An Analysis on Monthly Supply of New Houses in United States

Buse Uzun
Undergraduate Student
Statistics Department
Middle East Technical University
Ankara, Türkiye
buse.uzun@metu.edu.tr

***Abstract*— This paper analyses the Monthly Supply of New Houses in the United States using time series approaches to obtain the best forecasting method. Since the Monthly Supply of New Houses is an important topic for the economy and wealth of society, it is important to foresee the future behaviour of the market so that the necessary actions can be taken to prevent a related issue.**

Throughout this paper, several tests were conducted such as KPSS, ADF and Hegy to ensure stationarity, anomaly detection and cleaning were applied, diagnostic checks were conducted for the SARIMA model to see the behaviours of residuals. Several forecasting models were used such as ETS, Holt Winters' Exponential Smoothing, Prophet, TBATS and Neural Networks. The best model was chosen with respect to their accuracy and plot performance.

Keywords—time series analysis, supply, houses, united states, demand, forecast

I. Introduction

Society is growing and improving rapidly, and improvement comes with change. The transition of people indicates the transition of their demands. Demand and price certainly go hand in hand, this rule applies to the supply of new houses. As demand rises, new houses will be built and sold increasingly. Somerville (1999, January), states that house prices have an impact on macroeconomic conditions. This is a crucial topic to understand the well-being of a society as the supply of housing can easily indicate the purchasing power of the citizens. This indication is possible due to sheltering being a basic necessity of humanity. A crisis in the housing market or supply can have catastrophic effects on the world order. Hence, countries need to follow market patterns closely to take necessary actions. Similarly, according to Joseph Gyourko (2009, September); to understand the price differences between metropolitan areas, it is important to inspect the supply conditions' heterogeneity in marketplaces.

In this report, the investigation will be the change in the monthly supply of new houses in the United States. Supply means a ratio of new houses on sale to new houses sold. Assuming there are no new houses built; the months' supply shows the duration of existing new inventory on sale until all of them are sold out. A low months' supply means that the market is in favour of sellers as there will be a lower inventory, the prices will be determined by the seller. On the contrary, a high months' supply is favourable for buyers as a greater inventory will provide negotiation power to the purchaser.

The aim of this paper is, by having past data on the monthly supply changes of new houses in the United States, finding the best forecasting approach for obtaining future data points by using several methods.

The data for analysis is obtained from the data library of Federal Reserve Bank of St. Louis, the source of the data is from U.S. Census Bureau and U.S. Department of Housing and Urban Development. Monthly Supply of New Houses in the U.S. (MSACSR) is a time series data with monthly frequency from January 1963 to October 2024.

II. Methodology

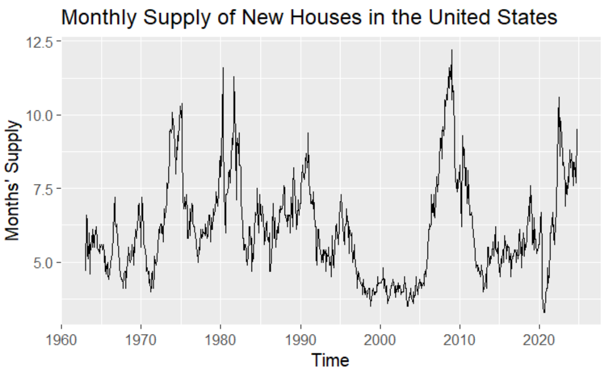
A. Data Cleaning and Interpretation

Before conducting the analysis, data was transformed from a data frame to a time series object. There were no NA values. Summary can be interpreted for a better understanding of the data (Fig 1).

