**Quarterly Mobile-home Shipments in the United States**

**ISDS 526-01 21352-: Forecasting for Analytical Decision Making**

**Submitted to:** Dr. Dawit Zerom

**Submitted by:** Ruchika Narang

Sanchit Singh

Sebastian Stuchetz

Abhinay Sariswal

**Table of Contents**

[**Executive Summary** 3](#_Toc531214054)

[**The Forecasting Problem** 4](#_Toc531214055)

[**Examining Data Patterns** 5](#_Toc531214056)

[**Determining patterns in time series using Autocorrelation Analysis** 6](#_Toc531214057)

[**No differencing applied** 6](#_Toc531214058)

[**First order simple differencing applied** 6](#_Toc531214059)

[**First – order seasonal differencing applied** 7](#_Toc531214060)

[**First – order simple and seasonal both differencing applied** 7](#_Toc531214061)

[**Analysis of Seasonality** 8](#_Toc531214062)

[**Trend Analysis** 9](#_Toc531214063)

[**Analysis of Cycle** 9](#_Toc531214064)

[**Analysis of Fitted Forecast** 10](#_Toc531214065)

[**The Forecast** 10](#_Toc531214066)

[**Box Jenkins** 10](#_Toc531214067)

[**Quadratic Regression Model** 11](#_Toc531214068)

[**Linear Regression Model** 11](#_Toc531214069)

[**Quadratic Vs Linear?** 11](#_Toc531214070)

[**Evaluation of Forecast Accuracy** 12](#_Toc531214071)

[**Conclusion and Recommendations** 13](#_Toc531214072)

**Table of Figures**

[**Figure 1: Framework showing goal, decision, and relationship between forecasting and decision making** 4](file:///C:\Users\CampusUser\Desktop\OmSaiRam\Third%20semester\ISDS%20526\Project%203\final%20report_155PM11282018.docx#_Toc531176716)

[**Figure 2:Mobile home shipment time series** 5](#_Toc531176717)

[**Figure 3: Mobile home shipment time series split into 3 parts for better understanding** 5](#_Toc531176718)

[**Figure 4: Correlogram with no differencing applied** 6](#_Toc531176719)

[**Figure 5: Correlogram with first order simple differencing applied** 7](#_Toc531176720)

[**Figure 6: Correlogram with first order seasonal differencing applied** 7](#_Toc531176721)

[**Figure 7: Correlogram with both simple and seasonal differencing applied** 7](#_Toc531176722)

[**Figure 8: Seasonal indices of MHS** 8](#_Toc531176723)

[**Figure 9: MHS vs Deseasonalized MHS Data** 8](#_Toc531176724)

[**Figure 10: Long term trend based on Center moving averages** 9](#_Toc531176725)

[**Figure 11: Cyclical Factor graph** 9](file:///C:\Users\CampusUser\Desktop\OmSaiRam\Third%20semester\ISDS%20526\Project%203\final%20report_155PM11282018.docx#_Toc531176726)

[**Figure 12: Graph showing fitted forecast with our forecast of MHS** 10](#_Toc531176727)

[**Figure 13: Forecast of cyclical factor** 11](#_Toc531176728)

# **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 four quarters of 2004 based on 16 years of quarterly data collected from the period 1988 quarter 1 through 2003 quarter 4. This forecast would help Kim and Larry manage their cash flow to open more such facilities. We have utilized the application “Forecast Pro” to get actionable insights from our data, and thus, created accurate and credible forecasts.

We utilized the multiplicative classical time series decomposition to forecast the data. We examined the data patterns within our time series to better understand 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.

