Proceeding Paper

A Dynamic Combination of Theta Method and ATA: Validating on a Real Business Case †

Yasin Tadayonrad * and Alassane Ballé Ndiaye

Qalinca-Labs, Ecole Polytechnique, Université Libre de Bruxelles, 1050 Brussels, Belgium; alassane.ndiaye@ulb.be
* Correspondence: yasin.tadayonrad@ulb.be; Tel.: +32-494276007
† Presented at the 8th International Conference on Time Series and Forecasting, Gran Canaria, Spain, 27–30 June 2022.

Abstract: In order to make better decisions and take efficient actions in any supply chain system, we need to have better estimation of uncertain parameters, especially the future demands of our customers. To do so we must use a forecasting model which gives the most useful and accurate forecasts. Time series forecasting methods are still one of the most popular approaches used in the business because of their simplicity. One of the most recent methods that caught the attention of researchers and practitioners is Theta method, which was first ranked in M3 competition. This method works based on the decomposition of the deseasonalized original demand data into two components. The first component represents the long-term trend, and the second component indicates the short-term behavior of the data set. ATA method is another method which has been introduced recently. ATA method works like exponential smoothing methods, but in ATA method the smoothing parameter is a function of time point. In this paper we have proposed a new form of Theta method in which we have benefited from the features of ATA and presented a combination of ATA method and Theta method. We have introduced a dynamic model which uses Theta method as the main model and selected from among some alternative methods such as ATA method, simple exponential, and Double Exponential smoothing methods to be used as the theta lines. Also, we optimize the parameters of each method used in the model. Finally, we have tested the mentioned model on a real data set and concluded that the combination of Theta and ATA methods has a better performance compared to the other alternatives in terms of forecast accuracy.

Keywords: forecasting; time series; Theta method; ATA method; simple exponential smoothing (SES); double exponential smoothing (DES); forecasting key performance indicator (KPI); combination

1. Introduction

One of the major concerns in most supply chain systems is finding the best ways to help the managers and decision-makers to make right decisions and take efficient actions [1,2]. Estimating the customers' demand in the most accurate way is one of the most critical issues that affects many decisions in a supply chain system. Therefore, having an efficient and reliable forecasting model plays an important role in this industry. In the past decades, time series forecasting methods have been at the core of attention of researchers and practitioners in this field [3–7]. These methods are usually simple and understandable. That is why those models are still popular in the industry.

1.1. Theta Method

Theta method was introduced and ranked as the first method with the best performance in M3 competition [8]. After that, Theta method has been the benchmark method in the forecasting projects, competitions, and research [9]. The classic Theta method is applied to seasonally adjusted demand data set by decomposing the deseasonalized data into theta lines, extrapolating the theta lines, combining the extrapolated theta lines, and finally

seasonalizing the combined line to obtain the forecasts for periods ahead [10]. In the classic form, Theta decomposes the seasonally adjusted data into two theta lines. These two lines refer to long-term and short-term behavior of the original data set. The implementation of classic form of Theta method could be summarized in the following algorithm:

1.1.1. Seasonality Detection

This step can be ignored if our original data set rep-resents demand data in a yearly basis. In case of having quarterly, monthly, weekly, and even daily data we need to detect seasonality patterns in the historical data.

There are several ways for seasonality detection but one of the most common ways is using autocorrelation method. Using the below equations, we can find the existence of seasonality and the seasonal lag in our data.

$$r_k = \frac{\sum_{i=1}^{n-k}\left(d_i - \bar{d}\right)\left(d_{i+k} - \bar{d}\right)}{\sum_{i=1}^{n}\left(d_i - \bar{d}\right)^2} \tag{1}$$

$$|r_m| \geq q_{\left(1-\frac{\alpha}{2}\right)}\sqrt{\frac{1}{n}\left(1 + 2\sum_{k=1}^{m-1} r_k^2\right)} \tag{2}$$

r_k: Autocorrelation coefficient for lag k.
d_i: Values of original time series at time i.
m: The periodicity of the data (four for quarterly data, twelve for monthly data).
The value of q refers to the $(1-\alpha/2)\%$ confidence level in a Normal distribution
For any value of m that the absolute value if r_m is greater than the right side of Equation (2), we can say there is a seasonal pattern with lag m. In case of having more than one value for m, we can take the largest value.