Min.	1st. Qu.	Median	Mean	3rd Qu.	Max.
3.3	4.9	5.9	6.133	7	12.2

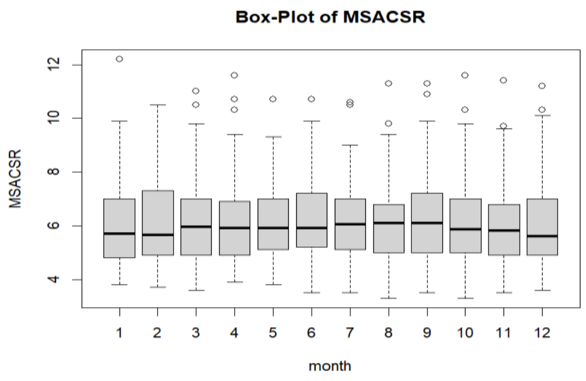
Fig. 1 Summary table of the MSACSR data

Minimum supply of new houses in a month is 3.3. First quartile value gives the 25th percentile of the data, monthly supply of new houses for the 25% of the data is 4.9 or less. Similarly, 50% of the data is equal to or below 5.9. 75% of the data is equal to or below 7, and lastly the maximum supply of new houses in a month is 12.2. Averagely, the monthly supply of new houses is 6.133.



Graph 1 Time series plot of MSACSR

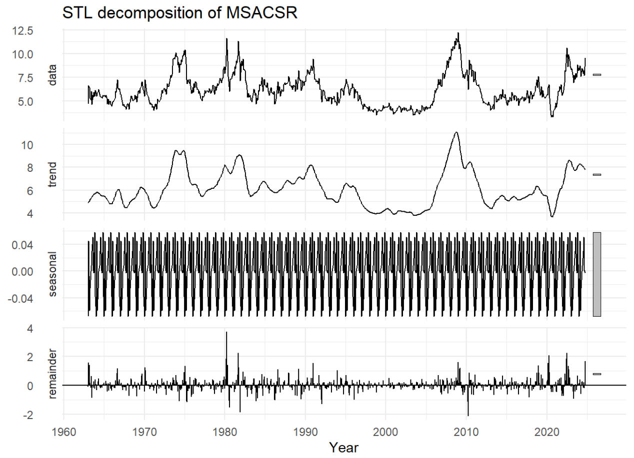
When graph 1 is observed, there seems to be seasonality. The mean varies lightly, variance is unstable. There can be a trend even if it is hard to inspect. A box plot of the data can be drawn to see seasonality.



Graph 2 Box plot of MSACSR

When the box plot of MSACSR for each month (Graph 2) is checked, means looks steady but there can be seasonality due to small variations.

STL can be used to split the series into components to make a more detailed observation of seasonality, trend and remainder.

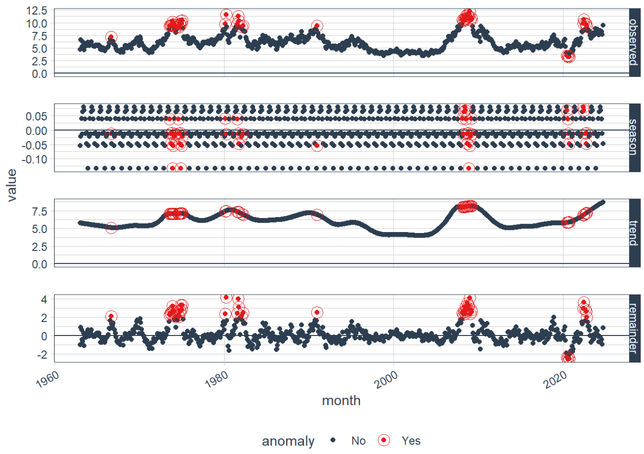


Graph 3 STL decomposition of MSACSR

When the seasonality graph is analysed, there is repetitive behaviour. Hence, there seems to be seasonality. From the remainder plot, some peaks can be observed which indicate outliers.

