance for the chosen θ. For time series, the splitting is usually done in consecutively index to preserve the order between observations (e.g. with N in-sample observations, the first $N - p$ in-sample are used for training while the remaining p is used for validation).

Just one measurement of validation performance is often not reliable since there could difference in the characteristics or distributions between the chosen validation samples and actual out-of-samples, especially for time series data which tends to be unstable. To enhance the robustness of a validation scheme, multiple "folds" of training and validation samples can be created. This way, the performance of the estimated θ can be evaluated several times, depending on the number of folds created (this is often referred to as "Cross-Validation" in the machine learning field). To preserve the order of the time series, k folds with forecast horizon h can be created from N observations as follows:

- Choose origin n_1, number of folds k and the number of movements from the origin m
- Define subsequent origins as $n_i = n_{i-1} + m$ for $i \in [2, k]$ such that $n_k < N$
- For each fold, define observations with time index $\leq n_i$ as training sample and n_i to $n_i + h$ as validation sample if $n_i + h \leq N$. For n_k, the validation sample will be $[n_k, N - n_k]$

This is exactly the validation scheme referred to as GROE used by the OTM where the loss function is simply an aggregation of the validation loss by each fold:

$$loss(\theta) = \sum_{i=1}^{k} \sum_{j=1}^{min(h, N-n_i)} g(X_{n_i+j}, \hat{Y}_{n_i+j|n_i}) \tag{12}$$

where $g(.)$ is the chosen loss function.

The optimal θ is then estimated recursively to minimise this loss. An important aspect of the estimation process is that the Theta-lines do not present great variation for a small neighbourhood of θ (Fioruci et al., 2015). Thus, the parameter space can be restricted and discretized for faster search.

3.2 Algorithm

Step 0. *Seasonality Testing:* the data is tested for seasonality behavior using an autocorrelation test.

Step 1. *Deseasonalization:* if needed, the data is deseasonalized using the classical decomposition method, assuming multiplicative relation of the seasonal component.

Step 2. *Deseasonalization:* estimate $\hat{\theta}$ by minimising (11)

Step 3. *Decomposition:* decompose the time series into two Theta-lines.

Step 4. *Extrapolation:* extrapolate the linear regression trend the usual way while the second line is extrapolated using SES.

Step 5. *Combination:* recombine the two forecast components according to (10).

Step 6. *Reseasonalization:* reseasonalize the data

These steps are exactly the same as the classical method with the exception of Step 3. and Step 5.

3.3 Performance

To obtain empirical support for the OTM, Fioruci et al. (2015) considers various GROE parameters, p, n_1, h, several loss functions as well as different extrapolation methods. The data used in their empirical study is the same as in section 2.3. Refer to Table 1 for more details. To compare the performance of the proposed OTM, the classical Theta method was benchmarked.

For the cost function g, squared error (SE), absolute error (AE) and symmetric Absolute Percentage Error (sAPE) were considered. As for the evaluation of the out-of-sample performance, two evaluation metrics were considered: sMAPE, as in (7), and the Mean Absolute Scaled Error (MASE). The MASE, is the mean absolute error of the forecast scaled by the mean absolute error of the in-sample one mean absolute error of the forecast values, divided by the mean absolute first difference (Hyndman & Koehler, 2006), i.e.,

$$MASE = \frac{\frac{1}{h}\sum_{i=1}^{h}|X_{n+1} - \hat{Y}_{n+1|n|}}{\frac{1}{n-1}\sum_{t=2}^{n}|X_t - X_{t-1}|} \tag{13}$$

The denominator can be interpreted as the absolute error associated with a naive forecast (e.g. using the actual value from the prior period as the forecast). Thus, the MASE has a nice interpretation of comparing relative forecast error against the naive method. In other words, a MASE value > 1 implies the forecast value under consideration performed better than the naive method. As such, it represents a comparative accuracy of forecasts.

Different GROE approaches were considered to obtain the optimal validation scheme in the estimation θ. h is fixed according to the forecast horizon set in the M3-Competition. The set of parameters for the GROE are listed in Table 3.

