

Time series Forecasting using Holt-Winters Exponential Smoothing

Prajakta S. Kalekar(04329008)
Kanwal Rekhi School of Information Technology

Under the guidance of
Prof. Bernard

December 6, 2004

Abstract

Many industrial time series exhibit seasonal behavior, such as demand for apparel or toys. Consequently, seasonal forecasting problems are of considerable importance. This report concentrates on the analysis of seasonal time series data using Holt-Winters exponential smoothing methods. Two models discussed here are the *Multiplicative Seasonal Model* and the *Additive Seasonal Model*.

1 Introduction

Forecasting involves making projections about future performance on the basis of historical and current data.

When the result of an action is of consequence, but cannot be known in advance with precision, forecasting may reduce decision risk by supplying additional information about the possible outcome.

Once data have been captured for the time series to be forecasted, the analyst's next step is to select a model for forecasting. Various statistical and graphic techniques may be useful to the analyst in the selection process. The best place to start with any time series forecasting analysis is to graph sequence plots of the time series to be forecasted. A sequence plot is a graph of the data series values, usually on the vertical axis, against time usually on the horizontal axis. The purpose of the sequence plot is to give the analyst a visual impression of the nature of the time series. This visual impression should suggest to the analyst whether there are certain behavioral "components" present within the time series. Some of these components such as *trend* and *seasonality* are discussed later in the report. The presence/absence of such components can help the analyst in selecting the model with the potential to produce the best forecasts.

After selecting a model, the next step is its specification. The process of specifying a forecasting model involves selecting the variables to be included, selecting the form of the equation of relationship, and estimating the values of the parameters in that equation.

After the model is specified, its performance characteristics should be verified or validated by comparison of its forecasts with historical data for the process it was designed to forecast. Error measures such as MAPE¹, RAE², MSE³ maybe used for validating the model. Selection of an error

¹Mean absolute percentage error

²Relative absolute error

³Mean square error

measure has an important effect on the conclusions about which of a set of forecasting methods is most accurate

Time-series forecasting assumes that a time series is a combination of a pattern and some random error. The goal is to separate the pattern from the error by understanding the pattern's trend, its long-term increase or decrease, and its seasonality, the change caused by seasonal factors such as fluctuations in use and demand.

Several methods of time series forecasting are available such as the Moving Averages method, Linear Regression with Time, Exponential Smoothing etc. This report concentrates on the Holt-Winters Exponential Smoothing technique as applied to time series that exhibit seasonality.

2 Basic Terminology

The following keywords are used throughout the report:

2.1 Additive and multiplicative seasonality

Often, time series data display behavior that is seasonal. Seasonality is defined to be the tendency of time-series data to exhibit behavior that repeats itself every L periods. The term season is used to represent the period of time before behavior begins to repeat itself. L is therefore the season length in periods. For example, annual sales of toys will probably peak in the months of November and December, and perhaps during the summer (with a much smaller peak). This pattern is likely to repeat every year, however, the relative amount of increase in sales during December may slowly change from year to year.

For example, during the month of December the sales for a particular toy may increase by 1 million dollars every year. Thus, we could add to our forecasts for every December the amount of 1 million dollars (over the respective annual average) to account for this seasonal fluctuation. In this case, the seasonality is additive.

Alternatively, during the month of December the sales for a particular toy may increase by 40%, that is, increase by a factor of 1.4. Thus, when the sales for the toy are generally weak, then the absolute (dollar) increase in sales during December will be relatively weak (but the percentage will be constant); if the sales of the toy are strong, then the absolute (dollar) increase in sales will be proportionately greater. Again, in this case the sales increase by a certain factor, and the seasonal component is thus multiplicative in nature (i.e., the multiplicative seasonal component in this case would be 1.4).

In plots of the series, the distinguishing characteristic between these two types of seasonal components is that in the additive case, the series shows steady seasonal fluctuations, regardless of the overall level of the series; in the multiplicative case, the size of the seasonal fluctuations vary, depending on the overall level of the series.