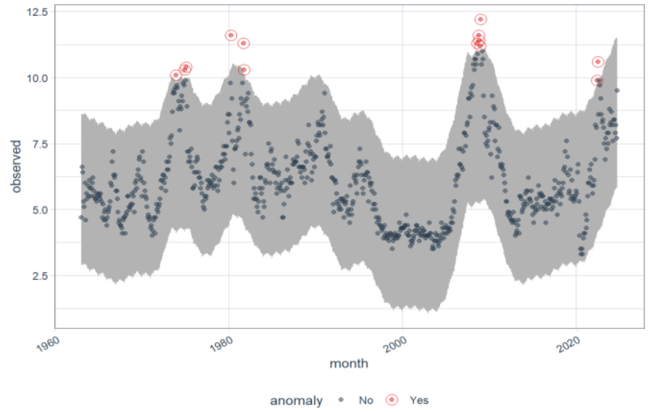
After these inspections, data is split into train and test sets for investigating the accuracy of forecasts in the following parts of the analysis. The last 12 data points are taken as the test set.

Analysis can be proceeded with anomaly detection to clean the data and investigate the outliers further.

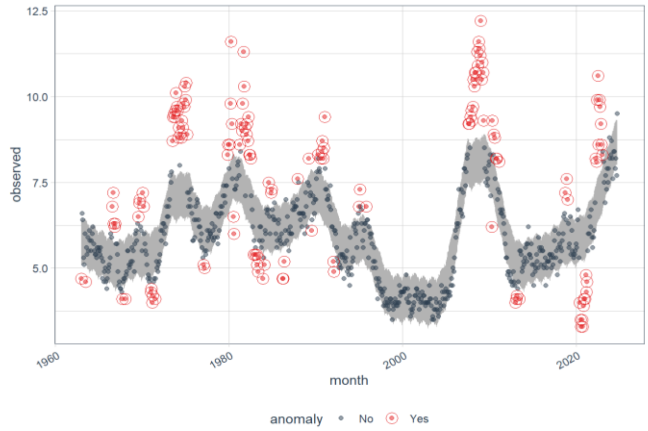


Graph 4 Anomaly plot of MSACSR

From graph 4, anomalies can be seen where the data points are red. There are several ways to show anomalies.

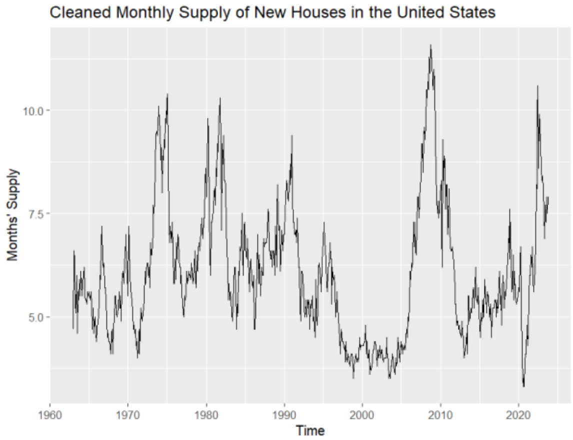


Graph 5 Anomaly plot of MSACSR (alpha=0.05)



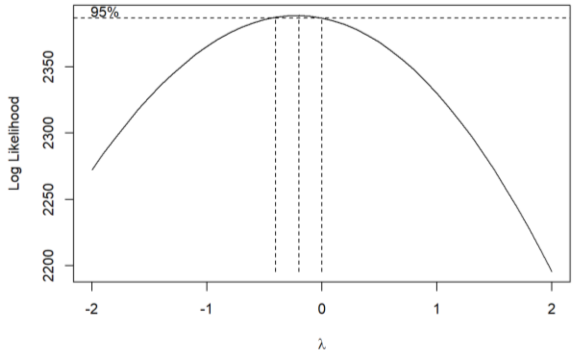
Graph 6 Anomaly plot of MSACSR (alpha=0.3)

As seen in graphs 5 and 6, for different alpha values anomalies will differ. Taking a bigger alpha value results in many outliers which is not desirable. The train set of MSACSR data is cleaned by using the tsclean function on R.