With regards to the performance of the OTM, the three cost functions obtain very similar results for both accuracy metric for the out-of-sample evaluation. When all the data are considered, the variation in the results were negligible (≤ 0.09 for sMAPE and ≤ 0.01 for MASE). As for the GROE parameters, (a) to (d) performed better than (e) to (h) with (d) providing the best results and (h) the worst. Based on the results for (a) to (d), the OTM performed better than the classical method for all data and across all time frequencies.

As a robustness test, the Multiple Comparisons with the Best test (MCB) was conducted to compare statistical significance of the eight approaches of the OTM when compared to the classical

Approaches	k	m	n_1
(a)	1	h	$n-h$
(b)	2	$\lfloor h/2 \rfloor$	$n-h$
(c)	3	$\lfloor h/3 \rfloor$	$n-h$
(d)	h	1	$n-h$
(e)	2	h	$n-2h$
(f)	4	$\lfloor h/2 \rfloor$	$n-2h$
(g)	6	$\lfloor h/3 \rfloor$	$n-2h$
(h)	h	h	$n-2h$

Table 3: Set of parameters used in the GROE

method. The idea to create rank intervals for each method and identify overlapping rank intervals between two methods. That is, if a pair of methods have overlapping rank intervals, then the null hypothesis of both methods having same accuracy cannot be rejected. With a significance level of 5%, approach (h) is the statistically the worse estimation approach amongst (a) to (h). However, when compared to the classical method, (h) is still statistically better. This is encouraging result that the OTM, through optimizing θ, can achieve better performance than the classical method.

In addition, three methods: SES, Holt and Damped were considered in the extrapolation of the second Theta-line in their study. The results for Holt are always than worse than the standard SES, while Damped produced mix results, with slight improvements for time series with 'others' frequency.

3.4 Evaluation

The extension of OTM is to use information in the data to choose the parameter θ. Since the key contribution of the classical Theta is its ability to exploit information from the time series in the decomposition step, it is natural that the choice of θ should somewhat be led by the characteristic of the data itself. Thus, the OTM explores an estimation process with a GROE validation scheme depending on the data we are dealing with. The empirical studies by Fioruci et al. (2015) offered two key insights. Firstly, the choice of θ can improve the forecasting performance of the classic method, significantly when ranks are taken into account. Secondly, the choice of extrapolation for the second Theta-line can affect the forecasting performance (although it was not the focus of their study).

An obvious drawback of the OTM relative to the classical method is the increase in computational resources since the estimation of θ requires a validation scheme and also a model selection process to produce the final forecast. In this sense, the restriction of the parameter space and choice of validation scheme is also dependent on human judgement and thus, estimating θ may be counterproductive under certain setups.

Furthermore, as the assumption of the linear long-term trend is maintained, even an optimized

θ may not be capable of handling time series with non-linear long-term components. Also, the assumption of an additive relation between the two Theta-lines and the assumption of multiplicative seasonality remain unquestioned. These assumptions may challenge the Theta method's ability in dealing with time series data that are not additively related or with an additive seasonal behavior.

4 AutoTheta

Apart from the adjusting the trend intensity via the parameter θ, 4Theta (Spiliotis & Assimakopoulos, 2018) proposes to transform the classical Theta into a generalized forecasting algorithm, capable of automatic extrapolation with enhanced flexibility and improved properties. The 4Theta algorithm was implemented in the M4-Competition and later on, developed into an open-source solution (in R environment) named AutoTheta[1] (Spiliotis et al., 2020).

This section examines the extensions on the decomposition framework of AutoTheta and evaluates its performance, compared to its classical form.

4.1 Framework

To generalize the classical Theta for more complex types of data, the AutoTheta relaxes the assumptions of a linear long-term trend, a multiplicative seasonality, and the additive relationship between the components of trend and level (represented by the Theta-lines). The key idea is to use a simplified model selection criteria to select the best model from a set of models, and then exploit it for extrapolation.

4.1.1 In-Sample Based Estimation of θ

There are many ways to select the parameter θ, one of which is the OTM discussed in Section 3. However, to reduce the computation cost of the estimation process, AutoTheta proposes using the Brent method, minimizing the Mean Absolution Error (MAE). The Brent method is a combination of the golden search section and the quadratic approximation, with additional checks in place to the quadratic approximation method to make it more robust. As such, it has the reliability of the golden search method but converges relatively quicker. Thus, the θ is chosen based on a value that best fits the training sample. This is in contrast to the OTM which chooses θ based on out-of-sample accuracy. The rationale is that using in-sample captures a long-term trend (Theta logic) as compared to using out-of-sample which may puts an emphasis on more recent trend (SES logic).