2.2 Linear, exponential, and damped trend

To continue with the toy example above, the sales for a toy can show a *linear upward trend* (e.g., each year, sales increase by 1 million dollars), *exponential growth* (e.g., each year, sales increase by a factor of 1.3), or a *damped trend* (during the first year sales increase by 1 million dollars; during the second year the increase is only 80% over the previous year, i.e., \$800,000; during the next year it is again 80% over the previous year, i.e., $\$800,000 * .8 = \$640,000$; etc.).

2.3 Seasonality Index (SI)

Seasonality Index of a period indicates how much this period typically deviates from the annual average. At least one full season of data is required for computation of SI.

2.4 Stationarity

To perform forecasting, most techniques require the stationarity conditions to be satisfied.

- **First Order Stationary**

A time series is a first order stationary if expected value of $X(t)$ remains same for all t . For example in economic time series, a process is first order stationary when we remove any kinds of trend by some mechanisms such as differencing.

- **Second Order Stationary**

A time series is a second order stationary if it is first order stationary and covariance between $X(t)$ and $X(s)$ is function of length $(t-s)$ only. Again, in economic time series, a process is second order stationary when we stabilize also its variance by some kind of transformations, such as taking square root.

3 Exponential smoothing

Exponential smoothing is a procedure for continually revising a forecast in the light of more recent experience. Exponential Smoothing assigns exponentially decreasing weights as the observation get older. In other words, recent observations are given relatively more weight in forecasting than the older observations.

3.1 Single exponential smoothing

This is also known as simple exponential smoothing. Simple smoothing is used for short-range forecasting, usually just one month into the future. The model assumes that the data fluctuates around a reasonably stable mean (no trend or consistent pattern of growth).

The specific formula for simple exponential smoothing is:

$$S_t = \alpha * X_t + (1 - \alpha) * S_{t-1} \quad (1)$$

When applied recursively to each successive observation in the series, each new smoothed value (forecast) is computed as the weighted average of the current observation and the previous smoothed observation; the previous smoothed observation was computed in turn from the previous observed value and the smoothed value before the previous observation, and so on.

Thus, in effect, each smoothed value is the weighted average of the previous observations, where the weights decrease exponentially depending on the value of parameter (α) . If it is equal to 1 (one) then the previous observations are ignored entirely; if it is equal to 0 (zero), then the current observation is ignored entirely, and the smoothed value consists entirely of the previous smoothed value (which in turn is computed from the smoothed observation before it, and so on; thus all smoothed values will be equal to the initial smoothed value S_0). In-between values will produce intermediate results.

Initial Value

The initial value of S_t plays an important role in computing all the subsequent values. Setting it to y_1 is one method of initialization. Another possibility would be to average the first four or five observations.

The smaller the value of α , the more important is the selection of the initial value of S_t .

3.2 Double Exponential Smoothing

This method is used when the data shows a trend. Exponential smoothing with a trend works much like simple smoothing except that two components must be updated each period - *level* and

trend. The level is a smoothed estimate of the value of the data at the end of each period. The trend is a smoothed estimate of average growth at the end of each period. The specific formula for simple exponential smoothing is:

$$S_t = \alpha * y_t + (1 - \alpha) * (S_{t-1} + b_{t-1}) \quad 0 < \alpha < 1 \quad (2)$$

$$b_t = \gamma * (S_t - S_{t-1}) + (1 - \gamma) * b_{t-1} \quad 0 < \gamma < 1 \quad (3)$$

Note that the current value of the series is used to calculate its smoothed value replacement in double exponential smoothing.

Initial Values

There are several methods to choose the initial values for S_t and b_t .

S_1 is in general set to y_1 .

Three suggestions for b_1

$$b_1 = y_2 - y_1$$

$$b_1 = [(y_2 - y_1) + (y_3 - y_2) + (y_4 - y_3)]/3$$

$$b_1 = (y_n - y_1)/(n - 1)$$

3.3 Triple Exponential Smoothing

This method is used when the data shows *trend* and *seasonality*. To handle seasonality, we have to add a third parameter. We now introduce a third equation to take care of seasonality. The resulting set of equations is called the "Holt-Winters" (HW) method after the names of the inventors. There are two main HW models, depending on the type of seasonality (Section 2.1).

