



Table of Contents

[**Executive Summary** 2](#_30j0zll)

[**The Forecasting Problem** 3](#_1fob9te)

[**Examining Data Patterns** 4](#_3znysh7)

[**Analysis of Seasonality** 5](#_2et92p0)

[**Trend Analysis** 6](#_tyjcwt)

[**Analysis of Cycle** 7](#_3dy6vkm)

[**Analysis of Fitted Forecast** 8](#_1t3h5sf)

[**Forecast for rest of 2004 - Period (March ‘04 to December ‘04)** 9](#_4d34og8)

[Box Jenkins 9](#_2s8eyo1)

[Regression 9](#_17dp8vu)

[**Evaluation of Forecast Accuracy** 10](#_3rdcrjn)

[**Conclusion and Recommendations** 11](#_26in1rg)

# **Executive Summary**

Kim Brite and Larry Short have developed a series of exclusive mobile-home parks. Each site occupied by these parks provides privacy and a pleasant living environment to its occupants. They want to open more such facilities.

In our project, we conducted a thorough analysis of mobile-home shipments/sales in the United States to forecast for the 4 quarters of 2004 based on detailed 16 years’ quarterly data collected from the period 1988 quarter 1 through 2003 quarter 4. This forecast would help Kim and Larry manage cash flow for opening more such facilities. We have utilized the application “Forecast Pro” to get actionable insights to our data, and thus, created accurate and credible forecasts.

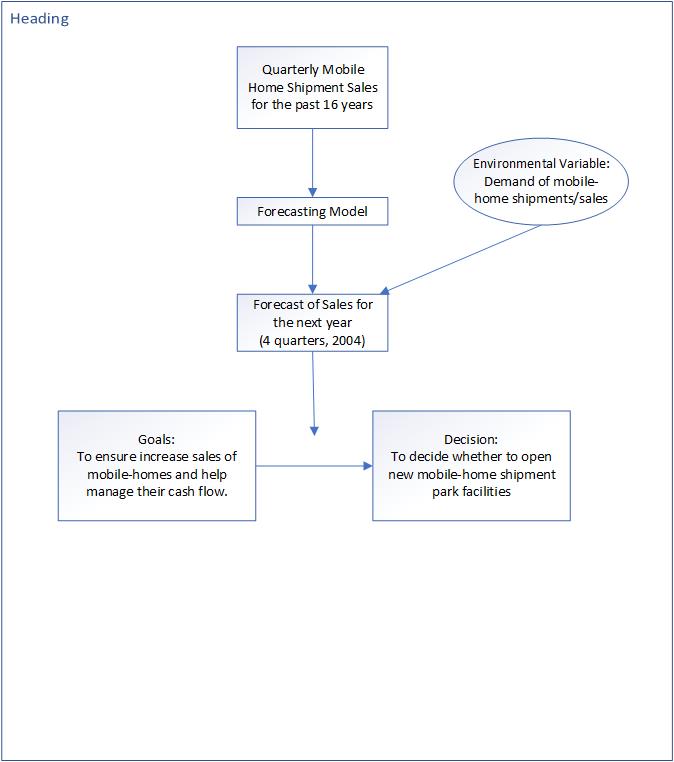
We utilized multiplicative classical time series decomposition to forecast the data. We examined the data patterns within our time series for better understanding the data and produce accurate forecasts. Differencing both seasonal and simple were implemented to identify if there is any trend or seasonality in our data. Once seasonality was identified, seasonal indexes were studied to determine the nature of seasonality. If producing more such mobile home shipment units would be profitable for Kim and Larry, we developed a long-term linear trend for the data and analyzed it properly.

Further, we used Box-Jenkins and Regression approaches to forecast the cyclical factors and thus produce our forecasts based on the observations. Finally, we used MAPE to measure the accuracy of our forecast model.