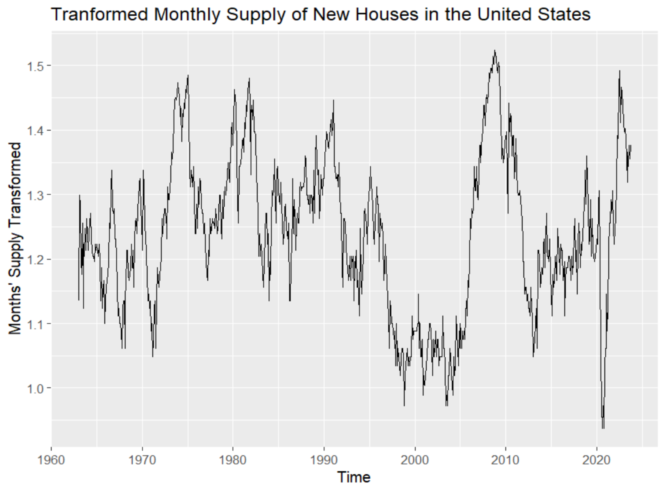
Graph 7 Plot of cleaned MSACSR Train

BoxCox plot should be checked to see if there is any need for transformation of data.



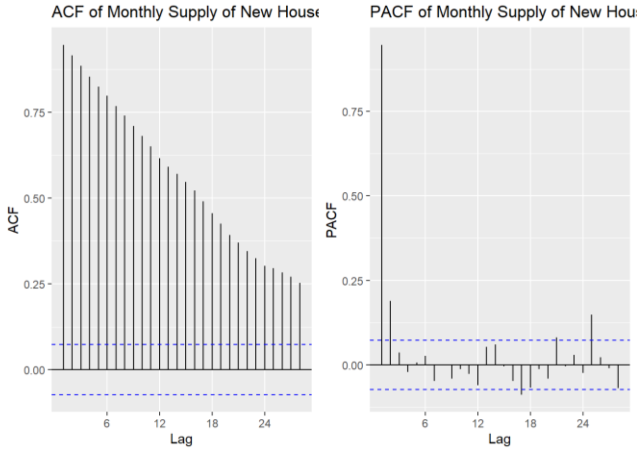
Graph 8 Box-cox plot of MSACSR Train

Since y axis has values with big differences on graph 8, a transformation should be applied. Lambda function of BoxCox helps finding the suitable lambda for transformation.



Graph 9 Plot of transformed MSACSR Train

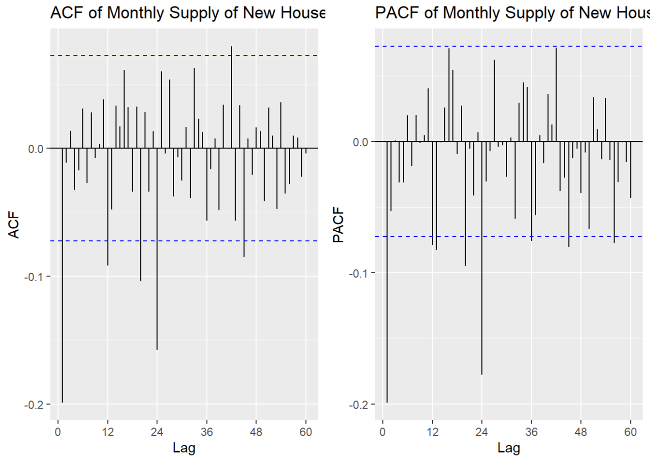
Data is cleaned and transformed. Hence, Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots can be interpreted.



Graph 10 ACF and PACF plots of MSACSR Train

Since the ACF plot has slow linear decay on graph 10, there is a non-stationarity problem. For inspecting stationarity and seasonality KPSS, ADF, HEGY and Canova and Hansen tests are applied. The results of these tests will be shown in the *Results and Findings* section. According to these tests, one regular difference is taken to make the series stationary. Graph 11 is obtained after stationarity was achieved.