1.1.2. Removal of Seasonality

One of the most well-known approaches for removing seasonality from the original data is decomposition methods. Classical multiplicative decomposition and classical additive decomposition are the most popular approaches to decompose seasonal patterns by which we can separate a time series into trend, seasonality, and remainder.

1.1.3. Theta Decomposition

In this step the seasonally adjusted series need to be separated into two components, short-term and long-term components. In Theta method θ is the main parameter that modifies the local curvatures of seasonally adjusted data. Using different values for θ results in different representation of the original data. $\theta = 1$ simply refers to the seasonally adjusted data. $\theta = 0$ refers to the series with no local curvatures, in other words it indicates the linear trend. $0 < \theta < 1$ results in reduction of the local curvatures and, as such, amplification of the long-term behavior. $\theta > 1$ results in amplification of the local curvatures and, as such, amplification of the short-term behavior.

In the classic version of Theta method, two theta lines are used, with $\theta 1 = 0$ and $\theta 2 = 2$, denoted as $t1$ and $t2$. As noted, $t1$ is the linear trend line of the data and corresponds to the long-term component, whereas $t2$ represents the series with double the local curvatures of the nonseasonal data.

The first theta line is $t1$. In the classic Theta method, simple regression is used to fit the nonseasonal data.

$$t1_t(0) = \hat{d} = b_0 + b_1 t \tag{3}$$

b_0: Intercept of seasonality adjusted data
b_1: Slope of seasonality adjusted data
t: Time period

The second theta line is $t2$.

$$t2_t(\theta) = \theta d_t + (1 - \theta)t1_t \tag{4}$$

with $\theta = 2$:

$$t2_t(2) = 2d_t - t1_t \tag{5}$$

1.1.4. Extrapolation

In the standard form, $t1$ is extrapolated by a normal linear regression line, while $t2$ is extrapolated using simple exponential smoothing. Extrapolation of the long-term component using normal linear regression line is as follow:

$$f1_{t+1} = b_0 + b_1(t + 1) \tag{6}$$

Extrapolation of the short-term component using SES is:

$$f2_{t+1} = \alpha t2_t + (1 - \alpha)f2_t \tag{7}$$

1.1.5. Combination

The final forecast is a combination of the forecasts of the two theta lines using equal weights. The forecasts of long-term and short-term theta lines should be combined (in the classic version, with equal weights; w_1, $w_2 = 0.5$).

$$f_{t+h} = 0.5\, f1_{t+h} + 0.5\, f2_{t+h} \tag{8}$$

For $\theta > 1$ we can use the following equation for optimizing the weights [11]:

$$w_2 = \frac{\theta_2 - 1}{\theta_2 - \theta_1}, \; w1 = 1 - w2 \tag{9}$$

Using the above formula, forecast could be obtained by:

$$f_{t+h} = w_1 f1_{t+h} + w_2 f2_{t+h} \tag{10}$$

1.1.6. Reseasonalization

If the series was identified as seasonal in step one, then the final forecasts are multiplied by the respective seasonal indices. Therefore, we must add/multiply the seasonality indices (calculated in step two) to the forecasts obtained in step five.

1.2. ATA Method

ATA method was introduced by a Turkish team in 2017 [12]. An alternative for two major forecasting approaches: exponential smoothing (ES) and autoregressive integrated moving average (ARIMA). The ATA method works by modifying the smoothing constants of various exponential smoothing models in a way that the smoothing constant becomes a function of t. This modification helps deal with initialization problems and makes the optimization process easier. Also, with ATA it is possible to assign past information equal weights. The ATA method has a similar form to ES, but the smoothing parameters are modified so that when obtaining a smoothed value at a specific time point, the number of observations that can contribute to the value being smoothed is taken into account when the weights among the observations are distributed. Therefore, the smoothing parameter for this method is a function of t, unlike ES where no matter where the value you are smoothing resides on the timeline, the observations receive weights depending only on their distances from the value being smoothed. The standard form of ATA method is given in the following aquations.