Based on the observations received from our forecast model, we were able to achieve our goal and thus, reach our final decision for producing more mobile home shipment units. Besides, we provided recommendations based on our model which would prove fruitful for Kim and Larry.

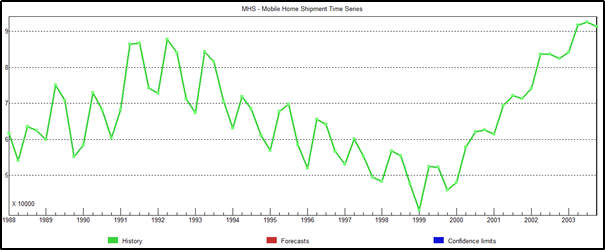
# **The Forecasting Problem**

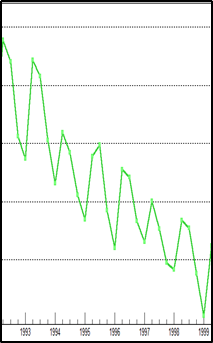
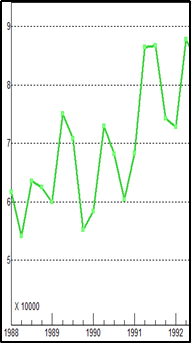
The problem which Kim Brite and Larry Short are facing is with the exclusive Mobile Home Shipments (MHS) data where they need us to forecast sales of mobile home shipments for the 4 quarters of 2004 to manage their cash flow. The current MHS appears to impact the vacancy rates and the momentum at which they can fill newly opened parks. The objective of this report is to provide a decision for and recommendations to Kim and Larry based on analysis of the time series pattern for the MHS data by addressing the trend, seasonality and cyclicity and hence, provide effective forecast according to our forecast model. With the help of this data, we would assist Kim & Larry to decide whether to open new mobile-home shipment park facilities or not. While recommending, we need to consider the external environmental factors such as the US economy and demand of mobile-home shipments/sales. The roadmap for our forecasting analysis of MHS is shown in figure 1(number all figures in the document in the end). This graphical framework also illustrates how our decision making is related to forecasting.

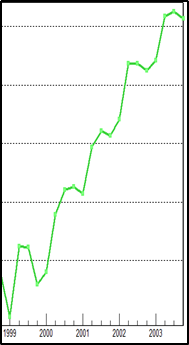


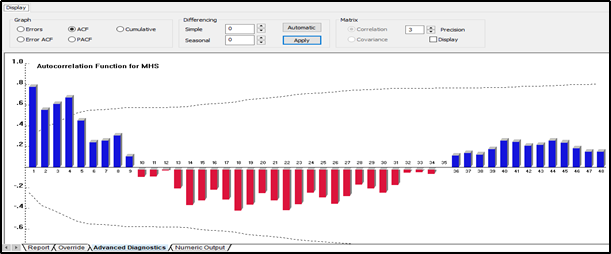
# **Examining Data Patterns**

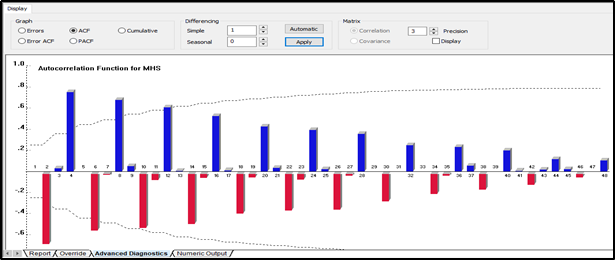
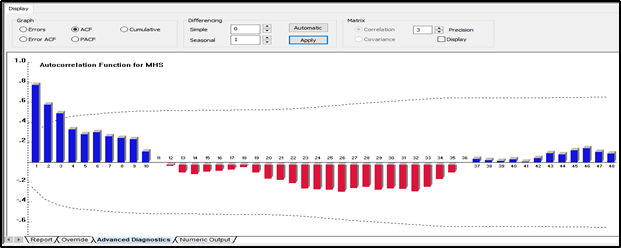
We have set forecast horizon to 0 in order to better visualize the time series. Forecast horizon is the length of time into the future for which forecasts are to be prepared. Setting forecast horizon to 0 indicates that we are currently understanding and exploring only the historical time series data which will help us in forecasting. The below graph depicts the time series for our dataset beginning from Q1 of 1988 to Q4 of 2003.