Similar to the OTM, the parameter space of θ is also limited, $1 \leq \theta \leq 3$. The restriction can be interpreted as a regularization technique to prevent over-fitting since θ is fitted on the training sample.

[1] Avaliable here: `https://github.com/vangspiliot/AutoTheta`

4.1.2 Non-linear Trends

Nonetheless, an optimized θ may not guarantee a proper fit especially if the assumption of a linear long-term trend is violated (both the classical method and the OTM maintain the assumption that the first Theta-line is linear, e.g. $\theta_1 = 0$). For instance, an extrapolation of a linear trend will likely produce under forecast for an exponentially trended time series with increasing growth rate over time. The extension proposed by AutoTheta is to simply replace the linear line with an alternative curve:

$$\text{Linear trend: } Y_1(\theta_1 = 0) = \hat{\alpha} + \hat{\beta}t \tag{14}$$

$$\text{Exponential trend: } Y_1(\theta_1 = 0) = \hat{\alpha}e^{\hat{\beta}t} \text{ , or } Y_1(\theta_1 = 0) = log(\hat{\alpha}) + \hat{\beta}t \tag{15}$$

where $\hat{\alpha}$ and $\hat{\beta}$ are obtained via OLS over $log(X_t)$ against time.

4.1.3 Multiplicatively Related Forecasts

The classical Theta assumes an additive relation between the individual forecast and produce the final forecasts by taking a weighted average. In contrast with classical forecasting methods, such as the exponential smoothing, the classical Theta do no consider variations in the combinations of the level, trend and seasonal components. In other words, the level and trend component represented by the Theta-lines are assumed to be independent from each other and additively related. The same applies to the deseasonalization and reseasonalization steps which remains isolated in the classical method. This limitation makes the classical Theta relatively static.

To address this shortcoming, AutoTheta extends the decomposition approach in (5) with the following modification:

$$\text{Additive Theta-line: } Y_{add}(\theta) = \theta X_t + (1 - \theta)Y(\theta = 0) \tag{16}$$

$$\text{Multiplicative Theta-line: } Y_{mul}(\theta) = \frac{X_t^{\theta}}{Y(\theta = 0)^{\theta - 1}} \tag{17}$$

where X_t^{θ} denotes the θth power of the original or seasonally adjusted data.

The final forecasts can then be obtained as follows:

$$\text{Additive combination: } \hat{X}_{add,t}(\theta) = \big(1 - \frac{1}{\theta}\big)Y(\theta = 0) + \frac{1}{\theta}Y_{add,t}(\theta) \tag{18}$$

$$\text{Multiplicative combination: } \hat{X}_{mul,t}(\theta) = Y(\theta = 0)^{\left(1 - \frac{1}{\theta}\right)} * Y_{mul,t}(\theta)^{\frac{1}{\theta}} \tag{19}$$

where the Theta-lines, $Y(\theta)$, are extrapolated separately. Furthermore, the deasonalization and reseasonalization can be simply extended to handle additive seasonality.

4.1.4 Generalization to Eight Models

Following with the suggestions of 4.1.2 and 4.1.3, the AutoTheta framework results in a total of 12 different modelling permutations (refer to Table 4). This extension follows follows closely to the innovations state space models for exponential smoothing. Following the notations, the classical Theta is obtained with AutoTheta(M,L,A).

Notation	Seasonality	Trend	Combination
(N,L,A)	None	Linear	Additive
(N,L,M)	None	Linear	Multiplicative
(N,E,A)	None	Exponential	Additive
(N,E,M)	None	Exponential	Multiplicative
(A,L,A)	Additive	Linear	Additive
(A,L,M)	Additive	Linear	Multiplicative
(A,E,A)	Additive	Exponential	Additive
(A,E,M)	Additive	Exponential	Multiplicative
(M,L,A)	Multiplicative	Linear	Additive
(M,L,M)	Multiplicative	Linear	Multiplicative
(M,E,A)	Multiplicative	Exponential	Additive
(M,E,M)	Multiplicative	Exponential	Multiplicative

Table 4: Twelve possible models that can be generated

It should be noted that multiplicatively expressed models are constrained to data with only positive value. In addition, additive seasonal adjustment and linear trend factors may drive the forecasts below zero, even with only positive input values. Thus, multiplicative expressed models with additive seasonal adjustments and linear trends are likely to be difficult to compute. Specifically, (N,L,M), (A,L,M), (A,E,M) and (M,L,M) are ignored by AutoTheta, decreasing the number of modelling permutation to eight (refer to Table 5).

