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

**ISDS 526: Forecasting for Analytical Decision Making**

**Submitted to:** Dr. Dawit Zerom

**Submitted by:** Ruchika Narang

Sanchit Singh

Sebastian Stuchetz

**Table of Contents**

[**Executive Summary** 2](#_Toc527421069)

[**Background** 3](#_Toc527421070)

[**Objectives** 3](#_Toc527421071)

[**Interpreting Error Statistics** 3](#_Toc527421072)

[**Determining accuracy of forecasts with Time series** 4](#_Toc527421073)

[**Exploring Data** 4](#_Toc527421074)

[**Determining patterns in time series using Autocorrelation Analysis** 5](#_Toc527421075)

[**No differencing applied** 5](#_Toc527421076)

[**First - order Simple differencing applied** 6](#_Toc527421077)

[**First - order Seasonal differencing applied** 6](#_Toc527421078)

[**First - order Simple and Seasonal both differencing applied** 7](#_Toc527421079)

[**Conclusion & Recommendations** 8](#_Toc527421080)

# **Executive Summary**

In our project, we conducted a thorough analysis for the demand of mobile-home shipments in the United States based on detailed quarterly data collected from the period 1988 quarter 1 through 2004 quarter 4. We have utilized the application “Forecast Pro” to get actionable insights to our data, and thus, created accurate and credible forecasts.

As per requirements, we withheld 2 years of data and performed Expert Selection Method to produce forecast. After that, we evaluated the fit and accuracy of our model. To evaluate the fit, we utilized the Within - Sample Statistics of Forecast Pro and by observing MAPE, which is a measure of predicting accuracy of a forecasting method, and RMSE, which is a measure of how well our model performed, we can say that our model is good. Similarly, we utilized out - of sample stats for accuracy measures of RMSE or MAPE, which suggests that our model performed poorly for our dataset. We also used the autocorrelation coefficient to get better insights from our data.

The autocorrelation coefficient, which tells us the relationship between a variable's current value and its past values, does not show a sign of rapid decline to zero during the initial lags. This suggests that the series is not stationary. Hence, we implemented differencing to remove a potential trend and potential seasonality, if any. First, we applied simple differencing to remove the trend and hence make the time series stationary. After first-order simple differencing, the resulting correlogram showed a peak and a trough at every 2-lag indicating that there is seasonality but the series is still not stationary. Second, we applied seasonal differencing to check that our series changes from one season to the next. The early autocorrelation coefficients are very large, and they drop toward zero very slowly over time, followed by an increase which shows a multiplicative seasonality. Lastly, we applied both first order simple differencing and seasonal differencing, and we observed that there is no autocorrelation left, making the data stationary.

**Background**

Our forecast is based on the history of 17 years of data available for quarterly mobile-home shipments in the United States between quarter 1, 1988 and quarter 4, 2004.

# **Objectives**

We aim to find the error statistics by forecasting the data using expert selection method in Forecast Pro, interpret these statistics, and identify if our forecasts are accurate. We also aim at determining the data patterns to answer the questions such as, is the data trended, is the data seasonal, or is the data stationary? We did so by looking at the time series graph and confirming those patterns using Autocorrelation Function.

# **Interpreting Error Statistics**

A forecast error is the difference between an observed value and its forecast and, it is used to measure the accuracy of forecasts. Therefore, the lower the errors, the more capable the model is to predict future values.

For our purpose, we determined both Fit and Accuracy error measures. The ‘fit’ indicates how close the observed data points are to the model’s predicted values and are evaluated using ‘Within - sample statistics.’ ‘Accuracy measures’ evaluate how well the model works outside the period used to develop the model which means that ‘Out - of sample statistics’ are used for evaluation.

The tables found below summarize our dataset’s error statistics; fit (within sample) and accuracy measures (out of sample).

|  |  |
| --- | --- |
| **Fit Measures** | |
| **RMSE** | 4247.06 |
| **MAPE** | 5.57% |

|  |  |
| --- | --- |
| **Accuracy Measures** | |
| **RMSE** | 30084.22 |
| **MAPE** | 45.97% |
| **GMRAE** | 0.46 |