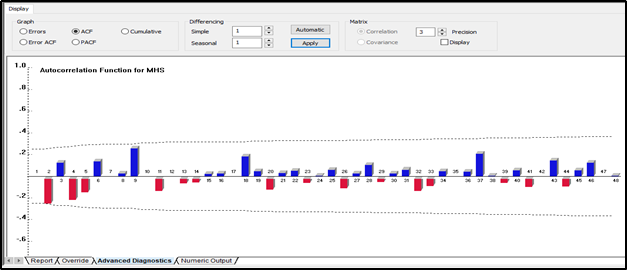
  
   
Below we have divided the series into 3 parts showing an increase and then decrease and then again an increase pattern in the series. The below visualization helps us better understand our data and thus, the series obtained from it.



 these 3 pictures are coming in one row in my downloaded document so nothing to worry about it. Once we all finish it, i’ll download the document and format it along with altering the front page.  
From the above graphs, we determine that there is an increase during quarter 2 of every year (upward peaks) and a decrease during quarter 1 of every year (downward peaks) and a similar pattern is followed throughout the time series. Moreover, the graph shows a linear increase for 5 years from Q1 1988 to Q2 of 1992 and then a linear decrease for 2.5 years up to Q4 1999 with again an increase for 3 years from Q1 2000 to Q3 2003. Additionally, the time series is not constant over mean and variance.  
  
These findings illustrate that our data is seasonal, has cyclical pattern, is non-stationary, and might be trended. Seasonality is supported by high rise during Q2s’ and depressions at Q1s’ annually. Also, since the seasonal influences increase or decrease proportionally with increases and decreases in the level of the series, we can infer that time series have multiplicative seasonality. Besides, above time series has trend since have linear increase for 5 years then decrease for 2.5 years followed by increase again. Now, since these rises and fall are not of fixed period, it has cyclical pattern. Non-Stationarity is depicted from the fact that the series is not constant over mean and variance.   
Determining patterns in time series using Autocorrelation Analysis  
The autocorrelation coefficients are plotted to demonstrate the autocorrelation function, ACF. The plot obtained by autocorrelation is the correlogram.  
Further, we utilized Autocorrelation Analysis without differencing, with first order simple differencing, with first order seasonal differencing, and both first order simple and seasonal differencing in order to confirm on the data patterns which we observed above by simply visualizing the time series graph.  
  
**No differencing applied**From the below correlogram, we examine that successive observations are highly correlated since the autocorrelation coefficients are significantly different from 0 for the first several time lags and then gradually drop toward zero as the number of lags increases. As a result, we can infer that the series might have a trend. Furthermore, the autocorrelation coefficient gradually drops to 0 instead of rapidly dropping to 0 indicates that the series is non-stationary.

  
   
**First - order Simple differencing applied**We removed the trend component from the series and can observe that there pertains a strong seasonality indicated by the rise or peaks at lags 4, 8, 12, 16, 20, and so on and depressions at lags 2, 6, 10, 14, 18 which occurs due to high Q2 and low Q1 respectively which is repeated annually across each Q2 and Q1. A significant autocorrelation coefficient (ACF) occurs in lags of 4 quarters.