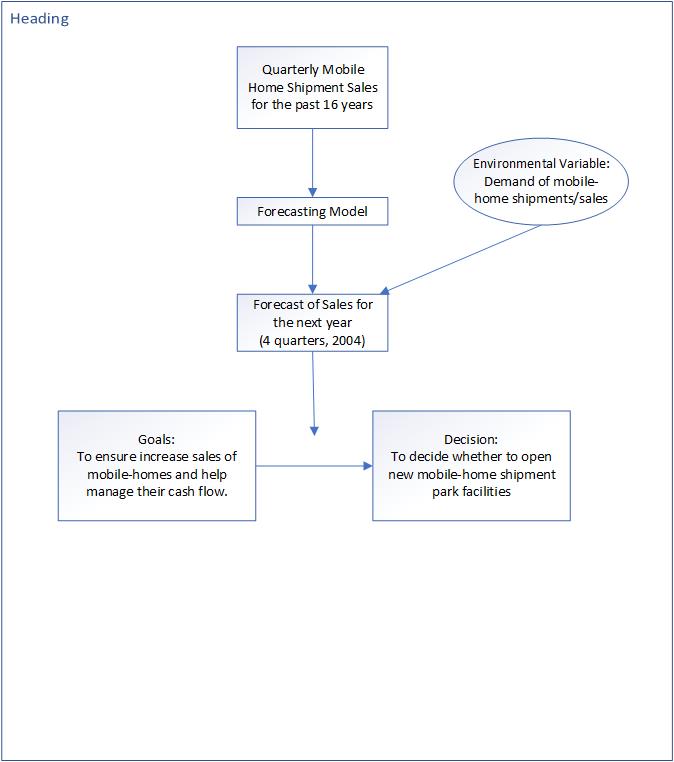
Further, we used Box-Jenkins and Regression approaches to forecast the cyclical factors and thus produce our forecasts based on the observations. Finally, we used MAPE to measure the accuracy of our forecast models.

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

# **The Forecasting Problem**

The problem that Kim Brite and Larry Short are facing is the exclusive Mobile Home Shipments (MHS) data, where they need to forecast sales of mobile home shipments for the 4 quarters of 2004 to manage their cash flow properly. The current MHS appears to impact the vacancy rates and the momentum at which they can fill newly opened parks.

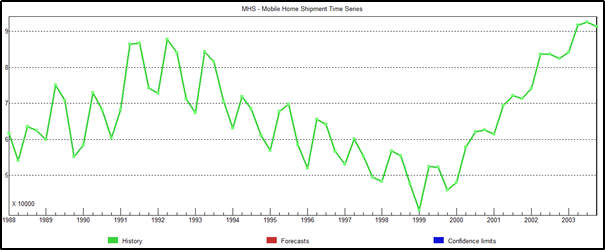
The objective of this report is to provide a decision for and recommendations to Kim and Larry based on analysis of the time series pattern for the MHS data by addressing the trend, seasonality and cyclicity and hence, provide effective forecast according to our forecast model. With the help of this 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 below. This graphical framework also illustrates how our decision making is related to forecasting.

******

**Figure 1: Framework showing goal, decision, and relationship between forecasting and decision making**

# **Examining Data Patterns**

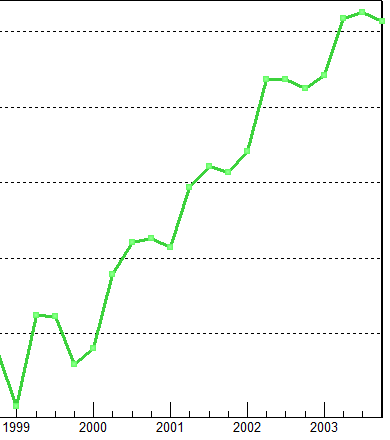
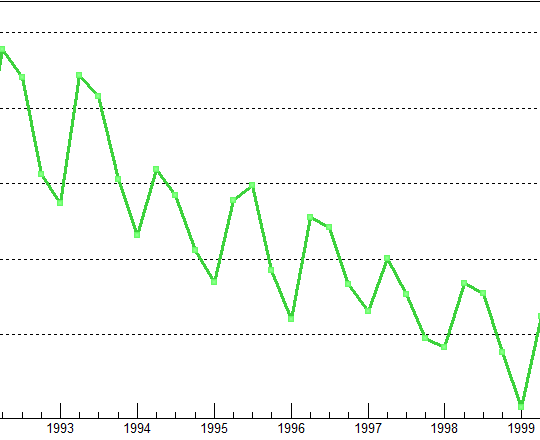
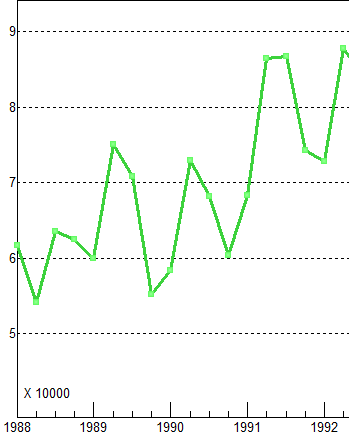
We have set the forecast horizon to 0 in order to better visualize the time series. Forecast horizon is the length of time into the future for which forecasts are to be prepared. Setting the forecast horizon to 0 indicates that we are currently understanding and exploring only the historical time series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



**Figure 2:Mobile home shipment time series**

Below (figure 3) we have divided the series into 3 parts showing an increase, and then a decrease, followed by another increase pattern in the series. The below visualization helps us better understand our data, and thus the series

obtained from it.