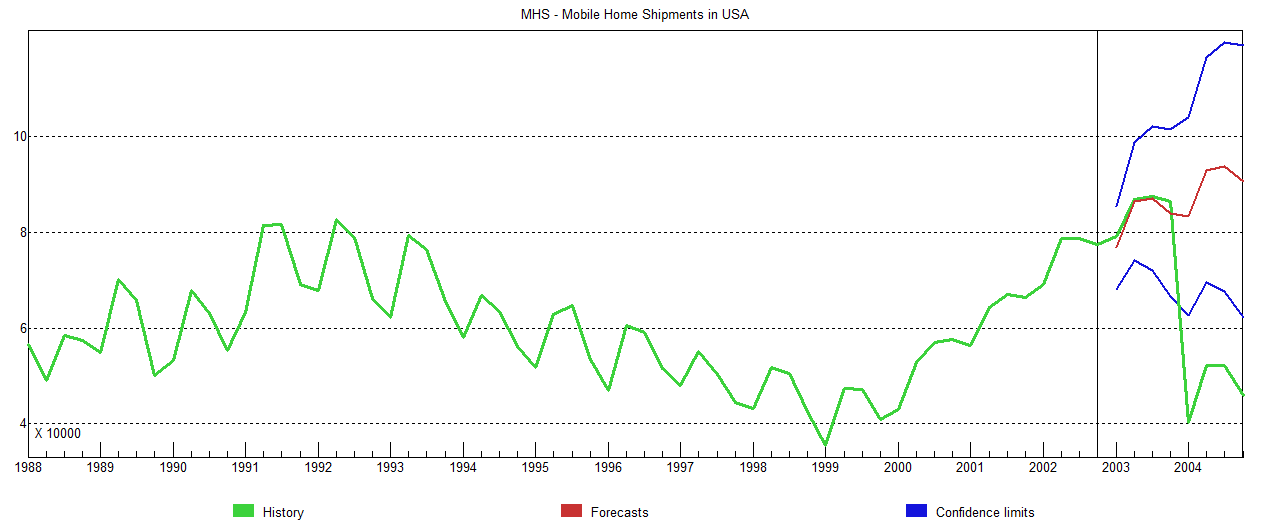
*‘MAPE, Mean Absolute Percentage Error’* measures deviation as a percentage of the actual data. A MAPE of 45.97% therefore, depicts that on average, quarterly forecast errors of our Expert Selection method are 45.97% of quarterly home-shipments. Thus, the accuracy of model is fairly bad. While considering the fit, lower value of MAPE, 5.57% depicts that on average, the deviation between the forecast of mobile home shipments and the actual value is 5.57% suggesting a good fit of the model.

*‘RMSE, Root mean squared errors’* measures squared forecast errors. By Squaring the errors, we can get more accurate results as the negative and positive errors don't cancel out each other and stay in existence till the end of commutation, thus adding more accuracy to the result (Quora, November, 2015). RMSE for our data is 30084.22 and 4247.06 respectively for accuracy and fit evaluation. (will here be units like shipments or not) which tells us that within the sample, the mobile home shipments deviate from the actual value by 30084.22 and out of sample by 4247.06. RMSE supports the evaluation results obtained from MAPE that the fit is good, however, the accuracy is undesirable.

*‘GMRAE, Geometric mean relative absolute error’* is used to compare the forecast performance of a given forecast model with that of the naïve model. GMRAE for our dataset is 0.46 which says that on average, quarterly forecast errors of the Expert Selection method are only 46% of the corresponding quarterly forecast errors of the Random Walk model (naïve model).

# **Determining the Accuracy of Forecasts with Time Series**

A time series is a sequence taken at successive equally spaced points in time. For our dataset, the below time series is a sequence taken at successive quarters beginning from quarter 1 of 1988 to quarter 4 of 2004 for mobile home shipments in the USA. The time series variable for the given data, I.e., Y is mobile-home shipments and the time period, I.e., t is 1988 – quarter 1 to 2004 – quarter 4. The **green** line is the historical data while the **red** line is the forecast in between the confidence limits line in **blue** color.



