ABSTRACT

The economic development in Indonesia is growing rapidly, competition is increasing. One of the things that are important in supporting the success of a company is to be able to meet the demands of various consumers. Activities that enable us to predict what will happen in the future are known as forecasting. Forecasting Theta methods, by Assimakopoulos Theta and Nikolopoulos. Initially, forecasting only use 2 theta weights parameters, then another theta parameter is added. Currently the Theta method use theta weight parameter. This research is to perform sales forecasting on car sales using Theta Method with five different weights using a system specially designed to perform the calculation. The error is minimized by minimizing the Mean Squared Error (MSE). As a result, using Theta weight 3 produces the smallest error.

Keywords: Theta method, forecasting, timeseries, Mean Squared Error

1. INTRODUCTION

The economic development in Indonesia is growing rapidly, competition is increasing. One of the things that are important in supporting the success of a company is to be able to meet the demands of various consumers. Activities that enable us to predict what will happen in the future are known as forecasting.

In addition to previously known forecasting methods, there is a method introduced by Assimakopoulos Theta and Nikolopoulos (2000). Initially, forecasting only use 2 theta weight parameters, later another theta parameter is added. Currently 5 theta weight parameter is used. By adding these weights, the error generated is smaller. This research is to apply the Theta Method to perform forecasting on car sales, minimizing the error using Mean Squared Error method.

1.1 Scope of the Research.

The scope of this research is as follows:
1. Forecasting method Theta.
2. The analyzed data are cars of type TBR, taking car sales from March 2008 until October 2008.
3. The method of forecasting accuracy is used to determine the theta method. The best theta is the one where the MSE (Mean Squared Error) is minimized.
4. Develop an application to determine the most appropriate method of forecasting. This application compares various methods. Using Theta by the number and weight of different theta and the best method will be obtained by considering the minimum MSE, and then calculates the value of forecasting.
5. The results of forecasting are only used as information for the company to make decisions.

This Research will discuss the following matters:
1. Analysis of factors - factors that affect the sales of cars such as the rate of inflation, the purchasing power of the people and other influences that directly or indirectly affect the sales.
2. Integration of applications created using systems running in the company.

1.2 Objectives

This research aims to find the best method to forecast sales using Theta by the value and theta different weights, and then the accuracy of forecast will be tested by minimizing MSE.

2. LITERATURE REVIEW

2.1 Theta Method

Design an application program that is used to compare Theta method with the value and weight of different theta.

Suppose \( \{X_1, \ldots, X_n\} \) is a univariate timeseries that is to be observed. From the series, a new series \( \{Y_1(\theta), \ldots, Y_n(\theta)\} \) is formed such that \( Y_t(\theta) = \theta X_t \) where \( X_t \) is the second difference of \( X_t \) and \( Y_t(\theta) \) is the second difference of \( Y_t(\theta) \) that has a solution as stated by (Box, 2013)

\[
Y_t(\theta) = a_0 + b_0 (t-1) + \theta X_t
\]

where \( a_0 \) and \( b_0 \) are constants. So \( Y_t(\theta) \) is equivalent to the linear function of \( X_t \) by adding a linear trend. Assimakopoulos and Nikolopoulos (2000) states that \( Y_t(\theta) \) is a “theta line” (Assimakopoulos, 2000).

For a fixed value of \( \theta \), given the value of \( Y_1(\theta) \) and \( Y_2(\theta) - Y_1(\theta) \) which simplifies to the sum of squared differences as follows:

\[
\sum_{i=1}^{t} \left[X_i - Y_t(\theta)\right]^2 = \sum_{i=1}^{t} \left[(1-\theta)X_i - a_0 - b_0 (t-1)\right]^2 \tag{2}
\]

This is equivalent to the sum of squares difference knowing \( a_0 \) and \( b_0 \). Therefore this become a simple regression mapping \( (1-\theta)X_i \) against time \( t-1 \) (Rob J. Hyndman, 2003). Thus the solution is:

\[
\hat{b}_{0,n} = \frac{6(1-\theta)}{n^2-1} \left(\frac{2}{n} \sum_{i=1}^{n} t X_i - (n+1) \bar{X}\right) \tag{3}
\]

and \( \hat{a}_{0,n} = (1-\theta)\bar{X} - \hat{b}_{0,n} (n-1)/2 \tag{4} \)

The fitted timeseries is as follows:

\[
\hat{Y}(\theta) = \hat{a}_{0,n} + \hat{b}_{0,n} (n-1)/2 + \theta \bar{X} = \bar{X} \tag{5}
\]

It is clearly seen that \( \frac{1}{2}[Y_t(1+p) + Y_t(1-p)] = X_t \) because \( \hat{a}_{1+p,n} + \hat{a}_{1-p,n} = 0 \) and \( \hat{b}_{1+p,n} + \hat{b}_{1-p,n} = 0 \).