$$a_t = \left(\frac{p}{t}\right)d_t + \left(\frac{t-p}{t}\right)(a_{t-1} + b_{t-1}) \tag{11}$$

$$b_t = \left(\frac{q}{t}\right)(a_t - a_{t-1}) + \left(\frac{t-q}{t}\right)b_{t-1} \tag{12}$$

$$f_t(h) = a_t + h\,b_t \tag{13}$$

$p \in \{1, 2, \ldots, n\}, q \in \{1, 2, \ldots, p\}, t > p \geq q$
For $t \leq p$, $a_t = d_t$
For $t \leq q$, $b_t = d_t - d_{t-1}$
$b_1 = 0$

where:

d_t: the value of the original series
a_t: the smoothed value at time t
b_t: the trend component
p: the smoothing parameter for level
q: the smoothing parameter for trend
$f_t(h)$: the h-step-ahead forecast value for $h = 1, 2, \ldots$

To be able to tackle with dampening trends in the historical patterns, we can us the dampening factor in ATA method as follows:

$$a_t = \left(\frac{p}{t}\right)d_t + \left(\frac{t-p}{t}\right)(a_{t-1} + \varphi b_{t-1}) \tag{14}$$

$$b_t = \left(\frac{q}{t}\right)(a_t - a_{t-1}) + \left(\frac{t-q}{t}\right)\varphi b_{t-1} \tag{15}$$

$$f_t(h) = a_t + \left(\varphi + \varphi^2 + \ldots + \varphi^h\right)b_t \tag{16}$$

$p \in \{1, 2, \ldots, n\}, q \in \{1, 2, \ldots, p\}, \varphi \in (0, 1], t > p \geq q$
For $t \leq p$, $a_t = d_t$
For $t \leq q$, $b_t = d_t - d_{t-1}$
$b_1 = 0$

where:

d_t: the value of the original series
a_t: the smoothed value at time t
b_t: the trend component
p: the smoothing parameter for level
q: the smoothing parameter for trend
$f_t(h)$: the h-step-ahead forecast value for $h = 1, 2, \ldots$
φ: dampening parameter

The algorithm of implementing ATA method could be summarized as three main steps.

Step 1: Deseasonalize the data by the classical decomposition methods, if necessary.

Step 2: Arbitrary models were used for different data types.

Find an optimal value for the parameter p by minimizing the in-sample one-step-ahead sMAPE using following models and obtain forecast values as desired.

Yearly: ATAadd(p, 1, φ) where $\varphi \in \{0.61, 0.62, \ldots, 1\}$

All other types: (quarterly, monthly, weekly, daily, hourly)

ATAadd(p, 0, 1) and ATAadd(p, 1, 1) were obtained and averaged.

("add" refers to using additive decomposition method in the first step)

Step 3: The forecasts are reseasonalized using seasonal indices from classical decomposition.

2. Materials and Methods

In this work Theta method has been considered as the base model. In Theta method other models need to be used to extrapolate theta lines. As explained in the first section, there are two theta lines in the standard form of Theta method representing the long-term and short-term components of the demand. In the standard form the values of theta parameters are considered as 0 and 2 for the first and second theta lines, respectively. In the proposed model we tested the Theta method with both two and three theta lines. Regarding the mentioned unique features of ATA method and its ability to project the trend in the time series, ATA has been selected to extrapolate the first theta line which represents the long-term component. Therefore, in the third and fourth steps of Theta method, we use the formulas of ATA method (ATA and ATA with damped trend) instead of simple regression model. It means in the proposed method we use Equations (11)–(16) instead of Equations (3) and (6).

Furthermore, exponential smoothing methods (SES and DES methods as well as DES with damped trend) are used as the other candidates to extrapolate the other theta lines.

In order to find the best parameters that result in the most accurate forecasts, all parameters are optimized dynamically. As the demand range varies for different stock keeping units (SKUs), we need to attribute different importance to different products in order to pay more attention to the errors on those products with higher demand. To do so, weighted mean absolute error (wMAE) has been selected as a KPI to measure the performance of the forecasting methods in which we normalize the error based on the scale of demand of each product. Therefore, best values for the related parameters are chosen among the potential values for each model at each iteration (see Appendix A). The possible values that we have considered for each parameter have been listed in Table 1.

Table 1. Potential values for different parameters.