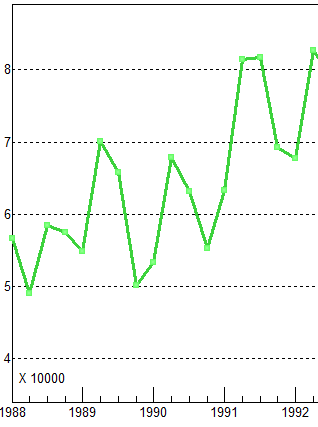
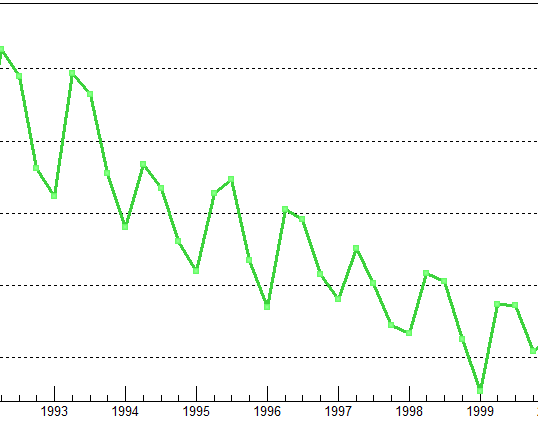
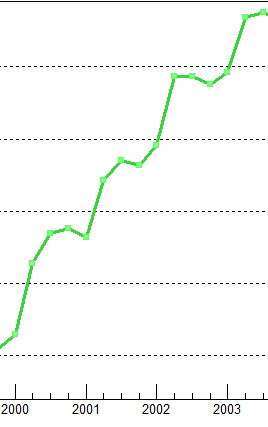
**Figure 1**

From the graph, we can clearly observe an unprecedented drop (green line) in the actual mobile home shipments from Q4 2003 to Q1 2004 in contrast to the forecasted increase (red line) during the same period. Thus, we can say that our forecasts are not completely accurate. This drop may be a reason for the high MAPE and RMSE which we evaluated earlier in the report.

As to why the number of mobile-home shipments in the US went down this rapidly resulting in undesirable accuracy of the model, there can be many reasons. According to a fact sheet, *‘Issues in Manufactured Housing’* published by ‘*AARP Public Policy Institute*,’ mobile homes are a major source of affordable housing for low or moderate income families. Therefore, this sudden depreciation might be due to the higher cost of mobile home shipments between quarter 4 of 2003 and quarter 1 of 2004 since after Q1 2004, there is again an increase in shipments. Another reason for this unexpected decline in mobile home shipments could have been concerns about quality and safety among customers and their changing tastes because mobile homes are not considered to be safe during abrupt events such as hurricanes. In fact, 2004 was predicted as the *‘Atlantic Hurricane Season*’ by *‘Colorado State University.’*

# **Exploring Data**

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.

**  **

**Figure 2**

From the above graphs, we determine that there is an increase during quarter 2 of every year (upward peaks) and a decrease during quarter 4 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 followed by a sudden drop which have been discussed earlier in report. Additionally, the time series is not constant over mean and variance.

These findings illustrate that our data is seasonal, trended, has cyclical pattern, and is non-stationary. Seasonality is supported by high rise during Q2s’ and depressions Q4s’ annually. Also, since the seasonal influences increase or decrease proportionally with increases and decreases in the level of the series, we can infer that the time series has multiplicative seasonality. Additionally, the time series shows cyclical properties and trends, since the shipment numbers increase for the first 5 years, then decrease for 2.5 years, followed by an increase again. Now, since these rises and falls are not of fixed periods, it can be inferred that a cyclical pattern exists. Non-Stationarity is depicted from the fact that the series is not constant over mean and variance.

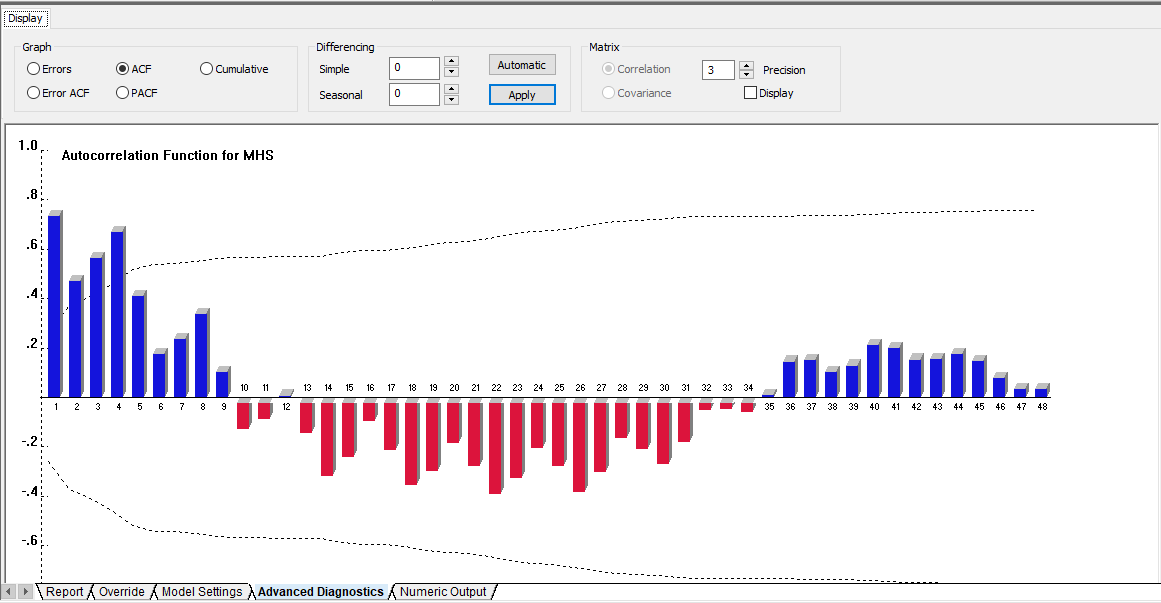
# **Determining patterns in time series using Autocorrelation Analysis**