**Figure 3: Mobile home shipment time series split into 3 parts for better understanding**

From the above figure we determine that there is an increase during quarter 2 of every year (upward slope) and a decrease during quarter 1 of every year (downward slope), 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.

These findings illustrate that our data is has seasonality, cyclical patterns, is non-stationary, and might also be trended. Seasonality is supported by a 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 has multiplicative seasonality. Additionally, the above time series has a trend since it has a linear increase for 5 years, and then a decrease for 2.5 years followed by a subsequent increase. Now, since these rises and falls are not of fixed period, it indicates a cyclical pattern (rather than a seasonal pattern, which would have fixed periods). Non-Stationarity is depicted by the fact that the series is not constant over mean and variance.

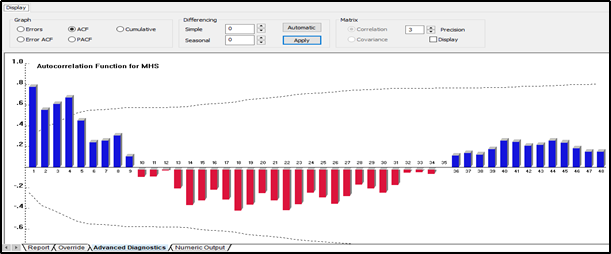
## **Determining patterns in time series using Autocorrelation Analysis**

The autocorrelation coefficients are plotted to demonstrate the autocorrelation function, ACF. The plot obtained by autocorrelation is called the correlogram.

Further, we utilized Autocorrelation Analysis without differencing, with first order simple differencing, with first order seasonal differencing, and with both first order simple and seasonal differencing in order to confirm the data patterns which we observed above by visualizing the time series graph.

### **No differencing applied**

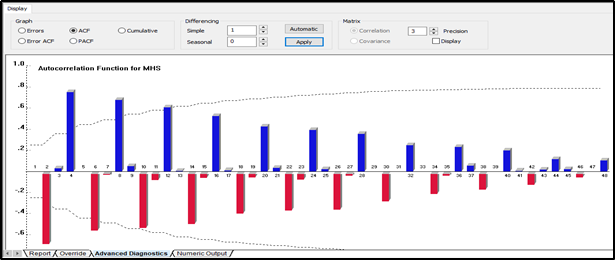
From the below correlogram (figure 4), we examine that successive observations are highly correlated since the autocorrelation coefficients are significantly different from 0 for the first several time lags, and then gradually drop toward 0 as the number of lags increases. As a result, we can infer that the series might have a trend. Furthermore, the autocorrelation coefficient *gradually* drops to 0 instead of rapidly dropping to 0, indicating that the series is non-stationary.



**Figure 4: Correlogram with no differencing applied**

### **First order simple differencing applied**

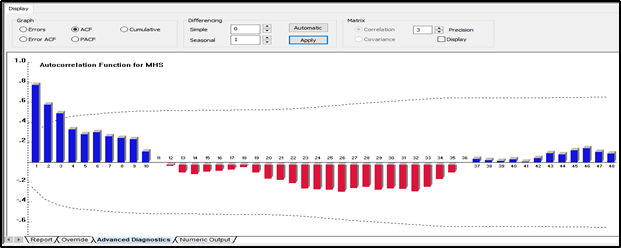
We removed the trend component from the series and can observe that there pertains a strong seasonality indicated by peaks at lags 4, 8, 12, 16, 20, and so on, and depressions at lags 2, 6, 10, 14, 18, which occur due to high Q2 and low Q1 respectively. This pattern is repeated annually across each Q2 and Q1 (figure 5 below). A significant autocorrelation coefficient (ACF) occurs in lags of 4 quarters.