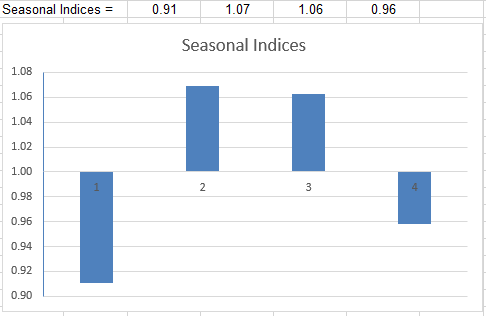
  
   
**First - order Seasonal differencing applied**We now removed the seasonality from the series to confirm the existence or non-existence of trend. We observe a similar pattern to what we observed from the correlogram with no differencing that successive observations are highly correlated since the autocorrelation coefficients are significantly different from 0 for the first several time lags and then gradually drop toward zero as the number of lags increases. This confirms the trend within the series. We also observe that this trend is of different periods, hence, the series is cyclic.  
  
 **First - order Simple and Seasonal both differencing applied**  
By applying both simple and seasonal differencing, we see below that the series is becoming stationary with constant mean and variance. Since few forecasting methods require the data to be stationary, we made the series stationary.



# **Analysis of Seasonality**

Almost all the organizations are interested in knowing how their sales performance on a seasonal basis compares to its normal variation. In **figure 2**, we can see sharp seasonal peaks and troughs when plotted from the MHS data from 1988 Q1 through 2003 Q3. Approximately all the peaks are occuring at Q2 and nearly all the troughs are observed at Q1 which implies that seasonality might exist in the data. However, this is not enough evidence to prove the seasonality.

Seasonal index is a measure of how a particular season in MHS through some cycle compares with the average season of that cycle of MHS. Applying seasonal indices will deseasonalize mobile homes data and, thereby, smooth it, to allow for forecasting of trends. Using deseasonalization, we are eliminating seasonal alterations in mobile-home sales or patterns. As a result, it would help us predict or approximate future demand.

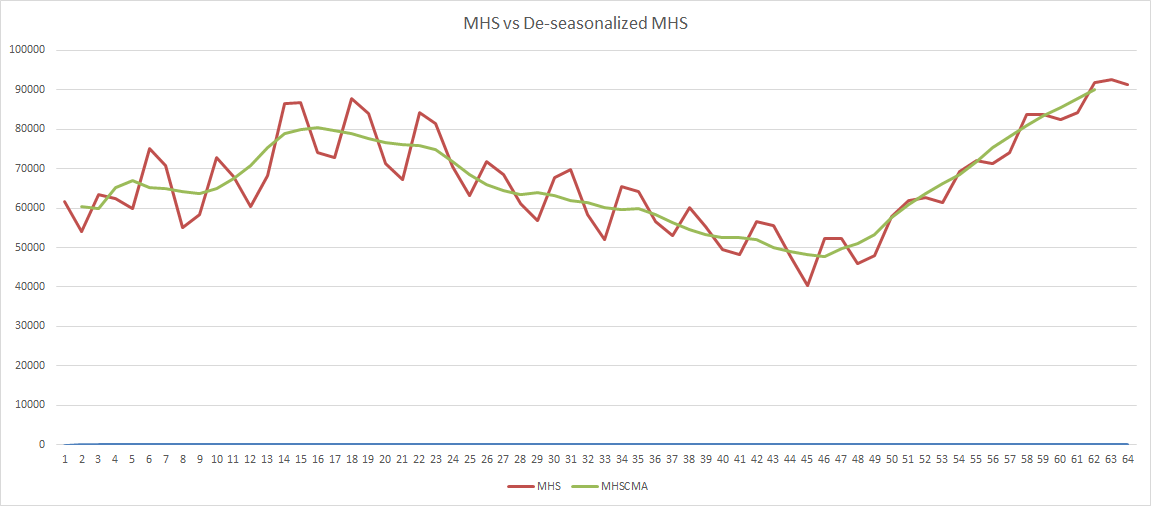
****

After seasonal indices are calculated, we need to measure seasonality in MHS by comparing the actual value of the MHS data with the equivalent deseasonalized value. Hence, the seasonality factor is calculated. Seasonality variations have been removed by using centered moving average (CMA).

