ISDS 526

Forecasting for Analytical Decision Making

Project 3

Forecasting Mobile Home Shipments to Guide Business Expansion Decision Making

Spring 2018 session

Professor: Dawit Zerom

Project Team

Aishwarya Chavan	893431940
Malvika Kulkarni	893215814
Prajval Ramachandran	893423897
Skandha Chinta	893216168

Table of contents

Executive Summary

Kim Brite and Larry Short have developed mobile-home parks which aims at providing privacy to the occupants in a pleasant environment. A decision has to be made whether to open new facilities for more profit and help Kim and Larry manage their cash flows. The forecasting report focuses on the analysis of the time series of quarterly data on mobile home shipments. The data available is of 16 years, beginning with 1988Q1 and ending with 2003Q4. The mobile home shipments are forecasted for the year 2004. The data is analyzed for any possible trend, seasonality or cyclical components and it is observed that all three components are present in the time series. The seasonality is first analyzed by observing the spikes and troughs in the time series and seasonal indices are calculated. The trend in the time series is analyzed and it is observed that the trend is declining over the period of years. The occurrence of cyclic component is also measured and the cyclic factors are computed. As the cyclic component has a huge impact on the forecast, the cyclic factor is forecasted using three approaches, namely, Box Jenkins forecasting method, Linear and Quadratic Regression method. Based on the forecasted cyclic factors, sales for mobile home shipments are calculated using time series decomposition method. Based on the analysis and forecasts, the following report explains each step taken to reach a decision and makes recommendations which are lucrative to Kim & Larry.

Forecasting Problem Set-up

The forecasting goal for Mobile Home Shipments (MHS) data is to predict the sales for the year 2004. The objective of this report is to analyze the time series for the presence of trend, seasonality and cyclicity. Using this data, the report should help Kim & Larry to make a decision on whether to open more mobile home parks. The external factors like demand for MHS and US economy should also be considered by the forecasting model. The forecast of cyclic components in the time series leads to the prediction of sales using time series decomposition method. The roadmap for the forecasting analysis of manufactured home shipments is shown in Exhibit 1.1, this graphical framework illustrates how decision making in this project was related to forecasting models.

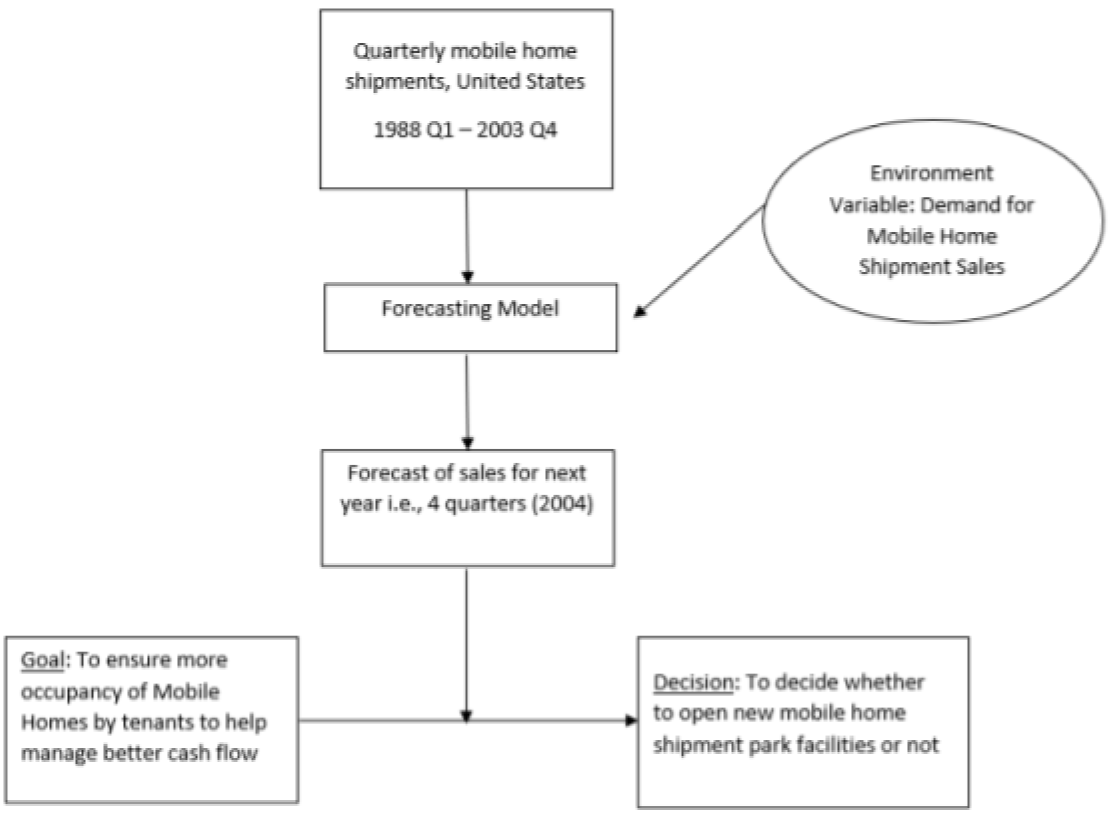
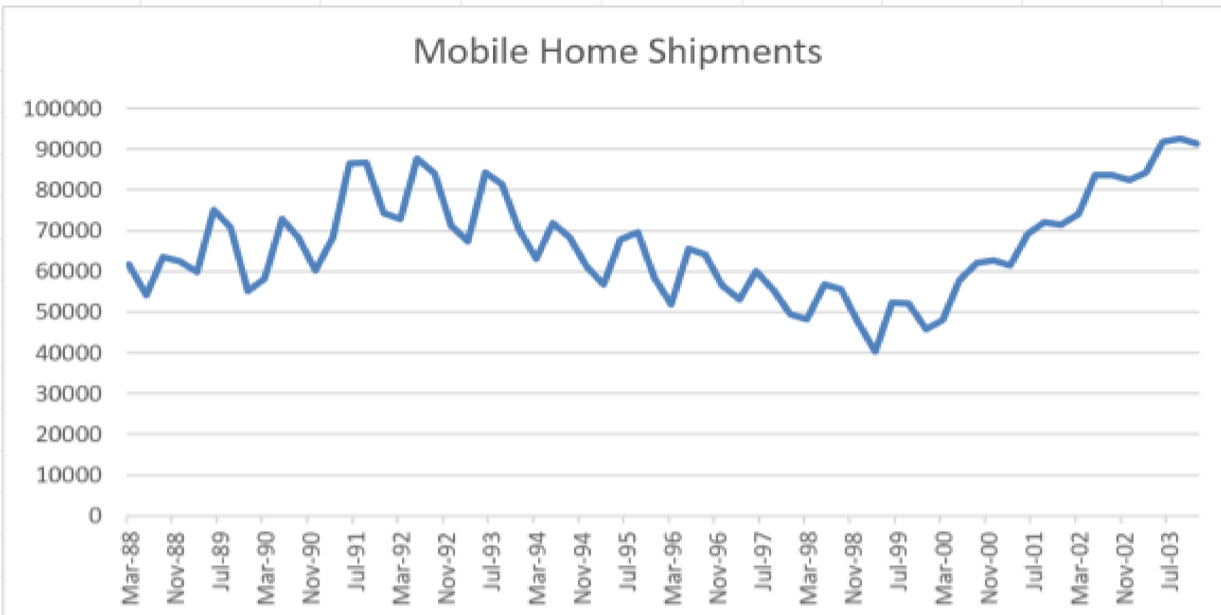