The forecast using Theta method is obtained by calculating the average weight of \( Y_t(\theta) \) using different values of \( \theta \). Assimakopoulos and Nikolopoulos (2000) explained how to get the forecast value for \( \theta = 0 \) and \( \theta = 2 \). They defined the following:

\[
X_{n+h} = \frac{1}{2} \left[Y_{n+h}(0) + Y_{n+h}(2)\right] \tag{6}
\]

where : 
- \( X_{n+h} \) = forecast value (Ft)
- \( Y_{n+h}(0) \) = forecast value from theta weighted 0
- \( Y_{n+h}(2) \) = forecast value from theta weighted 2
\( Y_{n+h}(0) \) is obtained by extrapolating linearly from the equations (2-6) and \( Y_{n+h}(2) \) is obtained using simple exponential smoothing (SES) on the timeseries \( \{Y_i(2)\} \). Thus,

\[
Y_{n+h}(0) = \widehat{a}_{0,n} + \widehat{b}_{0,n} (n + h - 1)
\]

(7)

is obtained by extrapolating linearly from the equations (2-6) and \( Y_{n+h}(2) \) is obtained using simple exponential smoothing (SES) on the timeseries \( \{Y_i(2)\} \). Thus,

\[
Y_{n+h}(2) = \alpha \sum_{i=0}^{n-1} \left( 1 - \alpha \right)^i Y_{n-i}(2) + (1 - \alpha)^n Y_i(2)
\]

(8)

where \( \alpha \) is the coefficient of smoothing for SES.

From the above result, it can be combined to get a simple forecast for \( X_{n+h} \). From equations (2-8),

\[
Y_{n+h}(2) = \alpha \sum_{i=0}^{n-1} \left( 1 - \alpha \right)^i \left[ \widehat{a}_{2,n} + \widehat{b}_{2,n} (n - i - 1) + 2 X_{n-i} \right] + (1 - \alpha)^n \left( \hat{a}_{2,n} + 2 X_i \right)
\]

\[
= \widehat{a}_{2,n} + \widehat{b}_{2,n} \left[ n - \frac{1}{\alpha} + \frac{(1 - \alpha)^n}{\alpha} \right] + 2 X_{n+h}
\]

(9)

where \( X_{n+h} \) is the SES forecast for the timeseries \( \{X_i\} \). Given that \( \hat{a}_{2,n} = -\hat{a}_{0,n} \) and \( \hat{b}_{2,n} = -\hat{b}_{0,n} \),

\[
X_{n+h} = X_{n+h} + \frac{1}{2} \hat{b}_{0,n} \left( h - 1 + \frac{1}{\alpha} - \frac{(1 - \alpha)^n}{\alpha} \right)
\]

(10)

For large \( n \),

\[
X_{n+h} = X_{n+h} + \frac{1}{2} \hat{b}_{0,n} \left( h - 1 + 1/\alpha \right)
\]

(11)

Thus this is the SES by adding trend where the slope of the trend is half of the trend line which is adjusted to the original timeseries.

2.2 Durbin-Watson Statistics

The Durbin-Watson test statistics tests the hypothesis that there is no auto correlation on the residue. Durbin-Watson statistics is explained below: (Keller, 2014):

\[
D-W = \frac{\sum_{t=2}^{n} (e_t - e_{t-1})^2}{\sum_{i=1}^{n} e_i^2}
\]

(12)

where : \( e_t \) is \( X_t - F_t \)

The Durbin-Watson distribution is symmetric around 2, which is the middle value. Therefore the confidence level can formed involving five regions as shown in figure 1.

![Durbin-Watson Distribution](image)

Figure 1. Durbin-Watson distribution

(Makridakis, Wheelwright, & McGee, 1983)
The five intervals are (Keller, 2014):
1. Less than $D_L$ means there is a positive autocorrelation.
2. Between $D_L$ and $D_U$ means that it cannot be determined.
3. Between $D_U$ and $4 - D_U$ means there is no autocorrelation.
4. Between $4 - D_U$ and $4 - D_L$ means that it cannot be determined.
5. Greater than $4 - D_L$ means that there is a negative autocorrelation.