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 visualizing the time series graph.

## **No differencing applied**

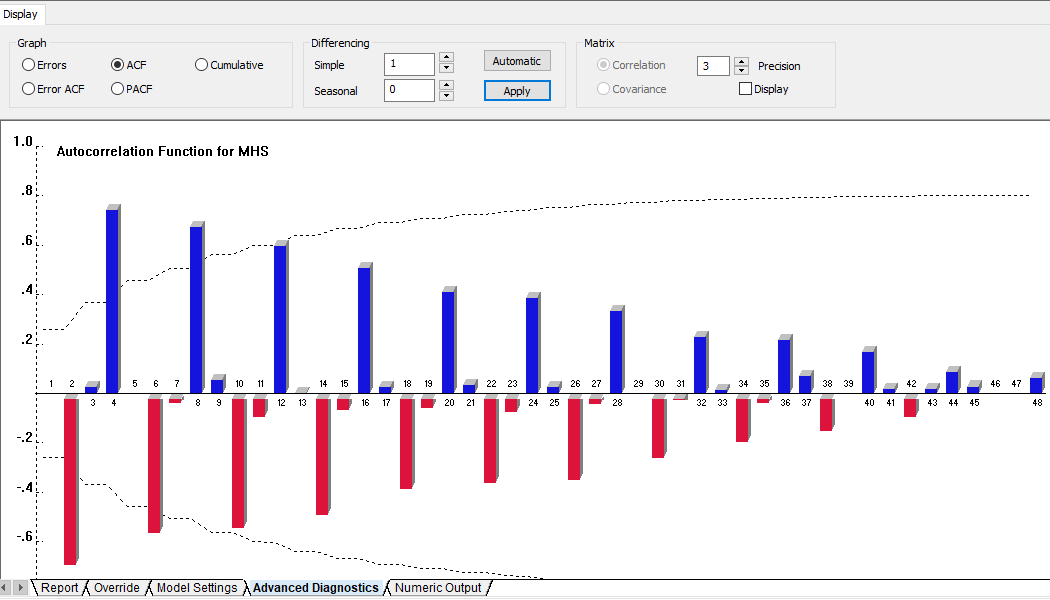
From the Figure 3 below, we examine that successive observations are highly correlated since the autocorrelation coefficients are significantly different from 0 for the first several time lags and then gradually drop toward zero as the number of lags increases. As a result, we can infer that the series has a trend. Furthermore, the autocorrelation coefficient gradually drops to 0 instead of rapidly dropping to 0 indicating that the series is non-stationary.