B. *Identifying a SARIMA Model*



Graph 11 ACF and PACF plots of MSACSR

Regarding the ACF and PACF plots in graph 11, some SARIMA models are suggested. Out of 9 models, one model is chosen with the lowest AIC value. Diagnostic checks are conducted on residuals of the fitted model to investigate their normality, autocorrelation, homoscedasticity and outliers.

C. *Forecasting*

- Forecasting is applied to the fitted SARIMA model.
- Simple Exponential Smoothing (SES) could not be used since data has seasonality.
- Double Exponential Smoothing (Holt's Exponential Smoothing) similarly could not be used since the data has seasonality.
- ETS is applied to find the best exponential smoothing, but since it conducted the model as ETS(M,Ad,N), seasonality was not captured. ETS should not be used due to this. However, ETS was applied for the representation of its forecast.

- Holt Winters' Exponential Smoothing is applied with additive and multiplicative methods since data had trend and seasonality. The symmetricity of the data made it possible to disregard non-normality.
- Prophet is applied since there is seasonality, normality is assumed with symmetricity.
- TBATS model is applied, and normality is assumed with symmetricity.
- NNETAR is applied

Accuracies and plots of the above methods are obtained to make a comparison between all models.

III. Results and Findings

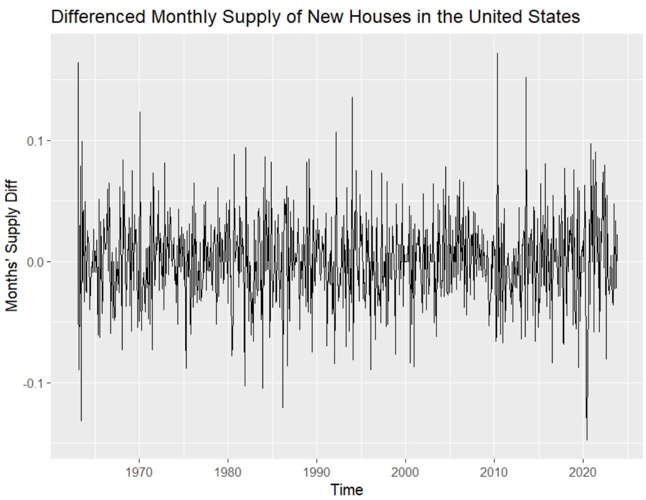
In this part of the paper, results of tests and taken actions according to them will be discussed.

A. *Obtaining Stationarity and Fitting SARIMA*

Making the data stationary is the primary concern. For this reason, the followings are conducted.

First, the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test for level and trend are found, according to these series was not stationary with the stochastic trend since their p-values were significant. Then, the Augmented Dickey-Fuller (ADF) Test was conducted, with a significant p-value according to ADF data was stationary. To see if there are unit roots Hegy Test was applied, this test resulted in no regular or seasonal unit roots in the data. Lastly, the Canova-Hansen Test was used to determine if the data was stochastic or deterministic. This test resulted in a non-significant p-value indicating that the series has deterministic and stationary seasonality.

Difference was taken since the series has a stochastic trend according to the KPSS test, some tests showed different results, but it is best to try as these tests might be misleading.



Graph 12 First order difference plot of MSACSR

After taking one regular difference, graph 12 is obtained. The graph seems stationary without any trend now. When the KPSS Test for Trend was checked non-stationarity problem was solved. ADF, Hegy and Canova-Hansen Tests conducted the same results as before. Now, after first-order differencing ACF and PACF plots can be interpreted to see seasonality and SARIMA model suggestions.