**Figure 5: Correlogram with first order simple differencing applied**

### **First-order seasonal differencing applied**

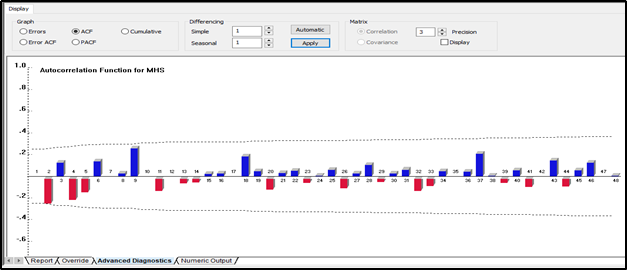
We now removed the seasonality from the series to confirm the existence or non-existence of trend. We observe a similar pattern to what we observed from the correlogram, with no differencing that successive observations are highly correlated, since the autocorrelation coefficients differ significantly from 0 for the first several lags, before gradually dropping toward zero as the number of lags increases. This confirms the trend within the series. We also observe that this trend is visible in different periods, hence, the series seems cyclic (figure 6 below).



**Figure 6: Correlogram with first order seasonal differencing applied**

### **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 (figure 7 below).

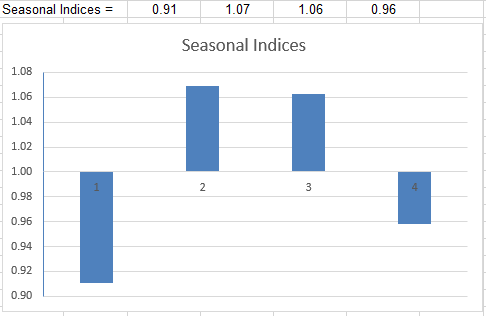


**Figure 7: Correlogram with both simple and seasonal differencing applied**

# **Analysis of Seasonality**

Most organizations are interested in knowing how their sales performance on a seasonal basis compares to its normal variation. In figure 8 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 occurring 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.

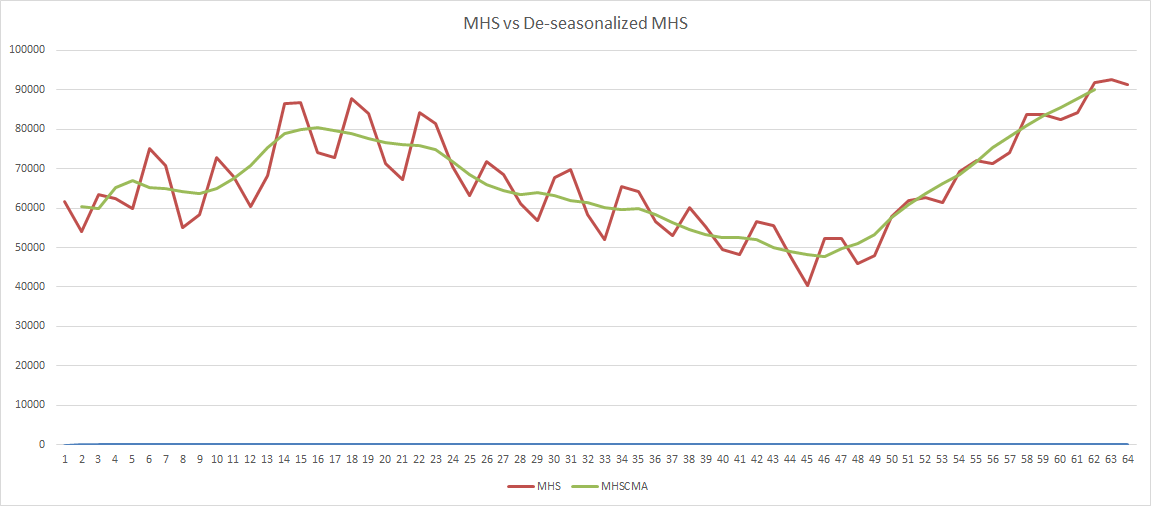
Seasonal index is a measure of how a particular season in MHS through some cycle compares to the average season of that cycle. Applying seasonal indices will deseasonalize mobile homes data, and thereby smooth it to allow forecasting of trends. Using deseasonalization, we are eliminating seasonal alterations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

****

**Figure 8: Seasonal indices of MHS**

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 the model is multiplicative, so the sum of seasonal factors should add up to 4. From figure 8, the total of seasonal indices calculated from the MHS data is 4. Therefore, no adjustment is required for the calculated indices.

As depicted in figure 8, 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 conclude that Q2 and Q3 have somewhat outperformed the seasonal average by 7% and 6%. However, Q1 and Q4 seem to have underperformed the seasonal average slightly by 9% and 4%. We plotted a graph, figure 9, 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 an increase or decline in the economy.