Notation	Seasonality	Trend	Combination
(N,L,A)	None	Linear	Additive
(N,E,A)	None	Exponential	Additive
(N,E,M)	None	Exponential	Multiplicative
(A,L,A)	Additive	Linear	Additive
(A,E,A)	Additive	Exponential	Additive
(M,L,A)	Multiplicative	Linear	Additive
(M,E,A)	Multiplicative	Exponential	Additive
(M,E,M)	Multiplicative	Exponential	Multiplicative

Table 5: Final eight models considered by AutoTheta

4.1.5 Model Selection

Choosing from a set of models is, of course, better than being restricted to just one. The advantage is a better generalization in dealing with a wider range of time series since some models perform better while other are worse for different series. In contrast, the classical Theta do not offer such flexibility. Naturally, a model selection criteria is needed to detect the best performing model. Given that AutoTheta's main objective is to generate forecast automatically, efficiency is prioritized. Thus, a computationally cheap model selection criterion is necessary.

Common practices include a goodness of fit measure, like the MAE or MSE, or an information, such as the AIC. The latter considers the complexity of models to produce a parsimonious model. In the case of AutoTheta, however, the use of information criteria may not add value since all models have identity complexity (the number of parameters estimated are fixed). As such, an error metric, like the MAE, may be more suited for evaluating the goodness of fit for each model. In addition, evaluating in-sample forecasting accuracy as compared to that of out-of-sample is less computation expensive and usually has lesser variation as number of out-of-samples observations may be limited. Thus, the AutoTheta uses the in-sample MAE to evaluate the goodness of fit for each model. For consistency, this criterion was selected to be the same as optimizing parameter θ as in section 4.1.1.

In short, for each time series, all possible models are estimated by AutoTheta before the MAE is applied to choose the most appropriate model for extrapolation. In the case of non-positive data, the pool of models is shrunk to exclude multiplicative an exponential models. Doing so speeds up the computation time. Lastly, a seasonality test is also conducted to further limit the choice of seasonal vs non-seasonal models.

4.2 Prediction Interval

In many applications, the uncertainty of the forecasts is as important as the point forecast itself. Uncertainty is usually quantified in terms of a prediction interval. Formally, given a significance level α, prediction intervals are designed such that $(1-alpha)\%$ of the future observations lie within the limits.

As the classical Theta can be expressed as a SES with drift, the prediction interval are given by:

$$\hat{X}_t(h) \pm q_{1-\alpha/2}\sigma\sqrt{(h-1)a^2+1} \tag{20}$$

where σ is the standard deviation of the residuals, a is the smoothing parameter of SES, and h is the forecast horizon.

However, (20) is not applicable to all models considered by AutoTheta since the interpretation of SES with drift is lost with different θ. AutoTheta considers a simplified formulation for i.i.d normal residuals (Makridakis et al., 1987), given as:

$$\hat{X}_t(h) \pm q_{1-\alpha/2}\sigma\sqrt{h} \tag{21}$$

AutoTheta builds on (21) by adjusting the spread of the prediction interval according to the frequency of the data as the forecasting uncertainty typically increases as the frequency of data decreases (Grushka-CockayneY & Jose, 2020). Intuitively, the forecasting uncertainty is expected to be greater for a one-year-ahead forecast as compared to that of a one-month-ahead forecast. Thus, the corresponding prediction intervals of the AutoTheta model is given by:

$$\hat{X}_t(h) \pm q_{1-\alpha/2}\sigma\sqrt{k(h-1)+1} \tag{22}$$

where k is the number of months in each period of the examined series. That is, one for monthly data, four for quarterly data and twelve for yearly data. Empirical observation from the M4-Competition also supports the results that uncertainty for yearly data tends to be underestimated as compared with quarterly and monthly series (Makridakis et al., 2020).