Exhibit 1.1: Forecast Problem Definition of MHS

Examination of Data Patterns

The data represented graphically below is the quarterly sales data for MHS from 1998 to 2003. The examination of data patterns will provide details about the behavior of the data. Also, Auto Correlation Analysis is conducted to base claims regarding the trend and seasonality.



MHS vs Time

Exhibit 2.1: Mobile Home Shipments over given time period.

The data exhibits recurring pattern in every quarter of the time series. Exhibit 2.1 displays peaks in Quarter 2 and lows are commonly observed in Quarter 1. The repetitive pattern in the time series indicates seasonality in the time series. Just by looking at the time series, the trend component cannot be confirmed. But, by looking more closely, there is an indication of cyclic behavior due to long term rise and decline of the time series that are not in a fixed time period.

A better way of analysing the components of the time series patterns is to observe the Correlogram obtained from Autocorrelation Function (ACF). ACF is the statistical approach to detect patterns in the time series. The graph of MHS data in exhibit 2.2 shows both trend and seasonality. The trend in the time series can be noticed as the lags gradually drop to zero by looking at the ACF graph. It is also observed that the time series has peaks and drops appearing every four quarters. Therefore, it can be confirmed that the time series is showing a seasonal pattern as well.

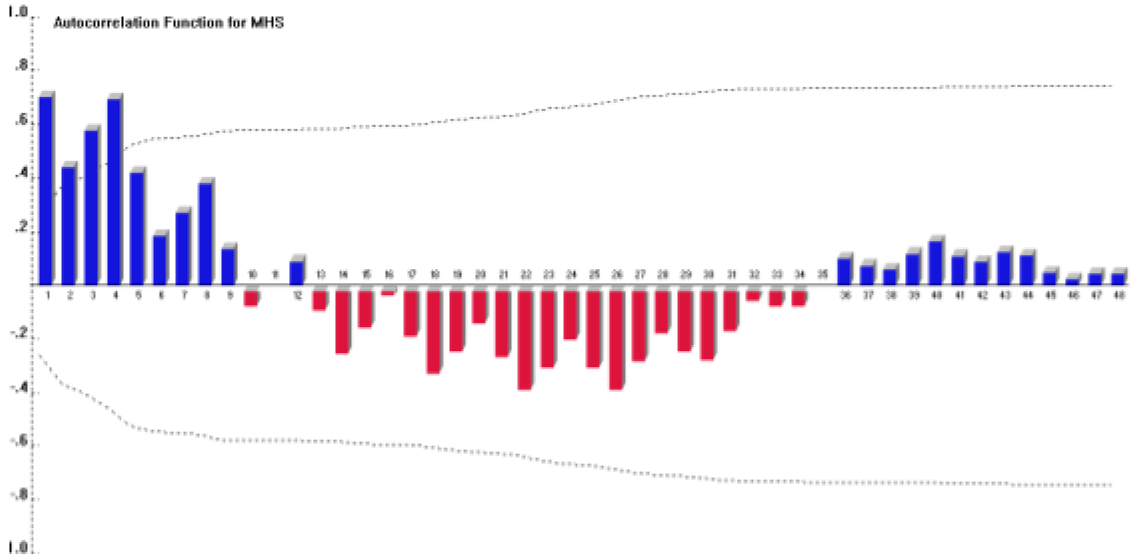


Exhibit 2.2: AutoCorrelation Function Analysis