As seen in graph 11, according to ACF plot there is a significant spike at lag 1, but it can also be taken as an exponential decay. ACF has seasonal spikes at 12th and 24th lags. On PACF plot there is a significant spike at lag 1, and seasonal spikes at lags 12 and 24 again. Models such as SARIMA(1,1,0)(2,0,2){12}, SARIMA(0,1,1)(2,0,2){12}, etc. are suggested, as indicated before, out of the significant models the one with the lowest Akaike Information Criterion (AIC) value is chosen which is SARIMA(0,1,1)(2,0,1){12}.

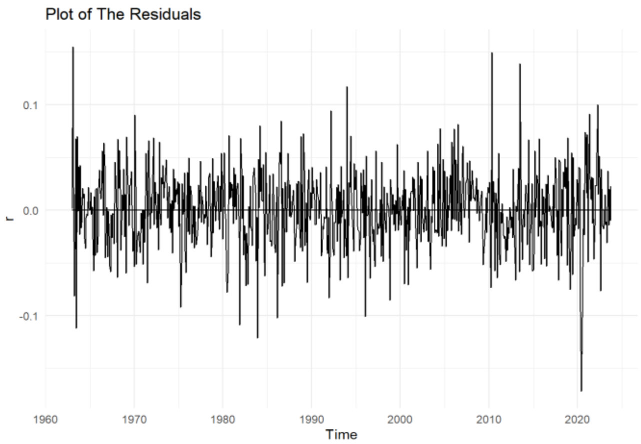
ARIMA(0,1,1)(2,0,1)[12]				
Coefficients:	ma1	sar1	sar2	sma1
	-0.2322	0.4763	-0.1126	-0.6295
s.e.	0.0367	0.1178	0.0490	0.1158
sigma^2 = 0.00138: log likelihood = 1367.13 [1]AIC=-2724.26 AICc=-2724.18 BIC=-2701.3				

Table 1: Summary table of SARIMA(0,1,1)(2,0,1){12}

Table 1 can be inspected as an example for summaries of SARIMA fits.

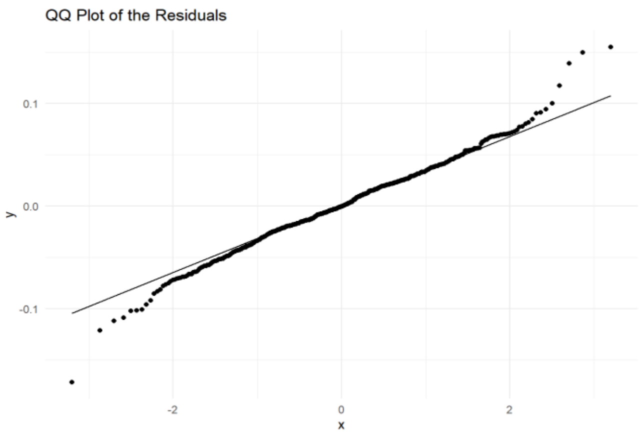
B. *Diagnostic Checking for SARIMA*

Residuals are obtained for diagnostic checking.



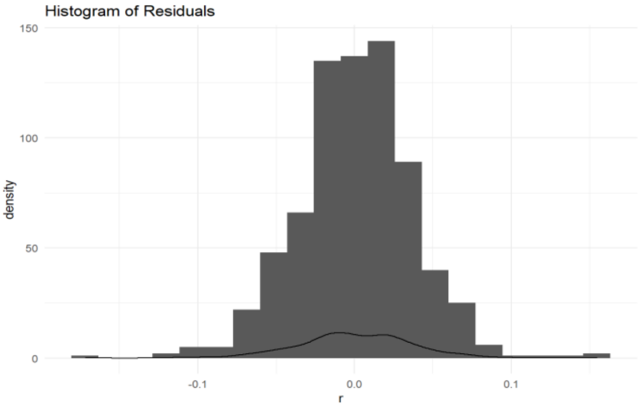
Graph 13 Plot of the residuals

As seen on graph 13, the residuals are scattered around zero. Hence, they have zero mean.