2.3 Mean Squared Error (MSE)

Makridakis, Wheelwright, and McGee uses several statistical standards to measure the accuracy of the forecast (Makridakis, Wheelwright, & McGee, 1983) The measurements shows the match between the model and the historical data. The comparison of the MSE value against forecast may provide indication of the accuracy of the model in the forecast:

$$e_i = X_i - F_i$$  \hspace{1cm} (13)

where: $e_t =$ error for period $t$.
$X_t =$ actual data for period $t$
$F_t =$ forecast for period $t$

If there is an observation and forecast for $n$ period of time, then there exists $n$ errors and the standard statistics measurement can be defined as Mean Squared Error,

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^{n} e_i^2$$  \hspace{1cm} (14)

2.4 File Input (Database)

In the program application, the input is obtained from Microsoft Excel (ekstensi .xls). The first two columns of this file is loaded. The first column contains the period and the second column contains the sales at the corresponding time. The data contain more than 30 rows such that a good forecast can be obtained.

2.4 Unified Modelling Language (UML)

UML is a standard language to create a blue print of software. UML can be used to visualize, determine and create documentation of the software design. There are several diagrams in UML:
1. Class Diagram to show the relationship among objects.
2. Object Diagram is the object and relation as a representation of the prototype.
3. Component Diagram is the components and relations that illustrate system implementation.
4. Deployment Diagram contains configuration from node and objects.
5. Use case Diagram is used to organize use cases and behaviors.
6. Sequence Diagram represents the time sequence of message and the object life line.
7. Collaboration Diagram represents the time sequence of message and the objects in the interactions.
8. Flow Diagram represents the work flow of the activities, focused on the operations between objects.
9. Activity Diagram is a diagram the represents the life cycle of the object in a transition from one state to the other state.

3.1 Research Data

Sales data is obtained from a car dealer, taking sales data for 30 weeks, from March 2008 until October 2008. The variable used is period ($t$), sales data ($X_t$) and forecast ($F_t$).

The hypothesis testing is using Theta Method with theta weight of 2, 3, 4 and 5 of which theta weight will provide the minimum MSE. The data has 30 periods, outliers is initially checked. When an outlier is detected, it will be corrected using Moving Average smoothing, followed by smoothing using theta weights and calculate the forecast where the weights has value 0 and 2 (for theta
weighted 0, forecast is performed using Simple Linear Regression, whereas for other forecast value use the simple exponential smoothing.

The result of the forecasting data, will be calculated using MSE in order to get the accuracy.

### 3.2 Application Design

The Use case diagram of the system can be seen in figure 2 below. The system performs three main functions, data correction, forecasting using manual theta weight and forecasting using the best theta weight.

![Use Case Diagram](image)

**Figure 2. Use Case Diagram of the System**

### 4 RESULTS AND DISCUSSION

#### 4.1 Results

The following table shows the sales data for 8 months.

<table>
<thead>
<tr>
<th>Period</th>
<th>Month</th>
<th>Week Number</th>
<th>Sales</th>
<th>Period</th>
<th>Month</th>
<th>Week Number</th>
<th>Sales</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>March</td>
<td>4</td>
<td>5</td>
<td>16</td>
<td></td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>2</td>
<td>April</td>
<td>1</td>
<td>5</td>
<td>17</td>
<td></td>
<td>4</td>
<td>13</td>
</tr>
<tr>
<td>3</td>
<td>April</td>
<td>2</td>
<td>5</td>
<td>18</td>
<td></td>
<td>5</td>
<td>13</td>
</tr>
<tr>
<td>4</td>
<td>April</td>
<td>3</td>
<td>8</td>
<td>19</td>
<td>August</td>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td>5</td>
<td>May</td>
<td>4</td>
<td>7</td>
<td>20</td>
<td></td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>May</td>
<td>1</td>
<td>11</td>
<td>21</td>
<td></td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>7</td>
<td>May</td>
<td>2</td>
<td>5</td>
<td>22</td>
<td></td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>8</td>
<td>May</td>
<td>3</td>
<td>13</td>
<td>23</td>
<td>September</td>
<td>1</td>
<td>14</td>
</tr>
<tr>
<td>9</td>
<td>June</td>
<td>4</td>
<td>6</td>
<td>24</td>
<td></td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>10</td>
<td>June</td>
<td>1</td>
<td>19</td>
<td>25</td>
<td></td>
<td>3</td>
<td>8</td>
</tr>
<tr>
<td>11</td>
<td>June</td>
<td>2</td>
<td>2</td>
<td>26</td>
<td></td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>12</td>
<td>June</td>
<td>3</td>
<td>2</td>
<td>27</td>
<td>October</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>13</td>
<td>July</td>
<td>4</td>
<td>8</td>
<td>28</td>
<td></td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>14</td>
<td>July</td>
<td>1</td>
<td>13</td>
<td>29</td>
<td></td>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>15</td>
<td>July</td>
<td>2</td>
<td>8</td>
<td>30</td>
<td></td>
<td>4</td>
<td>7</td>
</tr>
</tbody>
</table>


The sales data has maximum value 19 and minimum value 2. Standard Deviation is 3.9 and the mean is 7.5. Graphically, the trend is shown in figure 3.