If autocorrelation is high for the first few periods and slowly tends towards zero, it would indicate a strong trend in the data. But in exhibit 2.2 the spikes decrease after lag 1 and again at lag 3 and 4 the spikes increases, so the trend in the time series is difficult to confirm. Although it can be observed that the coefficients gradually drop to zero by looking at the Autocorrelation Function (ACF) graph. To confirm presence of any pattern, differencing is performed.

The trend component in the time series can be removed by applying first order simple differencing. By removing the trend, the Exhibit 2.3 still shows the presence of the seasonality in the time series. It can be confirmed by observing spikes for lags of 4, 8, 12 and so on, as well as significant negative spike for multiples of lag 2. Thus, the graph in Exhibit 2.3 is a time series with seasonal components.

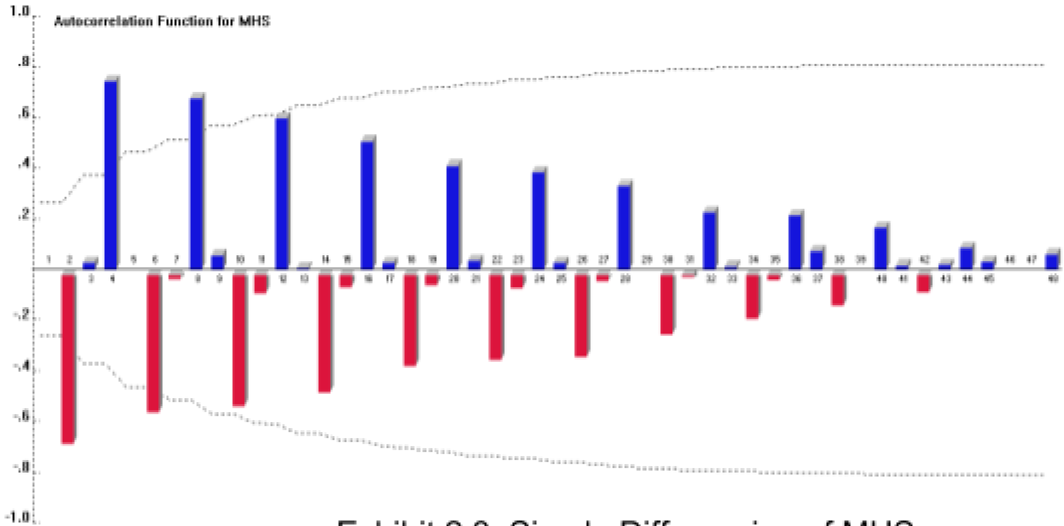


Exhibit 2.3: Simple Differencing of MHS

In the same way, by applying seasonal differencing on Exhibit 2.2, the seasonality from the time series can be removed, which is depicted in Exhibit 2.4. It is observed that the successive correlations are highly correlated and they gradually drop towards zero as the number of lag increases. Hence, the presence of trend component in the time series can be confirmed.

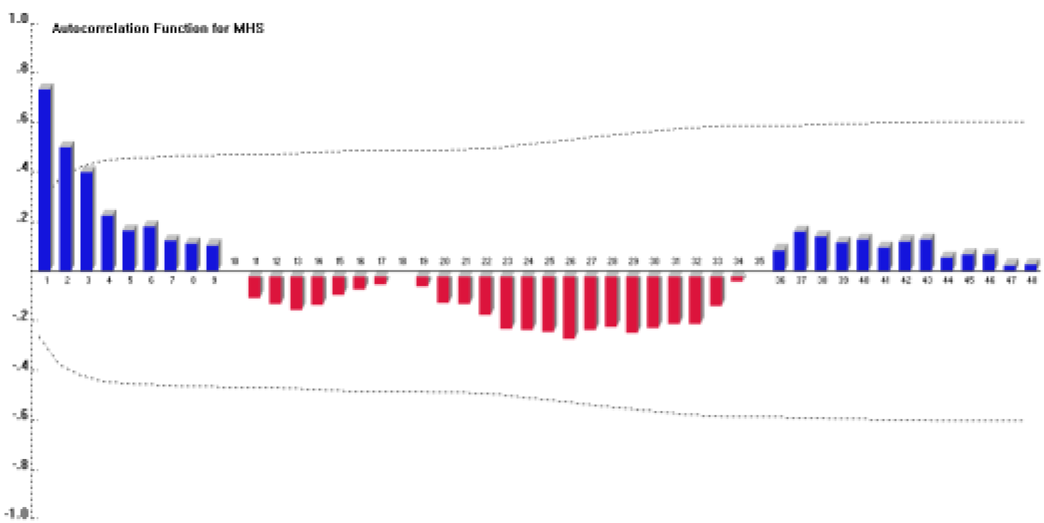


Exhibit 2.4: Seasonal Differencing of MHS

After applying both simple and seasonal differencing, i.e. removing the trend and seasonality from the time series, the end result is nothing but a stationary time series, shown in the Exhibit 2.5.