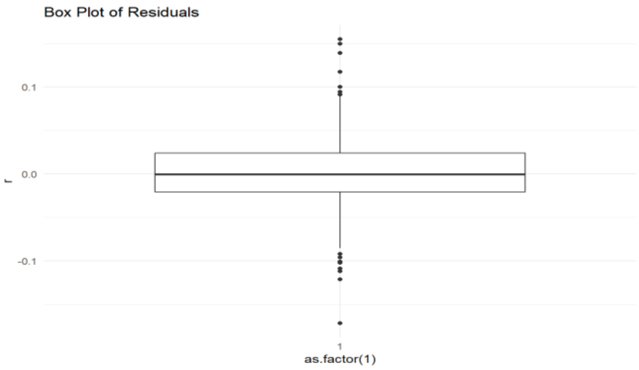
Graph 14 QQ Plot of the residuals

When graph 14 is interpreted, the residuals have a heavy tailed distribution due to the S shape. There can be some outliers too.



Graph 15 Histogram of residuals.

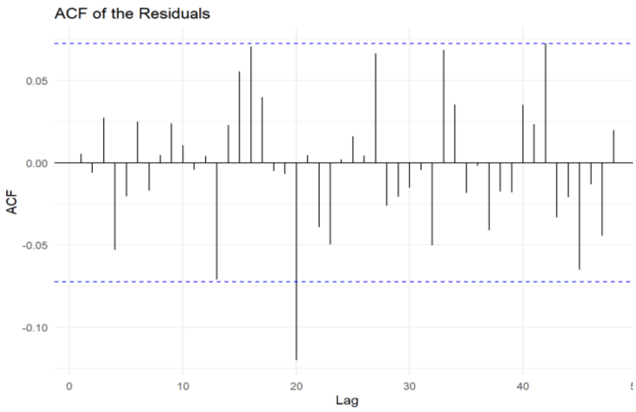
According to graph 15, there seem to be some outliers. The histogram does not look skewed.



Graph 16 Box-plot of residuals

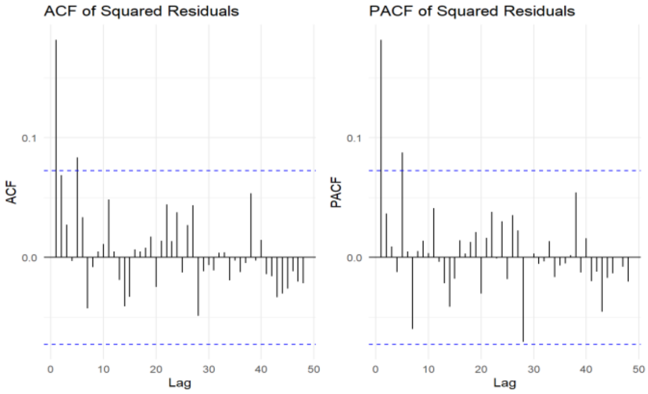
Since the box plot (graph 16) is quite symmetric data seems normal. There might be some outliers that can interfere normality of residuals.

Formal tests such as Jarque Bera and Shapiro-Wilk are applied to check the normality of the data besides previous plots. P-values of both tests were significant, it cannot be concluded that the residuals are normally distributed. However, since the histogram and the box plot show a symmetric distribution, analysis can proceed with forecasts even if they require a normal distribution.



Graph 17 ACF of residuals

Graph 17 is obtained to check the serial autocorrelation of the data. Since the majority of the spikes are inside the WN bands we can conclude that residuals seem uncorrelated. Formal tests which are Breusch-Godfrey, Box-Ljung, and Box-Pierce are also applied. It cannot be claimed that residuals are uncorrelated according to the Breusch-Godfrey Test. Both Box-Ljung and Box-Pierce Tests claim that the residuals are uncorrelated.