**Figure 9: MHS vs Deseasonalized MHS Data**

# **Trend Analysis**

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

The below graph is obtained by regressing the dependent variable, i.e. the centered moving averages for mobile home shipments data by the independent variable, time. We observe that the plot has a linear decreasing trend. This is indicative of the sales of mobile home shipment units decreasing in the future, which would not prove well for Kim’s and Larry’s business.

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

**Figure 10: Long term trend based on Center moving averages**

# **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)*).*

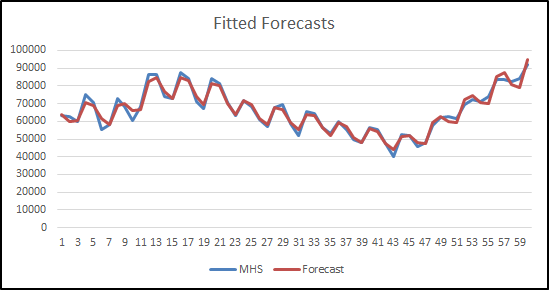
In the trend analysis section figure 10 above, where we analyzed the long-term trend, we could see that the trend data showed negative linearity with respect to the time. In this section, we analyze the impact of the cyclical factor on our historical data. The plot in figure 11 below shows the variation 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 long-term trend values determined.

**Figure 11: Cyclical Factor graph**

# **Analysis of Fitted Forecast**

In this section we plotted the observed forecast values versus the actual MHS values. The red line in figure 12 depicts the forecasted values while the blue one is for 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.

****

**Figure 12: Graph showing fitted forecast with our forecast of MHS**

# **The Forecast**

Forecasting is a method that utilizes historical information as the input to make educated quotes that are predictive in identifying the instructions of future patterns. Organizations use forecasting to figure out how to designate their budget or prepare for unexpected expenditures for a certain amount of time. In our project, forecasting MHS is constructed using 16 years quarterly historical data. From the data pattern analysis, we can perceive a presence of 3 components: trend, seasonality and cyclical factor. The forecast is created by multiplying all three components.

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, and 0.96 respectively.

We forecast the cyclic factor (CF) for the year 2004 using the Box Jenkins and Regression approaches described below. These CF values were then used in the calculation of the forecast values.

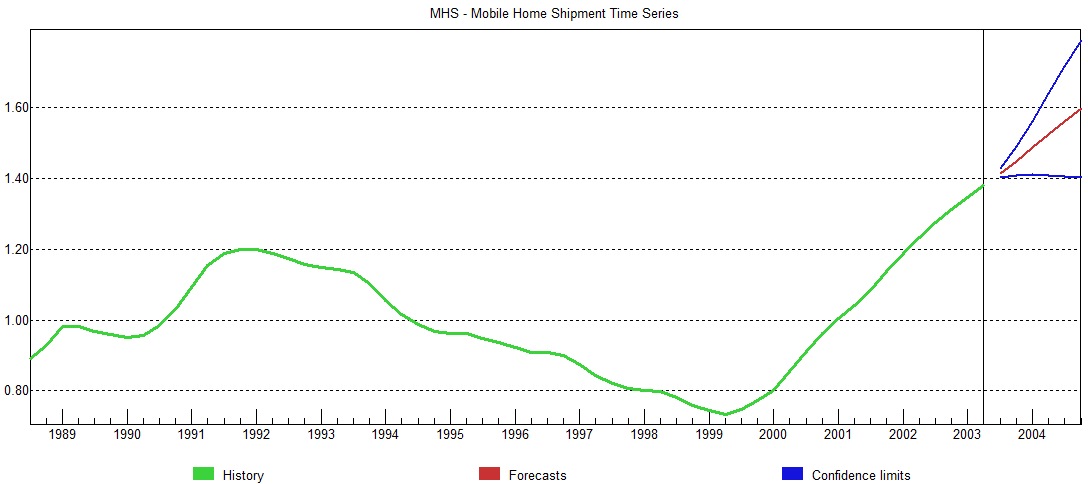
MHS forecast can be calculated from forecasted CF values as follows:

*Forecast = Trend (CMAT) \* Seasonality index (SI) \* Cyclic Factor (CF)*

## **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. We applied the Box-Jenkins forecasting method while forcing a constant to generate a forecast from Q3 2003 to Q4 2004 in Forecast Pro.

Figure 13 illustrates the plot of the forecasted values of CF.