- **Multiplicative Seasonal Model**
- **Additive Seasonal Model**

The rest of the report focuses on these two models.

4 Multiplicative Seasonal Model

This model is used when the data exhibits Multiplicative seasonality as discussed in Section 2.1.

4.1 Overview

In this model, we assume that the time series is represented by the model

$$y_t = (b_1 + b_2 t) S_t + \epsilon_t \quad (4)$$

Where

b_1 is the base signal also called the *permanent component*

b_2 is a linear trend component

S_t is a multiplicative seasonal factor

ϵ_t is the random error component

Let the length of the season be L periods.

The seasonal factors are defined so that they sum to the length of the season, i.e.

$$\sum_{1 \leq t \leq L} S_t = L \quad (5)$$

The trend component b_2 if deemed unnecessary, maybe deleted from the model.

4.2 Application of the model

The multiplicative seasonal model is appropriate for a time series in which the amplitude of the seasonal pattern is proportional to the average level of the series, i.e. a time series displaying multiplicative seasonality. (Section 2.1)

4.3 Details

This section describes the forecasting equations used in the model along with the initial values to be used for the parameters.

4.3.1 Notation used

Let the *current deseasonalized level* of the process at the end of period T be denoted by R_T .

At the end of a time period t , let

\bar{R}_t be the estimate of the deseasonalized level.

\bar{G}_t be the estimate of the trend

\bar{S}_t be the estimate of seasonal component (seasonal index)

4.3.2 Procedure for updating the estimates of model parameters

1. Overall smoothing

$$\bar{R}_t = \alpha \frac{y_t}{\bar{S}_{t-L}} + (1 - \alpha) * (\bar{R}_{t-1} + \bar{G}_{t-1}) \quad (6)$$

where $0 < \alpha < 1$ is a smoothing constant.

Dividing y_t by \bar{S}_{t-L} , which is the seasonal factor for period T computed one season (L periods) ago, *deseasonalizes* the data so that only the trend component and the prior value of the permanent component enter into the updating process for \bar{R}_T .

2. Smoothing of the trend factor

$$\bar{G}_t = \beta * (\bar{S}_t - \bar{S}_{t-1}) + (1 - \beta) * \bar{G}_{t-1} \quad (7)$$

where $0 < \beta < 1$ is a second smoothing constant.

The estimate of the trend component is simply the smoothed difference between two successive estimates of the deseasonalized level.

3. Smoothing of the seasonal index

$$\bar{S}_t = \gamma * (y_t / \bar{S}_t) + (1 - \gamma) * \bar{S}_{t-L} \quad (8)$$

where $0 < \gamma < 1$ is the third smoothing constant.

The estimate of the seasonal component is a combination of the most recently observed seasonal factor given by the demand y_t divided by the deseasonalized series level estimate R_t and the previous best seasonal factor estimate for this time period. Since seasonal factors represent deviations above and below the average, the average of any L consecutive seasonal factors should always be 1. Thus, after estimating S_t , it is good practice to re normalize the L most recent seasonal factors such that

$$\sum_{i=t-L+1}^t S_i = q$$

4.3.3 Value of forecast

1. *Forecast for the next period*

The forecast for the next period is given by:

$$y_t = (\bar{R}_{t-1} + \bar{G}_{t-1})\bar{S}_{t-L} \quad (9)$$

Note that the best estimate of the seasonal factor for this time period in the season is used, which was last updated L periods ago.

2. *Multiple-step-ahead forecasts (for $T < q$)*

The value of forecast T periods hence is given by:

$$y_{t+T} = (\bar{R}_{t-1} + T * \bar{G}_{t-1})\bar{S}_{t+T-L} \quad (10)$$

4.3.4 Initial values of model parameters

As a rule of thumb, a minimum of two full seasons (or $2L$ periods) of historical data is needed to initialize a set of seasonal factors. Suppose data from m seasons are available and let $\bar{x}_j, j = 1, 2, \dots, mL$ denote the average of the observations during the j^{th} season.