4.3 Performance

Spiliotis et al. (2020) evaluated the performance of the AutoTheta using times series from the M (Makridakis et al., 1982), M3 (Makridakis & Hibon, 2000) and M4-Competition (Makridakis et al., 2020). The consolidated dataset consisted of a total of 98,830 series (refer to Table 6 for details).

Frequency	M1	M2	M3	Total	Frequency
Yearly	181	645	23,000	23,826	6
Quarterly	203	756	24,000	24,959	8
Monthly	617	1,428	48,000	50,045	18
Total	1,001	2,829	95,000	98,830	-

Table 6: Data used for empirical evaluation by frequency and forecasting horizon

For each series, point forecasts and the 95% prediction intervals ($\alpha = 0.05$) were generated using ETS, ARIMA and the classical Theta. These three methods were used to benchmark against the AutoTheta.

The point forecasts are evaluated using the MASE, as defined in (13), and the prediction intervals are evaluated using three measures: the Mean Scaled Interval Score (MSIS), the coverage rate and the spread, which are defined as:

$$\text{MSIS} = \frac{1}{h}\frac{\sum_{t=n+1}^{n+h}(U_t - L_t) + \frac{2}{\alpha}(L_t - X_t)1\{X_t < L_t\} + \frac{2}{\alpha}(X_t - U_t)1\{X_t > U_t\}}{\frac{1}{n-m}\sum_{t=m+1}^{n}|X_t - X_{t-m}|}; \tag{23}$$

$$\text{Coverage} = \frac{1}{h}\sum_{t=n+1}^{n+h} 1\{X_t > L_t \,\&\, X_t < U_t\}; \tag{24}$$

$$\text{Spread} = \frac{\frac{1}{h}\sum_{t=n+1}^{n+h}(U_t - L_t)}{\frac{1}{n-m}\sum_{t=m+1}^{n}|X_t - X_{t-m}|} \tag{25}$$

where L_t and U_t are the lower and upper bounds of the prediction interval for period t respectively, and $1\{.\}$ is the indicator function. The MSIS takes the average of a summation of (1.) points where the future values are outside the specified bounds, and (2.) the width of the prediction interval and is then scaled by the mean absolute seasonal difference of the series. A lower MSIS indicates better prediction intervals. Coverage measures the percentage of times the true value lies within the prediction intervals while spread refers to the mean difference of the upper and lower bounds. Corresponding, a higher coverage and smaller spread is preferred.

Overall, the proposed AutoTheta algorithm outperforms its classical in terms of forecasting performance, with improvements ranging from 1% to 6% depending on the data frequency. AutoTheta also managed to outperform the ETS and ARIMA methods for the complete dataset (average of all series). However, when comparing scores across frequency, AutoTheta only managed to outperform for the yearly series with ETS and ARIMA providing slighly better forecast for the quarterly and monthly series (refer to Table 7 for more details).

Frequency	Theta	ETS	ARIMA	AutoTheta
Yearly	3.372	3.431	3.390	**3.175**
Quarterly	1.233	**1.165**	1.171	1.181
Monthly	0.968	0.947	**0.931**	0.958
Total	1.615	1.601	1.584	**1.549**

Table 7: Forecasting performance evaluated using MASE across data frequency

To test statistical significance of the achieved improvement, the MCB test was conducted (refer to section 3.3 for details on mechanism of the MCB test). Using the whole dataset, AutoTheta was the best-ranked performance but is not statistically different from that of the ARIMA. For yearly series, AutoTheta is statistically better than the other benchmarks. For monthly series, AutoTheta is only better than ETS and the classical Theta. For quarterly data, ETS and ARIMA provided better quality forecasts than AutoTheta. Nonetheless, AutoTheta is always significantly better than its classical form.

In terms of prediction intervals, AutoTheta performed better than its classical form across all data frequencies, as indicated by lower MSIS and coverage rate. Similar to the case of point forecasts, ETS and ARIMA achieved better prediction intervals for both quarterly and monthly series. The results also indicated that the prediction intervals by AutoTheta is wider than they should be as observed from relative larger spread across all frequencies. The results are replicated in Table 8.