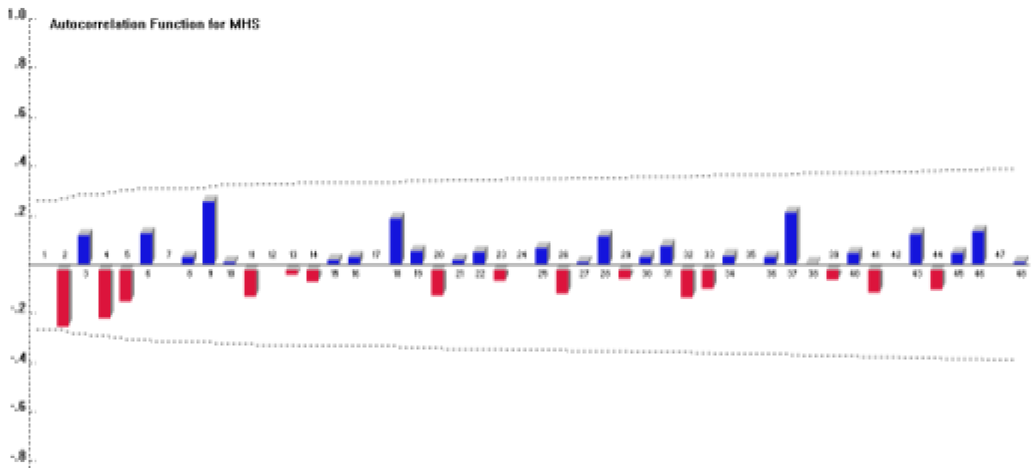


Exhibit 2.5: Simple & Seasonal Differencing of MHS

Analysis of Seasonality

Organisations handling seasonal variation are often interested in recognizing their performance relative to the normal seasonal variation. Plotting Mobile homes shipment data from 1988 to 2003 on a graph shows peaks and lows at regularly recurring intervals, as shown in Exhibit 2.1. For most of the data, peaks are commonly observed in the Q2, and lows are commonly observed in Q1. These hikes and drops lead us to believe that seasonality may exist in the data. However, this is not a conclusive enough evidence to point out seasonality.

Seasonal indices measure and showcase how a particular season or a quarter of mobile home shipments compare with the average season of that cycle of MHS. Eliminating the linearity of the data, by performing differencing shows some patterns of seasonality in the data, as shown in Exhibit 2.3. Seasonal indices are used to deseasonalize data and smoothen the data to allow for forecasting of the trend in the data. By deseasonalizing the data, seasonal fluctuations in the shipment data are removed to predict or approximate future values.

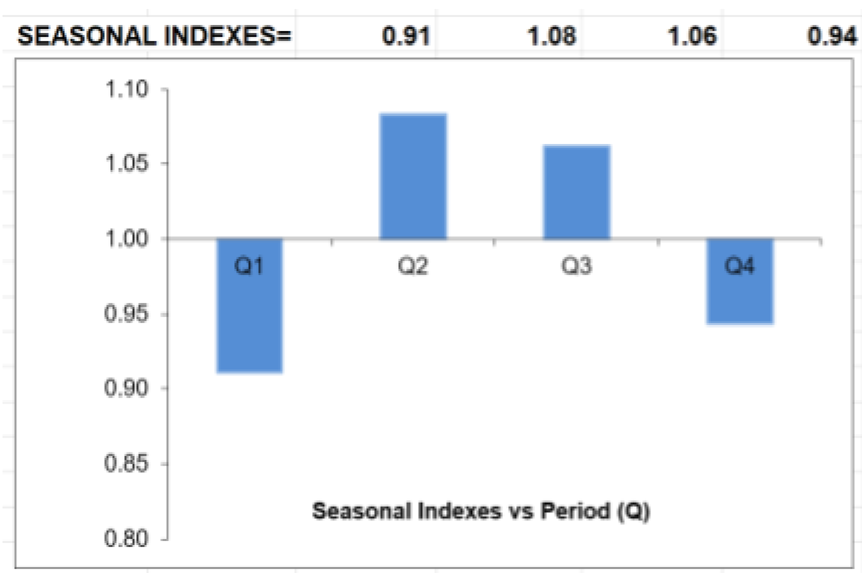


Exhibit: 3.1 Seasonal index for MHS

Seasonal variations have been removed using centered moving averages (CMA). Measuring seasonality in mobile home shipments is possible by comparing the actual value of the series in any given time period with the corresponding deseasonalized value. Thus, the seasonality factor is calculated. The sum of seasonal factors for the given periods should sum up to 4 since there are four quarters and it's a multiplicative model. However, since it totals 4.0003, a 0.9999 multiplier (n-period / sum of the unadjusted averages) is used to arrive at a new adjusted seasonal index for each period.

As observed from the exhibit 3.1, for Q1, Q2, Q3, and Q4 the seasonal indexes are 0.91, 1.08, 1.06 and 0.94 respectively. These numeric figures indicate that Q2 and Q3, seem to overperform by 8% and 6% when compared to the seasonal average, while Q1, and Q4 seem to underperform by 9% and 6%.

MHS and deseasonalized MHS are plotted to give a measure of the degree of seasonality. The rise and fall of Mobile Home Shipments may be attributed to the increase and decline in economy. It can also be attributed to the demand in Multi-family houses and apartments.