1. **Estimation of trend component**

Estimate the trend component by:

$$\bar{G}_0 = \frac{\bar{y}_m - \bar{y}_1}{(m-1)L} \quad (11)$$

2. **Estimation of deseasonalized level**

Estimate the deseasonalized level by:

$$\bar{R}_0 = \bar{x}_1 - \frac{L}{2}\bar{G}_0 \quad (12)$$

3. **Estimation of seasonal components**

Seasonal factors are computed for each time period $t = 1, 2, \dots, mL$ as the ratio of actual observation to the average seasonally adjusted value for that season, further adjusted by the trend; that is,

$$\bar{S}_t = \frac{\bar{x}_t}{\bar{x}_i - [(L+1)/2 - j]\bar{G}_0} \quad (13)$$

where \bar{x}_i is the average for the season corresponding to the t index, and j is the position of the the period t within the season. The above equation will produce m estimates of the seasonal factor for each period.

$$\bar{S}_t = \frac{1}{m} \sum_{k=0}^{m-1} \bar{S}_{t+kL} \quad t = 1, 2, \dots, L \quad (14)$$

$$\bar{S}_t(0) = \bar{S}_t \frac{L}{\sum_{t=1}^L \bar{S}_t} \quad t = 1, 2, \dots, L \quad (15)$$

5 Additive Seasonal Model

This model is used when the data exhibits Additive seasonality as discussed in Section 2.1.

5.1 Overview

In this model, we assume that the time series is represented by the model

$$y_t = b_1 + b_2t + S_t + \epsilon_t \quad (16)$$

Where

b_1 is the base signal also called the *permanent component*

b_2 is a linear trend component

S_t is a additive seasonal factor

ϵ_t is the random error component

Let the length of the season be L periods.

The seasonal factors are defined so that they sum to the length of the season, i.e.

$$\sum_{1 \leq t \leq L} S_t = 0 \quad (17)$$

The trend component b_2 if deemed unnecessary, maybe deleted from the model.

5.2 Application of the model

The additive seasonal model is appropriate for a time series in which the amplitude of the seasonal pattern is independent of the average level of the series, i.e. a time series displaying additive seasonality. (Section 2.1)

5.3 Details

This section describes the forecasting equations used in the model along with the initial values to be used for the parameters.

5.3.1 Notation used

Let the *current deseasonalized level* of the process at the end of period T be denoted by R_T .

At the end of a time period t , let

\bar{R}_t be the estimate of the deseasonalized level.

\bar{G}_t be the estimate of the trend

\bar{S}_t be the estimate of seasonal component (seasonal index)

5.3.2 Procedure for updating the estimates of model parameters

1. Overall smoothing

$$\bar{R}_t = \alpha(y_t - \bar{S}_{t-L}) + (1 - \alpha) * (\bar{R}_{t-1} + \bar{G}_{t-1}) \quad (18)$$

where $0 < \alpha < 1$ is a smoothing constant.

Dividing y_t by \bar{S}_{t-L} , which is the seasonal factor for period T computed one season (L periods) ago, *deseasonalizes* the data so that only the trend component and the prior value of the permanent component enter into the updating process for \bar{R}_T

2. Smoothing of the trend factor

$$\bar{G}_t = \beta * (\bar{S}_t - \bar{S}_{t-1}) + (1 - \beta) * \bar{G}_{t-1} \quad (19)$$

where $0 < \beta < 1$ is a second smoothing constant.

The estimate of the trend component is simply the smoothed difference between two successive estimates of the deseasonalized level.

3. Smoothing of the seasonal index

$$\bar{S}_t = \gamma * (y_t - \bar{S}_t) + (1 - \gamma) * \bar{S}_{t-L} \quad (20)$$

where $0 < \gamma < 1$ is the third smoothing constant.