Parameter	Model(s)	Values
θ	Theta	{1.2, 1.4, 1.6, 1.8, 2, 2.2, 2.4, 2.6, 2.8, 3.0}
α	SES, DES	{0.05, 0.1, 0.2, 0.3, 0.5, 0.8, 0.9}
β	DES	{0.1, 0.2, 0.3, 0.8, 0.9}
φ	ATA and DES with damped trend	{0.6, 0.7, 0.8, 0.9}
n	MA	{3, 4, 5, 6, 7}
p	ATA, ATA with damped trend	{2, 3, 4, 5, 6, 7, 8, 9, 10}
q	ATA, ATA with damped trend	{2, ... , p}

3. Results

The given method was implemented on python NumPy library. A large demand dataset from a major international company was used to test the model. The mentioned dataset includes the demand data of 54 months for 19,691 SKUs in their logistics network in the Europe. We implemented the proposed model using NumPy and Pandas libraries on Python 3.7.0 and ran it on the given data set.

In Tables 2 and 3, we have presented the results obtained from different models compared to the results from applying MA and SES (as the benchmarks) on the given data set. We have optimized the value of n and α in MA and SES for each SKU. The first column shows the number of periods that we have considered as the test set. Indeed, we do not train the models on this data set and it remained hidden so as to test the performance of models. The second and third columns show the comparison of the results from classic Theta method in which the first theta line is the simple regression and the second (and third) theta line have been dynamically chosen among SES, DES, and DES with damped trend. The fourth and fifth columns represent the results from ATA and ATA with damped trend methods compared to that of MA. And finally, the last two columns indicate the results from the proposed method in which the first theta line has been fitted and extrapolated using ATA with damped trend method with two and three theta lines, respectively. As

we can see in the last row of the tables, the proposed method has a better performance compared to MA and SES as the benchmarks by 5.50% and 3.73%, respectively.

Table 2. Comparing the results obtained from applying different methods on the given data set with the results of the first benchmark (optimized MA).

Number of Test Periods	wMAE Theta2 vs. MA	wMAE Theta3 vs. MA	wMAE ATA vs. MA	wMAE ATA Damped vs. MA	wMAE Theta2 & ATA Damped vs. MA	wMAE Theta3 & ATA Damped vs. MA
8	6.06%	3.03%	7.58%	7.58%	9.09%	7.58%
10	−1.56%	6.25%	9.38%	9.38%	6.25%	3.13%
12	0.00%	−3.57%	−1.79%	0.00%	1.79%	0.00%
14	−1.67%	0.00%	3.33%	5.00%	3.33%	3.33%
18	3.51%	1.75%	1.75%	1.75%	7.02%	5.26%
Average	1.27%	1.49%	4.05%	4.74%	**5.50%**	3.86%

Table 3. Comparing the results obtained from applying different methods on the given data set with the results of the second benchmark (optimized SES).

Number of Test Periods	wMAE Theta2 vs. SES	wMAE Theta3 vs. SES	wMAE ATA vs. SES	wMAE ATA Damped vs. SES	wMAE Theta2 & ATA Damped vs. SES	wMAE Theta3 & ATA Damped vs. SES
8	0.00%	−3.23%	1.61%	1.61%	3.23%	1.61%
10	−4.84%	3.23%	6.45%	6.45%	3.23%	0.00%
12	1.75%	−1.75%	0.00%	1.75%	3.51%	1.75%
14	−1.67%	0.00%	3.33%	5.00%	3.33%	3.33%
18	1.79%	0.00%	0.00%	0.00%	5.36%	3.57%
Average	−0.59%	−0.35%	2.28%	2.96%	**3.73%**	2.05%

4. Conclusions

One of the main concerns of managers and decision-makers in supply chain systems is having accurate estimation of uncertain parameters. Using time series forecasting methods has been always one of the reliable alternatives to obtain the forecasts. In this paper we have benefited from the advantages of some the well-known methods and, combining them, we concluded that we can improve the accuracy of the forecasts using this combined model in which we have used ATA and ATA with damped trend as the first theta line. By testing the performance of other alternative methods (SES, DES, and DES with damped trend), the best method(s) is selected to extrapolate the other theta line(s). The model is also capable of finding the best values for each parameter aiming at minimizing wMAE. We implemented the proposed model on Python and tested it on a real data set. For the future works, we can compare the results from the proposed model with that of machine learning algorithms.