There are four quarters and it's a multiplicative model so the sum of seasonal factors should sum up to 4. From **figure 7**, the total of seasonal indices calculated on the MHS data is 4. Therefore, no adjustment is required to the calculated indices.

As depicted in **figure 7**, for Q1, Q2, Q3, and Q4 the seasonal indices are 0.91, 1.07, 1.06, and 0.96 respectively. From this analysis, we can follow that Q2 and Q3 have relatively outperformed the seasonal average by 7% and 6%. However, Q1 and Q4 seem to have relatively underperformed the seasonal average by 9% and 4%.

We plotted a graph, **figure 8**, with deseasonalized sales data and superimposed it onto the original time plot for sales. It resembles that this deseasonalized time plot could be extremely helpful in forecasting future trends, as the rise and fall of Mobile Home Shipments may be attributed to the increase and decline in economy.



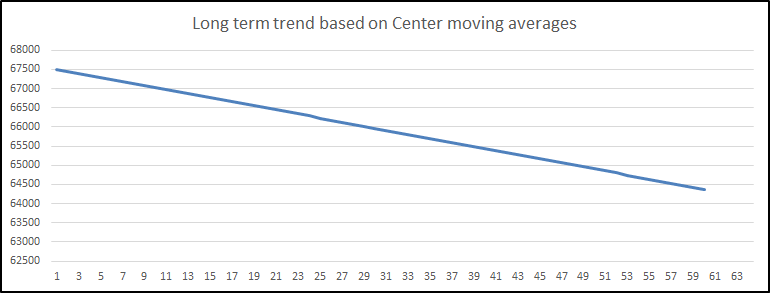
MHS vs Deseasonalized Data

# **Trend Analysis**

In our case, trend analysis is a useful way to look at past sales of mobile home shipment units and determine possible trends from that data and use the information to extrapolate what could happen in the future.

The below graph is obtained by regressing dependent variable I.e., the centered moving averages for mobile home shipments data by the independent variable I.e., time. We observe that the plot has a linear decreasing trend. This is indicative of the fact that the sales of mobile home shipment units could even reduce in future which would not prove to be a good factor for Kim and Larry.

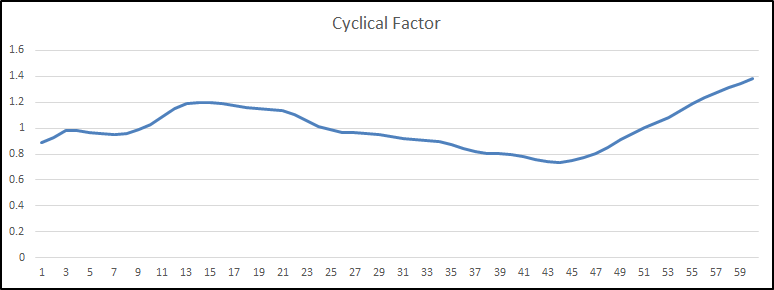
**Note:** All calculations for the purpose are done on deseasonalized data.

****

# **Analysis of Cycle**

The cyclical component of a time series refers to (regular or periodic) fluctuations around the trend, excluding the irregular component, revealing a succession of phases of expansion and contraction. *(QECD Glossary,* [*https://stats.oecd.org/glossary/detail.asp?ID=6694*](https://stats.oecd.org/glossary/detail.asp?ID=6694)*)*

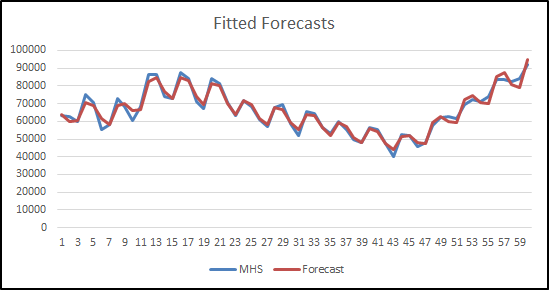
In the section Fig -- where we analyzed the long-term trend, we could see that the trend data showed negative linearity with respect to the time. In this section, we analyze the impact of the cyclical factor on our historical data. The plot in Fig -- shows the variation of the cyclical factor with the change in time. The cyclic factor shows increases and decreases in the values, which continue due to the cycle experienced by the time series observations. When we take a look at the plot for the last few observations, it shows a strong increase in the cycle, which indicates that if we remove the seasonality from the data, the MHS values obtained may be greater than the determined long-term trend values shown in the Fig --.