The estimate of the seasonal component is a combination of the most recently observed seasonal factor given by the demand y_t divided by the deseasonalized series level estimate \bar{R}_t and the previous best seasonal factor estimate for this time period. Since seasonal factors represent deviations above and below the average, the average of any L consecutive seasonal factors should always be 1. Thus, after estimating S_t , it is good practice to re normalize the L most recent seasonal factors such that

$$\sum_{i=t-q+1}^t S_i = q$$

5.3.3 Value of forecast

The forecast for the next period is given by:

$$y_t = \bar{R}_{t-1} + \bar{G}_{t-1} + \bar{S}_{t-L} \quad (21)$$

Note that the best estimate of the seasonal factor for this time period in the season is used, which was last updated L periods ago.

6 Experimental Results

Variations of the additive and multiplicative exponential smoothing techniques were applied to some standard time series data, the results of which are discussed in this section.

6.1 Test cases

The following tests were carried out:

- Multiplicative model
Two variations of the model are used:

1. Non-adaptive

In the non-adaptive technique, data of the first two years is used for building the model and establishing the parameters. Using these parameters, forecasts are made for the remaining data. Hence, once the values of the parameters α , β and γ are initialized, they are not modified later. The criterion used for selection of the parameters is *MAPE*.

- The advantage of this model is that the parameters are initialized only once. Hence, once the parameters have been established, the forecasting can proceed without any delay in re-computation of the parameters.
- Another advantage is that past data need not be remembered.

This method is typically suited for series where the parameters remain more or less constant over a period of time.

2. Adaptive

In the adaptive technique, the original parameters are established using the data of the first two years. However, the model keeps adapting itself to the changes in the underlying process. The parameters are adjusted after every two years. The newly computed parameters may be computed using

- All the data available till that point of time or
- Only k most recent data values.

The first approach requires all past data to be stored.

- Additive model

There are two variations of the model used. These are the same as that of the Multiplicative model, i.e. Adaptive and Non-Adaptive.

For a given time series, it is possible that the value of L ⁴ maybe unknown. For such a time series, the value of L was varied between 6 and 24.

6.2 Summary of Test results

A summary of test results are presented in this section. The tests were conducted on the following time series : *d1*, *ebrahim*, *rose*, *red-wine* and *paper*.

The multiplicative model gave better results as compared to the additive model for all the series.

For most of the series like *d1*, the adaptive multiplicative model gave better results as compared to the non-adaptive multiplicative model. For other series like *red-wine*, the non-adaptive multiplicative model gave better results.

Taking *d1* and *red-wine* as examples, the results are summarized below.

6.2.1 d1

The results obtained by using Multiplicative Adaptive technique are summarized in Table 1

L	Look back	MAPE
12	24	142.84827
12	∞	129.24403
13	26	113.98809
13	∞	111.99094
15	30	114.87889
15	∞	119.090744
17	34	120.50779
17	∞	115.598305
18	36	120.56211
18	∞	117.90184
23	46	123.13986
23	∞	109.389824

Table 1: Adaptive Multiplicative Technique for d1

The best result value of MAPE is obtained for value of $L = 23$, with an infinite look back. The MAPE for the non-adaptive technique is 10431.5, whereas that for adaptive is 109. This suggests

⁴Seasonality period

that the series is constantly changing. Consequently, the model parameters need to be constantly updated according to the newer data.

Figure 1 shows the comparison between the real and the forecasted values for $L = 23$ with an infinite look back.

With the best value of MAPE being 109, exponential smoothing technique does not appear to be the best forecasting technique for this data.

6.3 red-wine

In case of this time series, the multiplicative non-adaptive exponential smoothing technique gave better results as compared to the adaptive version of the same. The value of MAPE obtained was 9.01809 with $\alpha = 0.2$, $\beta = 0.1$ and $\gamma = 0.1$.

A summary of the results of Multiplicative Adaptive technique on the time series is shown in Table 2.