Appendix A

Algorithm A1. This algorithm represents the configuration of python code.

for *periods* in {8, 10, 12, 14, 18}:
    d_test = data_set [: , - periods:]
    d_train = data_set [: , : - periods]

    # 1. Theta classic with 2 theta lines
    t1 = regression
    for *t2* in [SES, DES, DES_damped]:
        for theta, *alpha, beta, phi* in thetas, alphas, betas, phis:
            f1 = theta2(d = d_train, t1 = regression, 2nd_theta_line = t2, theta = theta, alpha = alpha, beta = beta, phi = phi)
            kpi_train = wMAE(f1, d_test)
        select the best parameters based on minimum kpi
    select the best model for t2 based on minimum kpi
    f_theta2 = theta2(d=d_train, t1=regression, 2nd_theta_line=best_model, theta = best_theta, alpha = best_alpha, beta = best_beta, phi = best_phi)
    kpi_theta2 = wMAE(f_ theta2, d_test)

    # 2. Theta classic with 3 theta lines
    t1 = regression
    for *t2, t3* in [SES, DES, DES_damped]:
        for theta2, theta3, *alpha, beta, phi* in thetas, alphas, betas, phis:
            f2 = theta3(d = d_train, t1 = regression, 2nd_theta_line = t2, 3rd_theta_line = t3, theta2 = theta2, theta3 = theta3, alpha = alpha, beta = beta, phi = phi)
            kpi_train = wMAE(f2, d_test)
        select the best parameters based on minimum kpi
    select the best models for t2 and t3 based on minimum kpi
    f_theta3 = theta3(d = d_train, t1 = regression, 2nd_theta_line = best_model1, 3rd_theta_line = best_model2, theta2 = best_theta2, theta3 = best_theta3 alpha = best_alpha, beta = best_beta, phi = best_phi)
    kpi_theta3 = wMAE(f_ theta3, d_test)

    # 3. ATA
    for *p, q* in ps, qs:
        f3 = ata(d=d_train, p = p, q = q)
        kpi_train = wMAE(f3, d_test)
    select the best parameters based on minimum kpi
    f_ata = ata(d = d_train, p = best_p, q = best_q)
    kpi_ata = wMAE(f_ ata, d_test)

    # 4. ATA with damped trend
    for *p, q, phi* in ps, qs, phis:
        f4 = ata_damped(d = d_train, p = p, q = q, phi = phi)
        kpi_train = wMAE(f4, d_test)
    select the best parameters based on minimum kpi
    f_ata_damped = ata_damped(d = d_train, p = best_p, q = best_q, phi = best_phi)
    kpi_ata_damped = wMAE(f_ata_damped, d_test)

    # 5. Theta_ata with 2 theta lines
        t1 = ata
    for *t2* in [SES, DES, DES_damped]:
        for theta, *alpha, beta, phi, p, q* in thetas, alphas, betas, phis, ps, qs:
            f5 = theta2_ata(d = d_train, t1 = ata, 2nd_theta_line = t2, theta = theta, alpha = alpha, beta = beta, phi = phi)
            kpi_train = wMAE(f5, d_test)
        select the best parameters based on minimum kpi
    select the best model for t2 based on minimum kpi
    f_theta2_ata = theta2_ata(d = d_train, t1 = ata, 2nd_theta_line = best_model, theta = best_theta, alpha = best_alpha, beta = best_beta, phi = best_phi)
    kpi_theta2_ata = wMAE(f_ theta2_ata, d_test)

    # 6. Theta_ata with three theta lines
    t1 = ata
    for *t2, t3* in [SES, DES, DES_damped]:
        for theta2, theta3, *alpha, beta, phi, p, q* in alphas, betas, phis, ps, qs:
            f6 = theta3_ata(d = d_train, t1 = ata, 2^{nd}_theta_line = t2, 3^{rd}_theta_line = t3, theta2 = best_theta2, theta3 = best_theta3, alpha = alpha, beta = beta, phi = phi, p = p, q = q)
            kpi_train = wMAE(f6, d_test)
        select the best parameters based on minimum kpi
    select the best models for t2 and t3 based on minimum kpi
    f_theta3_ata = theta3_ata(d = d_train, t1 = ata, 2nd_theta_line = best_model1, 3rd_theta_line = best_model2, theta2 = best_theta2,   theta3 = best_theta3 alpha = best_alpha, beta = best_beta, phi = best_phi, p = best_p, q = best_q)
    kpi_theta3_ata = wMAE(f_ theta3_ata, d_test)