**X-SERIES SHOULD START WITH Q3 THAN Q1. CHANGE THE WORDS OR GRAPH**

# **Analysis of Fitted Forecast**

In this section we plotted the observed forecast values versus the actual MHS values. The red line depicts the forecasted values while the blue one are the actual values. For analyzing each parameter of the time series, we calculated the seasonal indexes values, long term trend values, cyclical factor values. We calculated the product of these values to obtain the forecast values shown in the Figure number below.

We can see that both the forecasted and actual are almost same which demonstrates a good fit within our data. Therefore, our forecast follows a historical pattern. The forecast using the decomposition of time series gives a great fitted model which shows that the decomposition method gives an approximately accurate forecast of the time series for the mobile home shipments data.

****

# **The Forecast**

Forecasting is a method that utilizes historical information as input to make educated price quotes that are predictive in identifying the instructions of future patterns. Organizations use forecasting to figure out how to designate their budget plans or prepare for unexpected expenditures for an amount of time. The forecasting of MHS is constructed using 16 years quarterly historical data. From the data pattern analysis, we can perceive a presence of trend, seasonality and cycle factor. The forecast is created by multiplying all three components along with the irregular component.

The Cyclic Factor (CF) is composed of seasonal and residual components and has a huge impact on the forecasts since it help define the path of industrial situation. Therefore, it is the most crucial and difficult component of the time series.

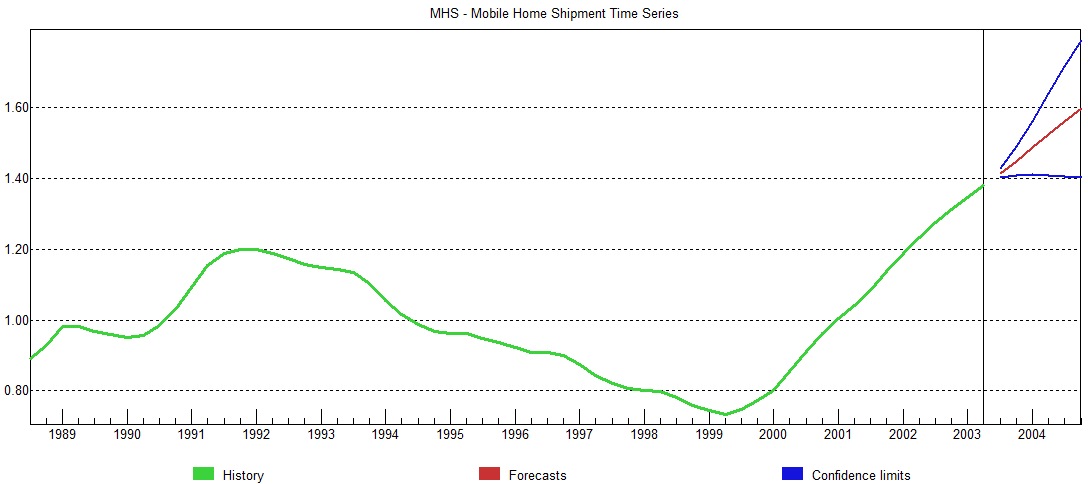
As illustrated in section “Analysis of Seasonality”, the calculated seasonal indices for Quarter 1, Quarter 2, Quarter 3, and Quarter 4 are 0.91, 1.07, 1.06, 0.96 respectively.