L	Look back	MAPE
10	20	21.62986
10	∞	22.011831
12	24	18.493525
12	∞	18.515306
24	48	17.48018
24	∞	17.722464

Table 2: Adaptive Multiplicative technique for red-wine

The results suggest that the periodic re-computation of the parameters may not necessarily improve the results. The time series may exhibit certain fluctuations. If the parameters are re-computed periodically, then it is possible that new set of parameters are affected by the fluctuations.

Figure 2 shows the comparison between the real and the forecasted values for the multiplicative non-adaptive technique with $\alpha = 0.2$, $\beta = 0.1$ and $\gamma = 0.1$.

7 Conclusion

The Holt-Winters exponential smoothing is used when the data exhibits both trend and seasonality. The two main HW models are *Additive model* for time series exhibiting additive seasonality and *Multiplicative model* for time series exhibiting Multiplicative seasonality.

Model parameters, α , β and γ are initialized using the data of the first two years. The error measure used for selecting the best parameters is MAPE. In the adaptive HW technique, these parameters are constantly updated in light of the most recently observed data. The motivation behind using the adaptive technique, as opposed to the non-adaptive technique is that, the time series may change its behavior and the model parameters should adapt to this change in behavior. Tests carried out on some standard time series data corroborated this assumption.

The value of L and the Look back size also play an important role in the performance of the adaptive model, hence these parameters were also varied for forecasting the different time series. It was observed that the adaptive technique with a larger look back in general improved the results (as compared to the non-adaptive version). However, adaptive technique with a smaller look back often performed worse than the non-adaptive technique. This suggests that re-computing the model parameters on the basis of only few of the most recent observations may be an unnecessary overhead which may lead to poor performance. Hence, the adaptive technique should be combined with a sufficiently large look back size in order to obtain good results.