Algorithm A1. *Cont.*

```
for periods in {8, 10, 12, 14, 18}:
    d_test = data_set [ : , - periods: ]
    d_train = data_set [ : , : - periods]
```

```
# 7. Moving average
for n in ns:
    f7 = ma(d=d_train, n=n)
    kpi_train = wMAE(f7, d_test)
select the best parameters based on minimum kpi
f_ma = ma(d = d_train, n = best_n)
kpi_ma = wMAE(f_ma, d_test)
```

```
# 8. Simple exponential smoothing
for alpha in alphas:
    f8 = ses(d = d_train, alpha = alpha)
    kpi_train = wMAE(f8, d_test)
select the best parameters based on minimum kpi
f_ses = ses(d = d_train, alpha = best_alpha)
kpi_ses = wMAE(f_ses, d_test)
```

Output kpi_theta2, kpi_theta3, kpi_ata, kpi_ata_damped, kpi_theta2_ata, kpi_theta3_ata, kpi_ma, kpi_ses

References

1. Hosseini, S.; Ivanov, D.; Dolgui, A. Review of quantitative methods for supply chain resilience analysis. *Transp. Res. Part E Logist. Transp. Rev.* **2019**, *125*, 285–307. [CrossRef]
2. Chen, H.Y.; Das, A.; Ivanov, D. Building resilience and managing post-disruption supply chain recovery: Lessons from the information and communication technology industry. *Int. J. Inf. Manag.* **2019**, *49*, 330–342. [CrossRef]
3. De Gooijer, J.G.; Hyndman, R.J. 25 years of iif time series forecasting: A selective review. Tinbergen Institute Discussion Papers. *SSRN Electron. J.* **2005**, 5–68. [CrossRef]
4. Hyndman, R.; Koehler, A.B.; Ord, J.K.; Snyder, R.D. *Forecasting with Exponential Smoothing: The State Space Approach*; Springer: Berlin/Heidelberg, Germany, 2008.
5. Hyndman, R.J.; Koehler, A.B.; Snyder, R.D.; Grose, S. A state space framework for automatic forecasting using exponential smoothing methods. *Int. J. Forecast.* **2002**, *18*, 439–454. [CrossRef]
6. Gardner, E.S., Jr.; McKenzie, E. Forecasting trends in time series. *Manag. Sci.* **1985**, *31*, 1237–1246. [CrossRef]
7. Makridakis, S.; Andersen, A.; Carbone, R.; Fildes, R.; Hibon, M.; Lewandowski, R.; Newton, J.; Parzen, E.; Winkler, R. The accuracy of extrapolation (time series) methods: Results of a forecasting competition. *J. Forecast.* **1982**, *1*, 111–153. [CrossRef]
8. Koning, A.J.; Franses, P.H.; Hibon, M.; Stekler, H.O. The M3 competition: Statistical tests of the results. *Int. J. Forecast.* **2005**, *21*, 397–409. [CrossRef]
9. Athanasopoulos, G.; Hyndman, R.J.; Song, H.; Wu, D.C. The tourism forecasting competition. *Int. J. Forecast.* **2011**, *27*, 822–844. [CrossRef]
10. Yapar, G.; Yavuz, I.; Selamlar, H.T. Why and How Does Exponential Smoothing Fail? An In Depth Comparison of ATA-Simple and Simple Exponential Smoothing. *Turk. J. Forecast.* **2017**, *1*, 30–39.
11. Fiorucci, J.A.; Pellegrini, T.R.; Louzada, F.; Petropoulos, F.; Koehler, A.B. Models for optimising the theta method and their relationship to state space models. *Int. J. Forecast.* **2016**, *32*, 1151–1161. [CrossRef]
12. Petropoulos, F.; Nikolopoulos, K. The Theta Method. *Foresight* **2017**, *46*, 11–17.