**Figure 13: Forecast of cyclical factor**

## **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 used Excel to run a regression analysis.

We regressed the dependent variable CF on the independent variables x and x2. The regression model obtained further assisted us in calculating the CF values for 2004 using the mobile home shipment numbers given in the project and in turn, the forecast of MHS. Table 1 below shows the adjusted R2 and coefficient values for the intercept, x, and x2. Hence, the quadratic equation of the estimated mode looks as follows:

*CF = -1.536 + 0.000011x - 1.2E-11x2*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **adjusted r2** | **Coefficient of Intercept** | **Coefficient of x** | **Coefficient of x2** |
| **Quadratic Regression** | 0.62 | -1.5356 | 1.1455 | -1.1994 |

**Table 1**

## **Linear Regression Model**

To determine the linear relationship, we regressed cyclic factor, CF on x and got the following equation:

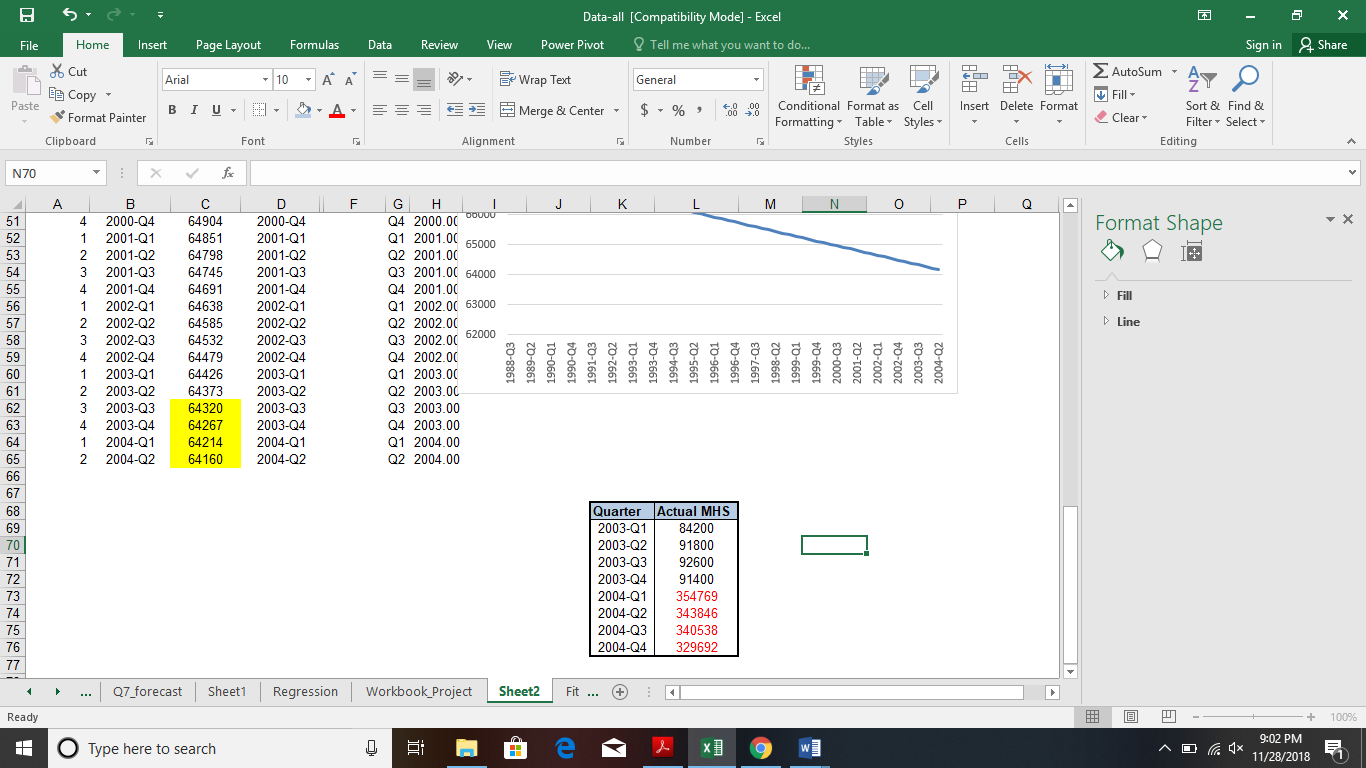
*CF = -0.07956 + 2.57059E-6x*

## 

## **Quadratic Vs Linear?**

## A quadratic regression model is the more appropriate model for our data compared to a linear regression model. This is the case because the cyclical factor - as seen in the “Cyclical Factor” figure in “Analysis of Cycle” - is rather curvilinear than linear (Source: “Regression Models: How do you know you need a polynomial?” Karen Grace – Martin, https://www.theanalysisfactor.com/regression-modelshow-do-you-know-you-need-a-polynomial). A quadratic model is generally used for predicting curvilinear (wave-like) patterns.