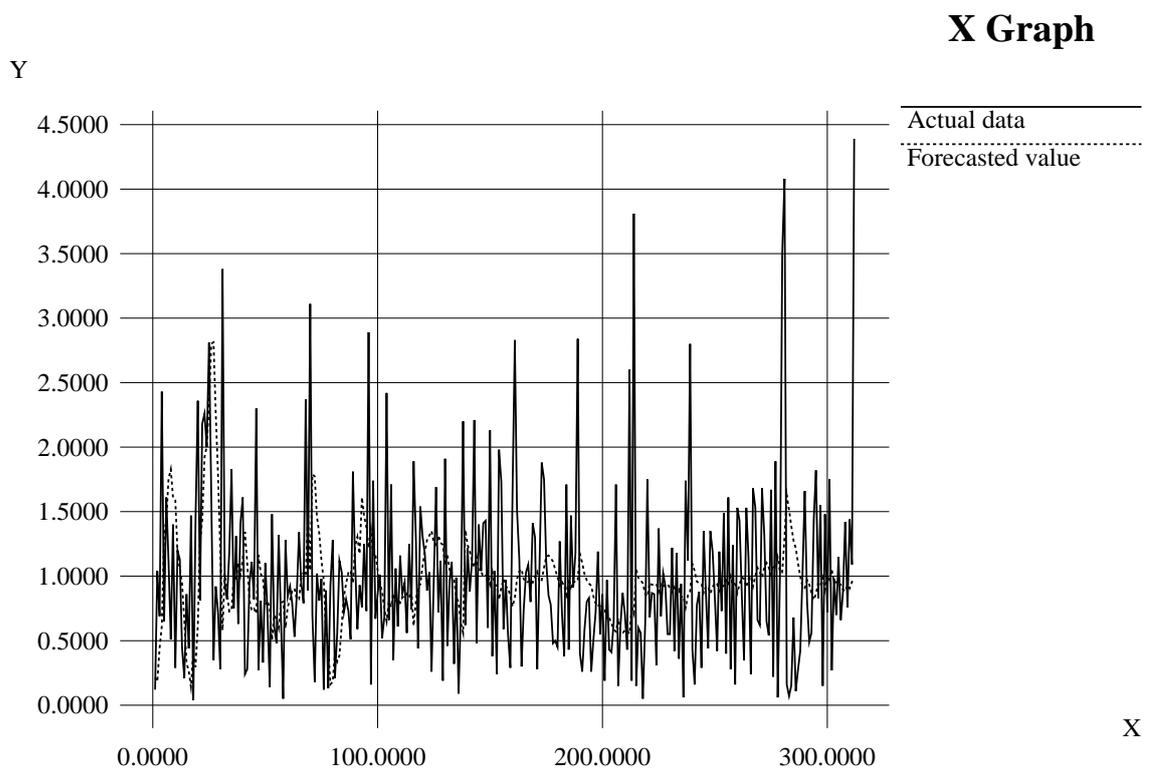


Figure 1: Multiplicative adaptive technique for d1 with $L = 23$

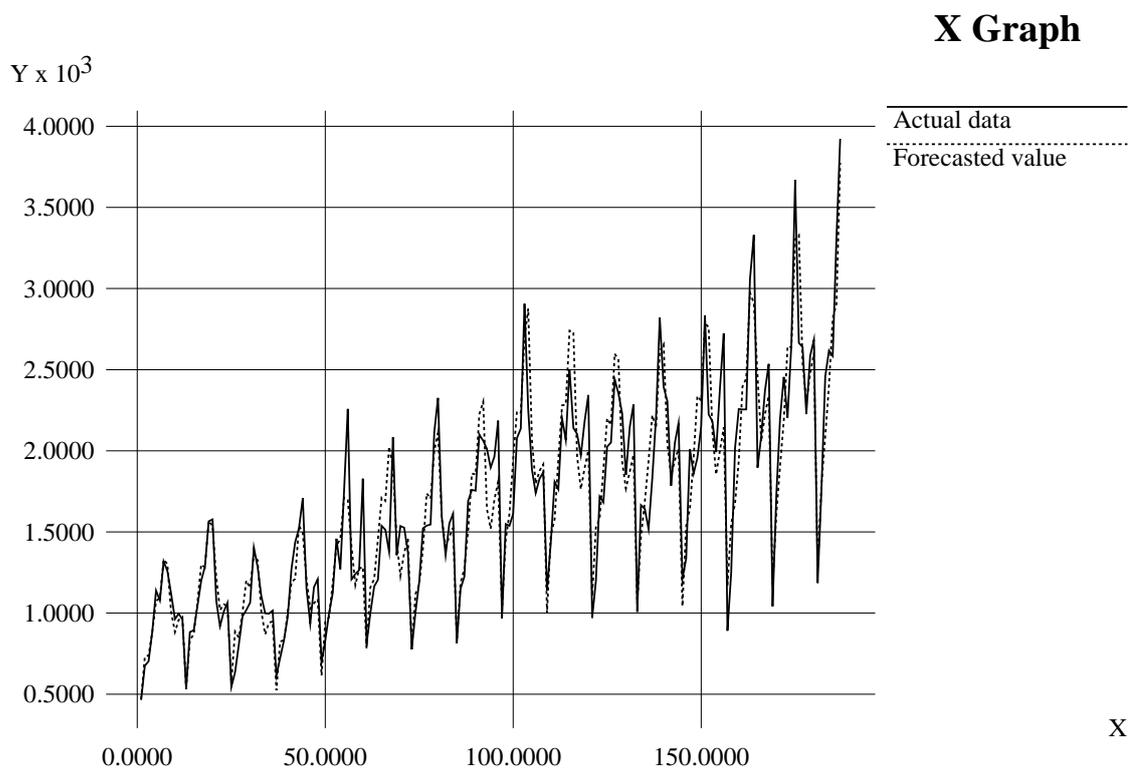


Figure 2: Multiplicative non-adaptive technique for red-wine with $\alpha = 0.2$, $\beta = 0.1$ and $\gamma = 0.1$

References

- [1] Introduction to time series analysis. <http://www.itl.nist.gov/div898/handbook/pmc/section4/pmc4.htm>.
- [2] Lynwood A. Johnson Douglas C. Montgomery and John S. Gardiner. *Forecasting and Time Series Analysis*. McGraw-Hill,Inc, 2nd edition edition, 1990.