In our case the requirement is to forecast the cyclic factor (CF) for the year 2004 using the approaches mentioned below. These CF values obtained will be used in the calculation of the forecast values.

* Box Jenkins
* Regression

## Box Jenkins

The forecast for Cycle Factor (CF) for 2004 is done by utilizing the calculated CF values in the section “Analysis of Cycle”. We used the CF values calculated until Q2 of 2003 to forecast 6 future quarters to cover 2004. First, we created the excel format of the CF values to be used by the Forecast Pro software. Then, we applied the Box-Jenkins forecasting method while forcing a constant to generate a forecast from Q3 2003 to Q4 2004.



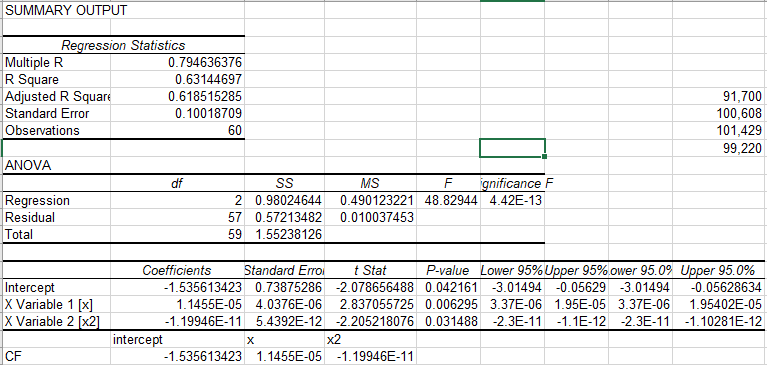
The **figure 9**, illustrates the plot of the forecasted values of CF. The values for the year 2004 Q1, Q2, Q3, and Q4 are 1.49, 1.52, 1.56, and 1.60 respectively.

## Linear Regression Model

## Quadratic Regression Model

To estimate a quadratic relationship between the cyclical factor from the decomposition as well as the number of unemployed people that claimed unemployment insurance, we will use Excel to run a regression analysis. [Chapter 5]

We regressed CF on x and x2 . The regression model obtained will help us to calculate the values of CF for 2004 using the actual mobile home shipment values given in the project requirement. The figure 33 below shows the the highlighted coefficient values for the intercept, x, and x2. Hence, we can calculate CF using the formula, CF = -1.536 + .000011 ***x*** - 1.2 X 10e-11 ***x2***



The whole process of calculating the CF values using the Box-Jenkin’s and Regression methodologies are to calculate the forecast values for 2004. The method that we used in the section Analysing long term trend gave us a linear equation(by regressing centred moving average over time) for the long term trend. The equation enabled us to calculate the trend values for the four quarters of 2004. We also have the values of seasonal indexes obtained from the section of analyzing seasonality. Now we have everything required to calculate the Forecast, which can be represented as,   
 Forecast = Trend (CMAT) X Seasonality index (SI) X Cyclic Factor (CF)

# 

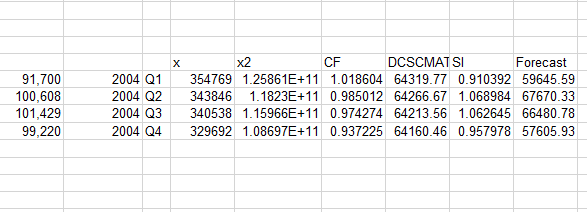
# 

# 

# 

# **Evaluation of Forecast Accuracy**

We used the CF values obtained using the Box-Jenkin’s and regression method to generate 2 different set of forecasts for the year 2004. Since we have been provided with the Actual values of 2004, we will be able to evaluate the accuracy of both the models. We will use the mean absolute percentage error(MAPE) method as the accuracy measure to evaluate both the models. The Table 22 below gives you the statistical result of the evaluation process.



# **Conclusion and Recommendations**