![Sales Trend graph for 30 periods.](image1)

After performing correction, the trend can be seen in figure 4.

![Sales Trend graph after correction](image2)

### 4.2 Discussions

The forecasting method is using theta weighted 2, 3, 4 and 5. Each method will calculate the forecast using the best alpha for each and every methods. The use of alpha is based on experiments, starting from alpha = 0.1 until alpha = 0.9. The theta weight used is from -3.0 until 2.0, however, the weight 0 and 2 is definitely used because it serves as the standard for theta method and limited such that there are no duplicate weights used. The following shows the results in tabular and graphic forms.
Table 2. The comparison of using Alpha against the value from the Accuracy Method MSE.

<table>
<thead>
<tr>
<th>Alpha</th>
<th>2 theta</th>
<th>3 theta</th>
<th>4 theta</th>
<th>5 theta</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MSE</td>
<td>MSE</td>
<td>MSE</td>
<td>MSE</td>
</tr>
<tr>
<td>0.1</td>
<td>14.04292</td>
<td>12.13138</td>
<td>12.13420</td>
<td>12.13825</td>
</tr>
<tr>
<td>0.2</td>
<td>14.44100</td>
<td>12.16304</td>
<td>12.18197</td>
<td>12.18598</td>
</tr>
<tr>
<td>0.3</td>
<td>15.02405</td>
<td>12.32256</td>
<td>12.39456</td>
<td>12.39844</td>
</tr>
<tr>
<td>0.4</td>
<td>15.63675</td>
<td>12.57458</td>
<td>12.72433</td>
<td>12.72803</td>
</tr>
<tr>
<td>0.5</td>
<td>16.28638</td>
<td>12.88094</td>
<td>13.12333</td>
<td>13.12682</td>
</tr>
<tr>
<td>0.6</td>
<td>17.01492</td>
<td>13.21252</td>
<td>13.55568</td>
<td>13.55894</td>
</tr>
<tr>
<td>0.7</td>
<td>17.87514</td>
<td>13.55093</td>
<td>13.99294</td>
<td>14.00225</td>
</tr>
<tr>
<td>0.8</td>
<td>18.93384</td>
<td>13.88413</td>
<td>14.40105</td>
<td>14.44290</td>
</tr>
<tr>
<td>0.9</td>
<td>20.28449</td>
<td>14.20283</td>
<td>14.86855</td>
<td>14.87103</td>
</tr>
</tbody>
</table>

From 2 it can be seen that at alpha = 0.1 for different theta gives the smallest MSE value. Therefore the forecasting is set to use alpha = 0.1.

Table 3. Durbin-Watson Statistics

<table>
<thead>
<tr>
<th>Statistics</th>
<th>Forecasting method using alpha = 0.1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2 theta</td>
</tr>
<tr>
<td>D-W</td>
<td>1.73</td>
</tr>
</tbody>
</table>

With n=30, k=1 and 5% level of significance, the value $d_L = 1.34$ and $d_U = 1.49$ and $4 - d_U = 2.51$, where the value of $d_U < D-W < 4 - d_U$. Therefore it is concluded that the error is random.

4.3 Hypothesis Testing

Table 3. Comparison of Forecast Accuracy

<table>
<thead>
<tr>
<th>Accuracy Method</th>
<th>Forecasting Method with alpha = 0.1</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSE</td>
<td>2 theta</td>
</tr>
<tr>
<td></td>
<td>14.04292</td>
</tr>
</tbody>
</table>

It shows that theta weight 3 produces the smallest error. Although theta weight 4 and 5 also shows a small error. Theta weight 2 shows a significantly larger error. Thus, theta weight 3 can be used as input in making decision.

5. CONCLUSION

Based on the result and the comparison of forecasting Theta with weight 2, 3, 4 and 5, it concluded that:

[a]. The best method to use for forecasting the car sales of type TBR is Theta Method with Theta weight 3. This method however has a weakness that it can only predict for one period in the future, therefore need the actual sales to predict the next period.
[b]. The result using Forecasting Theta Method is able to follow the actual pattern.
[c]. Forecasting using Theta Method has a weakness in which this method is not able to explain short term patterns.
[d]. From Durbin-Watson test and graphs it can be concluded that the error has a random behavior, which means that the
error does not have a residue after the forecasting model is applied.

6. RECOMMENDATION

This method can serve as a reference in making decisions on sales prediction. As a recommendation, researches using different weights can be performed and different method for finding the accuracy can be selected.

7. REFERENCES


   b. Akses : 10 Agustus 2008