Frequency	Theta	ETS	ARIMA	AutoTheta
		MSIS		
Yearly	54.579	34.970	45.102	**31.761**
Quarterly	12.277	**9.587**	11.240	11.057
Monthly	9.563	**8.258**	8.747	9.433
Total	21.101	**15.033**	18.141	15.226
		Coverage		
Yearly	68.997	83.842	72.487	**90.500**
Quarterly	84.753	92.913	86.261	**95.366**
Monthly	91.565	**93.286**	90.958	91.808
Total	84.404	90.915	85.319	**92.391**
		Spread		
Yearly	**7.777**	14.426	8.934	19.900
Quarterly	5.189	6.410	**4.611**	8.840
Monthly	5.698	5.215	**4.344**	5.637
Total	6.071	7.737	**5.518**	9.884

Table 8: Forecasting performance in terms of prediction intervals across data frequency

4.4 Evaluation

The results of the empirical study discussed in Section 4.3 observed AutoTheta achieving better forecasting performance then its classical form, for both point forecasts and prediction intervals. This indicates that the modification proposed were helpful in improving forecasting accuracy and predicts uncertainties better.

Overall, AutoTheta also managed to achieve better forecasting accuracy than well-known automated forecasting algorithms. This results support AutoTheta's capability in producing accuracy, robust and reliable predictions and thus, practising forecasters should consider adding AutoTheta to their existing collection of algorithms. In fact, AutoTheta, ETS and ARIMA each achieved best forecasting performance for yearly, quarter and monthly data respectively. This indicates that each method has its own strength in dealing with different types of time series and together, it represents diversity in terms of algorithm options. For instance, when dealing with different times series, it is always beneficial to assign each series with a specialized model, rather than a one-size-fit-all solution. In this aspect, the concept of combining forecasts from different models to produce a final forecast may be useful in improving forecasting accuracy too. In the field of machine learning, this strategy is often referred to as ensembling. In the simplest case, taking a simple average of diverse learners can often improve prediction performance as well as reducing prediction variance. This concept is evident in many well-known algorithm such as random forest or boosting tree where multiple weak learners are combined to produce a final prediction.

The study of such a proposal was also considered by Spiliotis et al. (2020). Specifically, the significance of the combination of all three algorithms were tested using the MCB test. The results indicated that a simple average of the forecasts by the three different algorithms leads to significat improves for both point forecasts and prediction intervals across all data frequencies. Thus, the contribution of the AutoTheta is the addition of a forecasting algorithm with its own special characteristics to existing methods. This increases the choice of algorithms a forecaster has under belt.

Finally, it should be noted that the AutoTheta comes with additional computational cost, measuring about 40 times expensive with the classical Theta. Yet, this increase in computation cost is relative reasonable since it is similar to that of ETS with ARIMA running 130 times slower.

5 Concluding Remarks

In its classical form, the Theta method is simply a decomposition approach which splits the original time series into two series with modified second differences. Being a simple and intuitive decomposition technique, the classical Theta method offers value in scenarios where forecasting procedures may have to be explained (e.g. consultant-client engagement). However, since the forecasts are ultimately extrapolated using existing methods, the effects of the decomposition approach on forecasting accuracy are ambiguous in its classical form.

Later, the empirical studies of the OTM confirm that the additional decomposition step has direct effect on forecasting accuracy. That is, under identical setups, different forecasting accuracy are obtained with different values of θ. Encouraging results from the OTM also revealed that optimizing θ has significant forecasting improvement as compared to the classical method. This indicates that carrying out the decomposition approach proposed before extrapolation can be beneficial to the forecasting accuracy. It was also noted that other extrapolation methods can be used with the Theta decomposition and not just SES or linear regression. Finally, the Theta method is generalised as an automatic forecasting algorithm with the capability of dealing with exponential trend and multiplicative relationship, named AutoTheta. AutoTheta managed to achieve significant improvement relative to its classical form and perform on par with popular algorithms, ETS and ARIMA. The observation that AutoTheta, ETS and ARIMA each work best with different data frequencies also indicates that AutoTheta is sufficiently diverse to complement existing work in the field of forecasting.