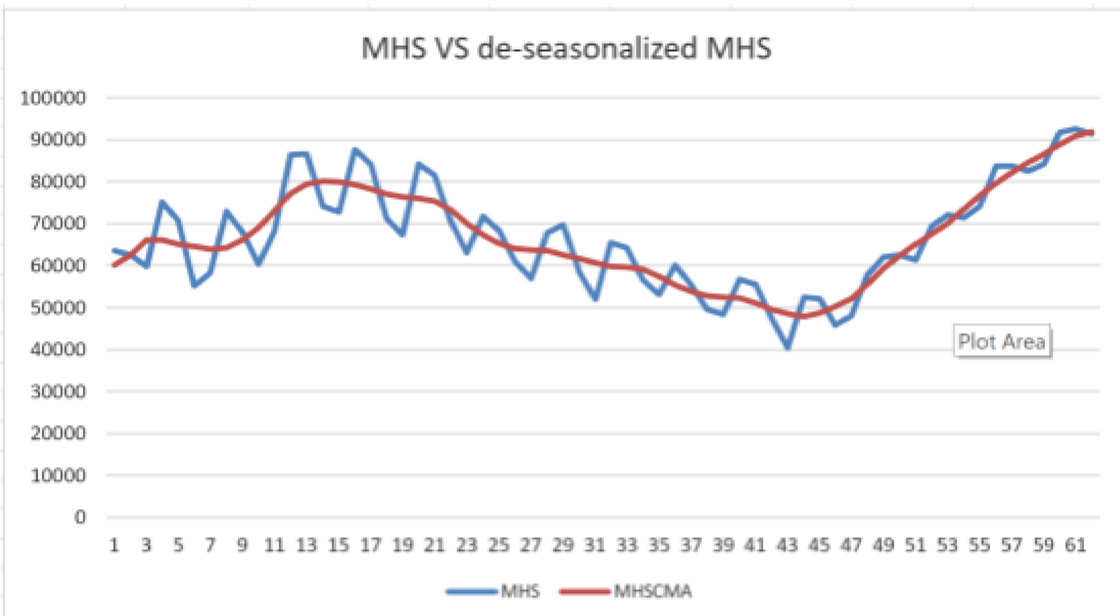


Exhibit 3.2: Deseasonalized data against MHS

Trend Analysis

A trend analysis allows the organization to predict what will happen with the shipments in the future. It is developed using the deseasonalized data of mobile home shipments. The Centered Moving Average is the average of shipments centered at (imaginary) point, a full period ahead of the average center. A linear equation of centered moving averages of MHS is used to obtain the trend.

A regression analysis is performed taking the deseasonalized data as the dependent variable and Time as the independent variable. The regression output gives us the linear relationship between the variables. This output is called centered moving average trend, as its formulation is time-dependent. All available historical time is taken into the equation and a long-term trend is developed by plotting a graph between the centered moving average trend and Time. As perceived from the plotted graph in exhibit 4.1, a declining trend is observed, which is not really a promising future for Brite and Short.

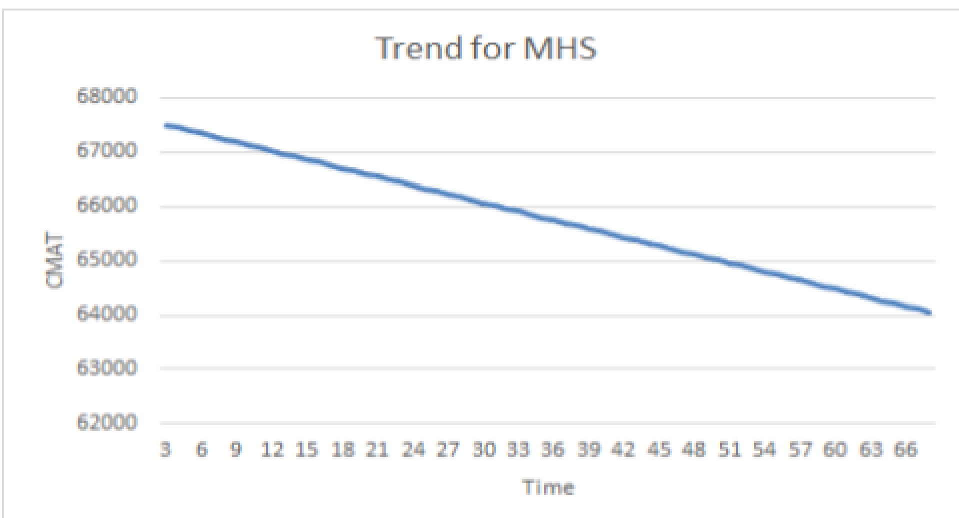


Exhibit 4.1: Trendline for MHS

Analysis of Cycle

The Mobile Home Shipment (MHS) data showed seasonality and weak trend after calculating Centered Moving Average. Then regression is performed on the actual data to perform trend analysis. The deseasonalized data provided insights on the underlying pattern in the data. However, even after deseasonalizing the MHS data, there are fluctuations about the long term trend. Any remains of wavelike pattern in deseasonalized data is nothing but cyclicity. The cyclical component of a time series is measured by a cycle factor (CF). The cycle factor is calculated using the ratio of the centered moving average to the centered moving average trend.