**Figure 3**

## **First - order Simple differencing applied**

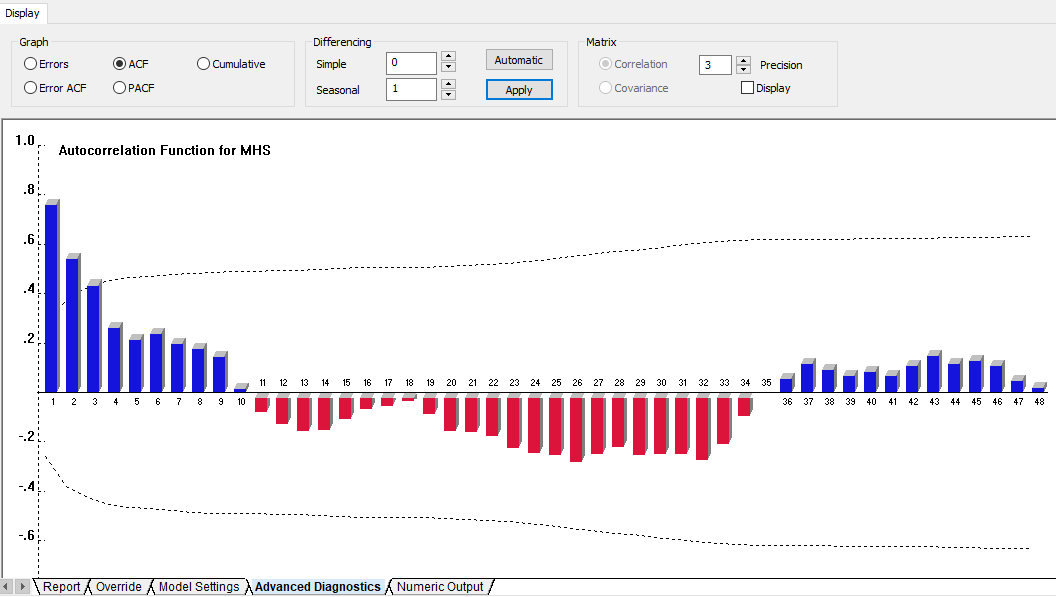
We can observe that there is a strong seasonality indicated by the high Q2 which is repeated annually across each Q2. A significant autocorrelation coefficient occurs in lags of 4 quarters.

****

**Figure 4**

## **First - order Seasonal differencing applied**

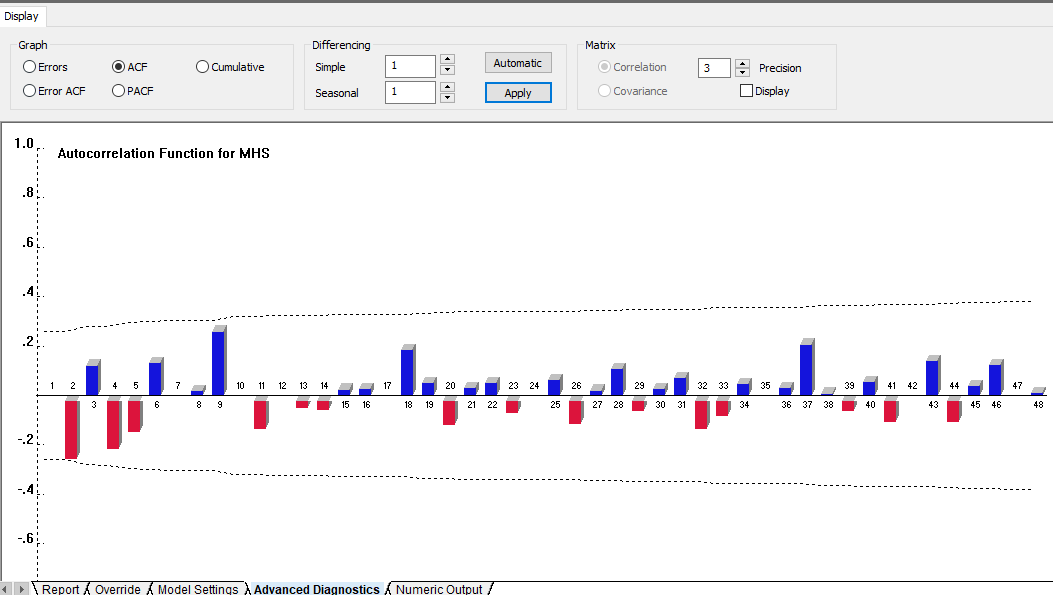
We observe that there is trend in the below figure which is of different periods, the series is cyclic.

****

**Figure 5**

## **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 are made the series stationary.

****

**Figure 6**

# **Conclusion & Recommendations**

Proper demand forecasting enables better planning and utilization of resources for business to be competitive. We have learnt through our analysis that to better understand which forecasting technique is best, we compared error measurements such as RMSE, MAPE, etc. and check how each method perform in order to fit the prediction with the time series patterns. Since we have seasonal and cyclical trend and also comparatively high RMSE and MAPE, we can eliminate naive model, simple exponential smoothing, hold exponential smoothing, simple moving average. We are only left with Box-Jenkins method and since our time-series outcome is stationary, this model is more appropriate to handle our data.

Our first recommendation would be to move at more promising location to find prospective customers for mobile homes. By investing correctly and with daily consistent effort and planning, a higher shipment numbers may be likely; however, this is only possible with a correct understanding of the market. Another recommendation is to apply energy-efficient and modern upgrades, thus, fulfilling the advanced technological demands and adding long-term value to the mobile homes without burning a hole in the pocket.