If we take a look at the statistical values generated via linear regression of CF on x (in the table below), we see that the adjusted R2 value for the model is 0.59. On the other hand, when we perform a quadratic regression, the adjusted R2 is 0.62, which is greater, and therefore better performing than the linear regression model. This indicates that our model gives a better calculated forecast than the linear regression. If we look at the actual values of 2004 MHS in Table 2, we can see that there is a significant drop in the numbers; roughly from 90K to 35K during Q4 2003 to Q1 2004. This kind of nature is very difficult to predict. The forecast for 2004 generated using linearly regressed CF looks better than the quadratic forecasted CF model because the actual values of MHS are in favor of the linear model by chance. Hence, the linear model is a biased model and should be avoided.



**Table 2**

# **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.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **2004** | **Q1** | **Q2** | **Q3** | **Q4** |
| **Actual values** | 354769 | 343846 | 340538 | 329692 |
| **Box-Jenkins** |  |  |  |  |
| **Forecasted CF** | 1.49 | 1.52 | 1.56 | 1.60 |
| **Forecast values** | 87248.77 | 104424.10 | 106448.50 | 98342.00 |
| **Quadratic Regression** |  |  |  |  |
| **Forecasted CF** | 1.02 | 0.99 | 0.97 | 0.94 |
| **Forecast values** | 59645.59 | 67670.33 | 66480.78 | 57605.93 |
| **Linear Regression** |  |  |  |  |
| **Forecasted CF** | 0.91 | 0.88 | 0.88 | 0.85 |
| **Forecast values** | 53401.31 | 60723.13 | 59732.95 | 52091.24 |

**Table 3**

We will use the mean absolute percentage error (MAPE) method as the accuracy measure to evaluate both the models. The Table 4 below gives us the statistical result of the evaluation process.

|  |  |
| --- | --- |
| **Method** | **MAPE %** |
| Box-Jenkins | 132.31 % |
| Quadratic Regression | 48.31 % |
| Linear Regression | 33.29 % |

**Table 4**

Looking at Table 4, we can observe that the Box-Jenkins Mean Average Percentage Error is 133%, which is very high and thus unacceptable. It means that the forecast error is greater in magnitude than the actual value of the mobile home shipments in 2004. This is clearly not the forecasting model that Brite and Short would like to consider. However, the error percentage for the regression model is 48%, which is a much better performance than the Box-Jenkins method. It is evident that the incorporation of the initial claim of unemployment insurance benefits has played a major role to improve the forecast accuracy. The Box-Jenkins method has not considered the effect of the claim of unemployment benefits on the cyclical factor. Incorporating this in the regression model helped to increase the accuracy of the forecasting model, which leads to it being more considerable for Brite and Short. They can strategize the approach based on the fact that the forecasting errors can be as much as 48% of the actual mobile home shipments. In our conclusion, we will discuss the advantages and the recommendations for Brite and Short.

# **Conclusion and Recommendations**

Proper demand forecasting enables better planning and utilization of resources for businesses to be competitive, which leads to the search of the best fitting models for any given situation.

The linear regression model performed the best on our data. However, we have to keep in mind that the monthly shipment values in 2004 plummet down with an unprecedented severity, for which many potential explanations may exist (e.g. economic downfall, recess, natural catastrophe, etc.). Because of that, in our case, it can confidently be said that the linear regression performs better than the quadratic regression method by chance. That said, we still recommend to use a quadratic approach in data problems like this. We do so because such downfalls or spikes are fairly unusual (especially considering the amount of historic data). There is no denying that the quadratic model would have performed better than the linear model if the mobile home shipment values would not have changed this drastically from Q4 2003 to Q1 2004.

Currently, for 2004, we do not recommend adding further mobile-home parks, since 2004 saw a massive decline in mobile home shipments. This might have been caused by an economic downfall or other factors. With fewer mobile homes being sold, it would not make sense to add additional parks for mobile homes, since they would not be occupied sufficiently