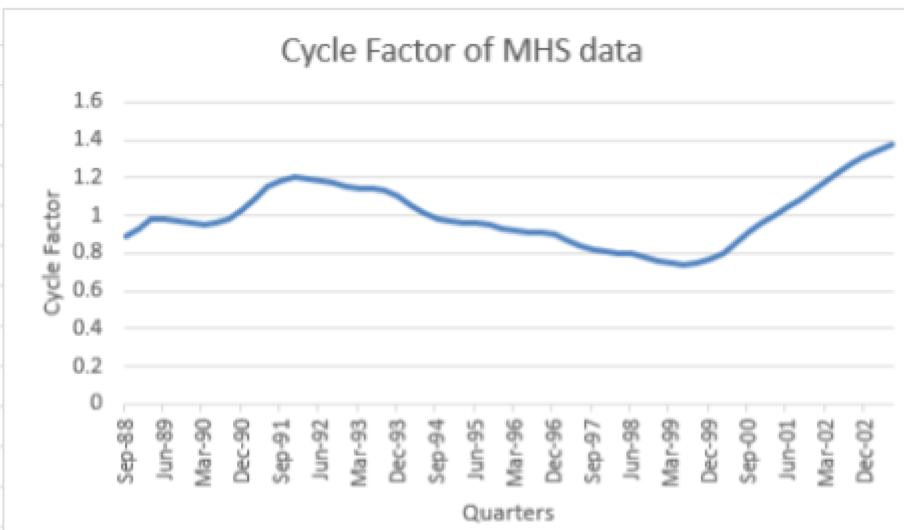


Exhibit 5.1: Cyclic Factor over time

The analysis of cycle factor helps to project where the industry is headed or any unexpected events in the cycle. From exhibit 5.1, it is observed that the CF starts to increase beyond 1.2 starting from Q1 of 2002. This indicates the occurrence of another Cyclic pattern along the time series. CF moves along the line at exactly 1.00 and suddenly goes up during the month of December 2002. The cycle factor increases as we go ahead in time.

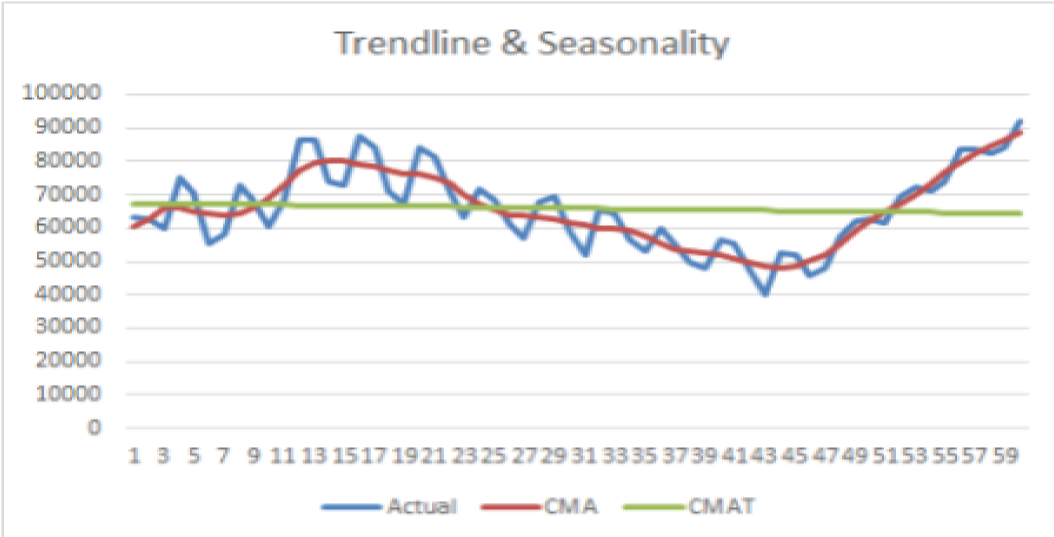


Exhibit 5.2: Plot of deseasonalized data along trendline

The cycle factor is plotted in the graph above to show how computing CF explains the movement of deseasonalized data of MHS around the trendline in exhibit 5.2. The plotted graph in exhibit 5.1 shows that as the CF increases in the quarters between Q3 of 1991 and Q2 of 1992, the MHS values also increases in exhibit 5.2. The decreasing value in Q4 of 1991 and Q1 of 1992 can be accounted by the effect of seasonality occurring every year. So it is evident that MHS and cyclic fluctuations are related.

The Cycle Factor varies around the values 1 which means that a cycle greater than 1 indicates that the deseasonalized value for that period is above the long-term trend of the data. If CF is less than 1, the deseasonalized value is below long-term trend of the data for that period. The plot above explains this. The trend line moves as the CF is moving around the line 1.

Analysis of Fitted Forecast

To compare the actual MHS data with forecasted MHS data, decomposition method is used to forecast the MHS sales data and is plotted in the graph shown in Exhibit 6.1. From this graph it can be seen that the forecasted values are fitting the actual data seamlessly.