Moving forward, there are still various aspect of the Theta method that can be explored. Firstly, AutoTheta optimizes θ and selects model according to in-sample MAE. In this sense, it may be sub-optimal if out-of-sample forecasting accuracy is the main objective. Although the authors cited that in-sample evaluation offered more observations to estimate the long-term trend, evaluations based on in-sample often result in over-fitting. One modification could be the use of an validation scheme which mimic the required forecasting horizon to proxy the out-of-sample forecasting accuracy. Furthermore, different metrics can be consider depending on the type of data. For instance, if outliers are of concern, then the MSE could be considered. Secondly, the prediction interval formulated in (22) fixes k depending on the data frequency. This may limit its performance in dealing with the different characteristics of data of the same frequency. Similarly, a possible extension would be to estimate k using validation samples evaluated by metrics (23), (24) and (25). This way, the parameter k has more flexibility in fitting to the data based on its overall characteristics, rather than on its frequency only. Also, only two Theta-lines are considered with weights chosen such that its combination result in the original time series as in (18) and (19). A potential modification would be remove the constraint (9) altogether in order to decouple ω and θ. That way, multiple Theta-lines can be constructed by fixing different values of θ. The final forecasts can then be produced by assigning each Theta-line optimal weights found by minimizing the out-of-sample forecasting metric. This alternative approach may be more robust for time series with structural breaks in the interested forecasting horizon since the Theta-lines are not longer restricted by the in-sample time series (whose characteristic are of lesser interest).

To conclude, the *No Free Lunch Theorem* suggests that there is no one model that work best for every problem. In context, models that work well with some time series may not work well with other. In the field of forecasting, I believe that any model/method with the potential to improve forecasting accuracy should be explored. Thus, despite the lack of strong statistical theory, the Theta method has provided a new, useful tool for practising forecasters to deal with the increasingly data focused era.

References

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

Fioruci, J. A., Pellegrini, T. R., Louzada, F., & Petropoulos, F. (2015). *The optimised theta method*.*

Grushka-CockayneY, Y., & Jose, V. R. R. (2020). Combining prediction intervals in the M4 competition. *International Journal of Forecasting*, *36*, 178–185. doi: https://doi.org/10.1016/j.ijforecast.2019.04.015

Hyndman, R. J., & Billah, B. (2003). Unmasking the theta method. *International Journal of Forecasting*, *19*(2), 287–290. doi: https://doi.org/10.1016/S0169-2070(01)00143-1

Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Jornal of Forecasting*, *22*(4), 679–688. doi: https://doi.org/10.1016/j.ijforecast.2006.03.001

Makridakis, S., Andersen, A., Carbone, R., Fildes, R., Hibon, M., Lewandowski, R., ... Winkler, R. (1982). The accuracy of extrapolation (time series) methods: Results of a forecasting competition. *Journal of Forecasting*, *1*(2), 111-153. doi: 10.1002/for.3980010202

Makridakis, S., & Hibon, M. (2000). The M3-Competition: results, conclusions and implications. *International Journal of Forecasting*, *16*(4), 451–476. doi: https://doi.org/10.1016/S0169-2070(00)00057-1*

Makridakis, S., Hibon, M., Lusk, E., & Belhadjali, M. (1987). Confidence intervals: An empirical investigation of the series in the M-competition. *International Journal of Forecasting*, *3*(3-4), 489–508. doi: https://doi.org/10.1016/0169-2070(87)90045-8

Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2020). The M4 competition: 100,000 time series and 61 forecasting methods. *International Journal of Forecasting*, *36*(1), 54–74. doi: https://doi.org/10.1016/0169-2070(87)90045-8

Nikolopoulos, K., Assimakopoulos, V., Bougioukos, N., Litsa, A., & Petropoulos, F. (2011, 01). The theta model: An essential forecasting tool for supply chain planning. *Lecture Notes in Electrical Engineering*, *123*, 431-437. doi: 10.1007/978-3-642-25646-2_56

Spiliotis, E., & Assimakopoulos, V. (2018). *4theta: Generalizing the Theta method for automatic forecasting*.*

Spiliotis, E., Assimakopoulos, V., & Makridakis, S. (2020). Generalizing the Theta method for automatic forecasting. *European Journal of Operational Research*, *284*(2), 550–558. doi: https://doi.org/10.1016/j.ejor.2020.01.007*

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

Note: Main source text are indicated with * at the end of the reference.