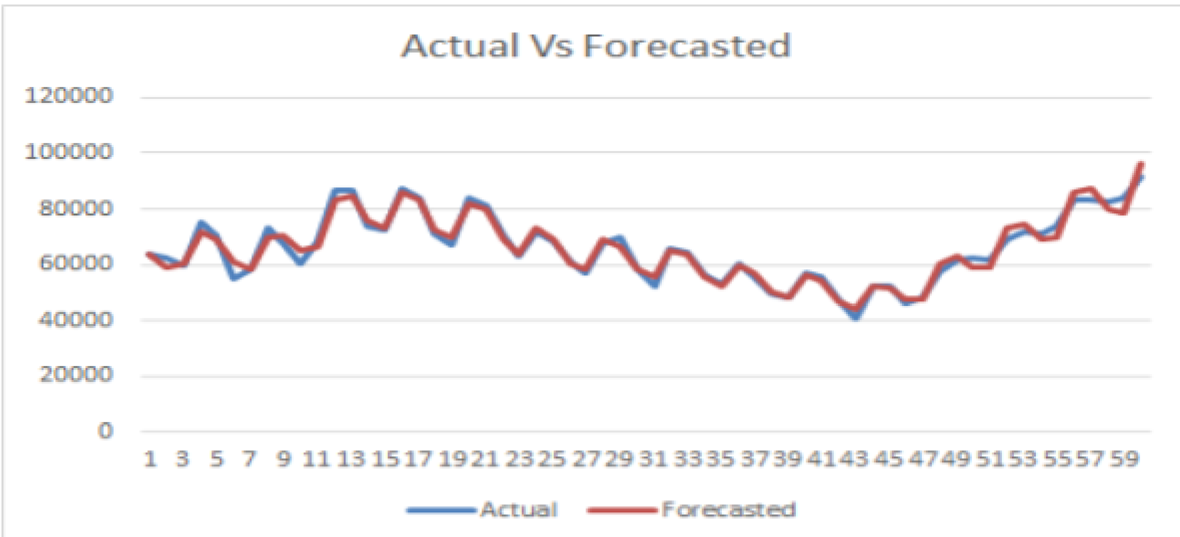


Exhibit 6.1: Fitted forecast plot for MHS

The accuracy of this model is determined using measures such as MAPE (Mean Absolute Percentage Error). This gives the percentage error between the actual values and the forecasted values. For this model, the forecasted values are just 0.12% off by the actual values. So this concludes that the forecast values fit well with the actual values of Mobile Home Shipments.

The Forecasts

Forecasting is the technique of predicting the future values by using the historical data. The forecast for the year 2004 is generated using quarterly historical data of 16 years. The time series confirms the presence of trend, seasonality and cyclic patterns. The forecast is generated by multiplying all three components along with the irregular component.

Centered Moving Average Trend (CMAT) is generated to find the long term trend, which is an estimate of trend value of Centered Moving Average (CMA). The values calculated for CMAT for the year 2004 are shown in Exhibit 7.1

Quarter - Year	Time Period	CMAT
Q1 - 2004	65	64213.56
Q2 - 2004	66	64160.46
Q3 - 2004	67	64107.36
Q4 - 2004	68	64054.26

Exhibit 7.1: CMAT for the year 2004

As described in the analysis of seasonality, Seasonal Indexes (SI) are calculated as 0.91, 1.08, 1.06, and 0.94 for Quarter 1, Quarter 2, Quarter 3 and Quarter 4 respectively.

Irregular component is considered as 1, as no anomalies are expected to occur.

The cyclic component of the time series, i.e. the Cyclic Factor (CF) has a deep impact on the forecasts as it offers an understanding of where the industrial situation is headed. That is why it is the most difficult component of the time series to forecast.

The following approaches have been used to calculate forecasts for CF:

- Box Jenkins forecasting method
- Linear Regression
- Quadratic Regression

Approach 1: Box Jenkins forecasting

To forecast the CF values for the year 2004, the historical CF values are calculated using deseasonalized MHS data. The values are then loaded into Forecast Pro to forecast the CF for the periods beginning from Q3 - 2003 to Q4 2004.

Approach 2: Linear Regression

The linear relationship is estimated between the CF and number of unemployed people who claimed the unemployment insurance (x). In this case, x is the independent variable for mobile house shipments, because, when people are unemployed, people look for more cheaper accommodation options. Cyclic Factor is calculated based on the independent variable for the year 2004.

R-square measures how much the dependent variable can be explained by the independent variables. In this case, how much CF can be attributed to number of unemployed people who claimed the unemployment insurance. For the linear regression model, the R-square value is 0.6, which indicates that the model is good. Also, the p-value is less than the significance level (α), which indicates that the relationship between the CF and number of unemployed people who claimed the unemployment insurance (x) is significant.

Approach 3: Quadratic Regression

In this method, a quadratic relationship is estimated between the CF and number of unemployed people who claimed the unemployment insurance (x). R-square for quadratic regression model is 0.63, which also indicates that the model is good. Also, the p-value is less than the significance level (α), which indicates that the relationship between the variables is significant.

By plotting the cyclic factors and number of unemployed people who claimed the unemployment insurance (x) in a scatter plot, it is observed that there is a correlation between these two variables. This shows that the cycle factor is influenced or related to the unemployed people who claimed the unemployment insurance. The scatter plot shows that the relationship between the two variables is not linear. After a while, as number of unemployment insurance claims increases, the value of CF appears to be stagnant. After certain point, the US economy gets so bad that the curve starts to descend and the CF decreases.

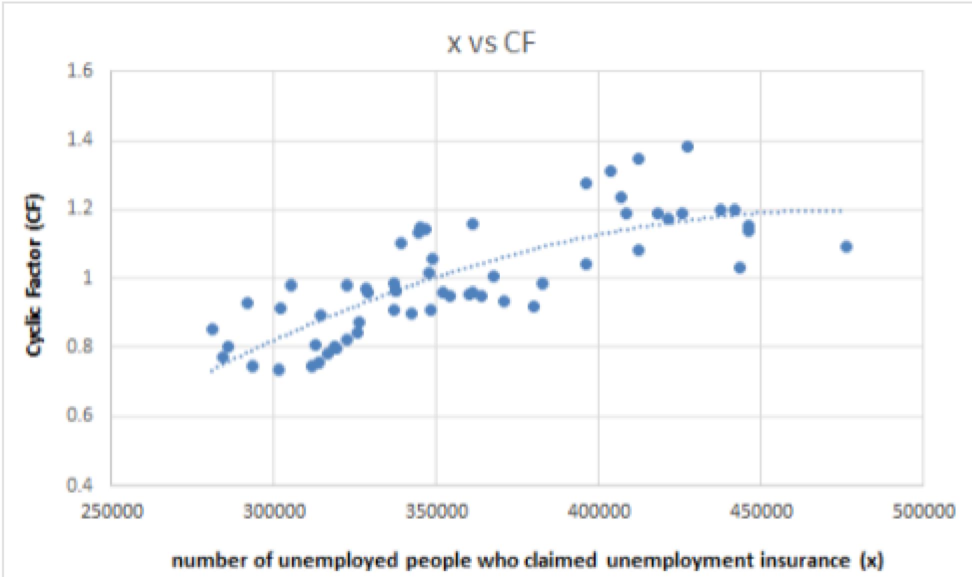


Exhibit 7.2 Scatter Plot x vs CF

Hence, it is concluded that Quadratic regression is the best model for determining the cyclic factor compared to Box Jenkins and Linear regression model.

The forecasted results for the CF values for the year 2004 by using three approaches are as follows -

Quarter - Year	Box Jenkins Method	Quadratic Regression	Linear Regression
Q1 - 2004	1.487	1.019	0.992
Q2 - 2004	1.524	0.985	0.963
Q3 - 2004	1.56	0.974	0.955
Q4 - 2004	1.597	0.937	0.927

Exhibit 7.3: Forecasted CF for the year 2004

It is known that Box Jenkins method does not account for the economic business cycle, which is a major factor in determining the cycle. For this reason, Box Jenkins method is ruled out. After comparing the linear and quadratic equation, it is observed that the R-square and p-value are low for the Quadratic regression model than linear regression. Also, by observing Exhibit 7.2, it can be implied that the quadratic model takes an account of the external factors like the US economy due to which the quadratic regression is a superior fit than the linear one.

Forecast

By keeping CMAT, SI and I as common and using CF from the three approaches, the following forecasts are generated for the year 2004 -

Quarter - Year	Forecast 1 (CF calculated using Box Jenkins method)	Forecast 2 (CF calculated using Quadratic Regression)	Forecast 2 (CF calculated using Linear Regression)
Q1 - 2004	87009.770	59602.209	58017.909
Q2 - 2004	105932.255	68524.184	66968.753
Q3 - 2004	106213.270	66333.836	65017.908
Q4 - 2004	96529.618	56649.977	56035.857

Exhibit 7.4: Forecasted MHS for the year 2004

Evaluation of Forecast Accuracy

Mean Absolute Percentage Error (MAPE) is used for evaluating forecasting accuracy. It is the measure of how different the forecast values are when compared to the actual values, in percentage terms. Lower MAPE for a forecasting model implies better forecasting accuracy.

The Cyclic Factor is the most difficult factor to predict and can easily skew the forecasts. By anticipating this, three methods are used to predict the cycle. The summary of the MAPE values are given below -

Method	MAPE (%)
Box Jenkins	132.7
Quadratic Regression	48.07
Linear Regression	45.06

Exhibit 8.1: Calculated MAPE values

Box Jenkins forecasting method gives the highest MAPE value of 132.7%. As this value is high and unacceptable, linear and quadratic regression analysis have been carried out that included number of unemployed people who claimed the unemployment insurance. Using quadratic regression, MAPE is derived to be 48.07% . In the same way, by using linear regression, MAPE is derived to be 45.06%. Based on these calculations, it is observed that number of unemployed people who claimed the unemployment insurance is correlated with the cyclic factor which forms a crucial part in the forecast.

Conclusion and Recommendation

The exploration of data shows sudden decline in sales due to the outside factors like drop in economy and interest rates during the year 2003 - 2004. It is recommended not to expand more facilities at this point of time. The analysis of trend for the historical data shows that the long term trend have been gradually decreasing. The cycle, which is related with number of unemployed people who claimed unemployment insurance has been increasing over time, it is expected to decline as well.

Hence, it is concluded that the investment in new parks would not be profitable and would result in negative cash flows. As per the analysis, the sales of the mobile home shipments would decline over the years. Therefore it is highly recommended that Kim & Larry should wait to invest more money in additional mobile home parks until the US economy is recovered from recession.