Graph 18 ACF and PACF plots of squared residuals

Graph 18 is obtained for detecting the heteroscedasticity of the residuals. Squared residuals are taken to obtain this plot. Breusch-Pagan and White Formal tests are conducted to check heteroscedasticity as well. Both of these tests conducted an insignificant p-value indicating that the residuals are homoscedastic. Engle’s ARCH test is obtained to determine if there is an ARCH effect present in the data. With an insignificant p-value, there are no ARCH effects according to this test.

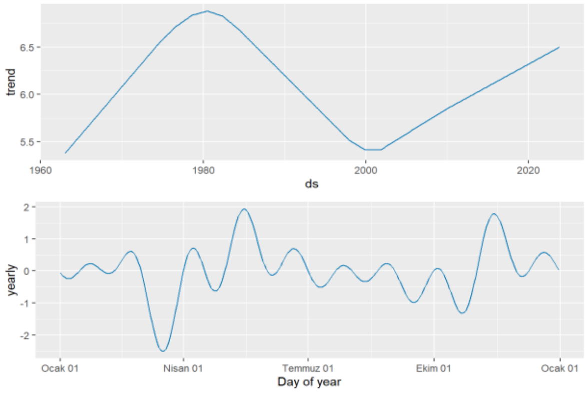
Since there does not seem to be heteroscedasticity in the data, there is no need to apply ARCH and GARCH methods. The analysis can be continued with other forecasting methods.

C. *Forecasting*

As stated in the methodology part, SES and Holt’s Exponential Smoothing methods are not used due to violation of assumptions. ETS should not be applied, but it was applied to show its forecast. The model ETS(M, Ad, N) was conducted with alpha=0.77, beta=3e-04, and phi=0.8.

Since data has a trend, normality (assumed by symmetricity), and seasonality Holt Winters’ exponential smoothing method is used. It is quite hard to determine the seasonality type of the data, hence, both additive and multiplicative methods are applied to see which is better. From additive model, smoothing parameters alpha = 0.79, beta = 1e-04, gamma = 1e-04; and from multiplicative model, alpha = 0.72, beta = 1e-04, gamma = 1e-0 are obtained.

Prophet and TBATS models work for seasonal and normal data, therefore they are applied. Due to Prophets’ bad performance, hyperparameter tuning is applied. It is found that changepoint.prior.scale=0.9, seasonality.prior.scale=0.1, and changepoint.range=0.9 gives the min RMSE, so Prophet parameters were changed accordingly which gave a much better forecast. BATS(0.007, {0,0}, 0.962, -) was derived from TBATS.



Graph 19 Broken down forecast of Prophet method

Graph 19 shows the forecast of Prophet broken down into yearly seasonality, weekly seasonality and trend.

Lastly, NNETAR was applied which is a Neural Network model. NNAR(25,1,13)[12] were collected, but its' neuron number in hidden layers was changed from 13 to 15 as it gave better accuracy results. When residuals of ETS, Holt Winters', Prophet, and TBATS was checked for normality with Jarque Bera and Shapiro Wilk tests, none of them showed normality.

All the above models can be compared based on their accuracy.

	ME	RMSE	MAE	MPE	MAPE	MASE
SARIMA	0.022	0.499	0.373	-0.132	6.071	
ETS	-0.001	0.537	0.391	-0.433	6.328	0.345
Ad. HW[a]	0.012	0.537	0.394	-0.210	6.378	0.347
M. HW[b]	-0.012	0.538	0.393	-0.645	6.376	0.347
Prophet	-0.007	0.811	0.598	-1.965	10.009	
Prophet[c]	-0.007	0.813	0.599	-1.987	10.034	
TBATS	0.001	0.512	0.380	-0.632	6.215	0.339
NN	-0.001	0.178	0.132	-0.204	2.337	0.116

Table 2: Train set accuracy of model forecasts

a. Holt Winters' Additive Model

b. Holt Winters' Multiplicative Model

c. Prophet with Hyperparameter Tuning

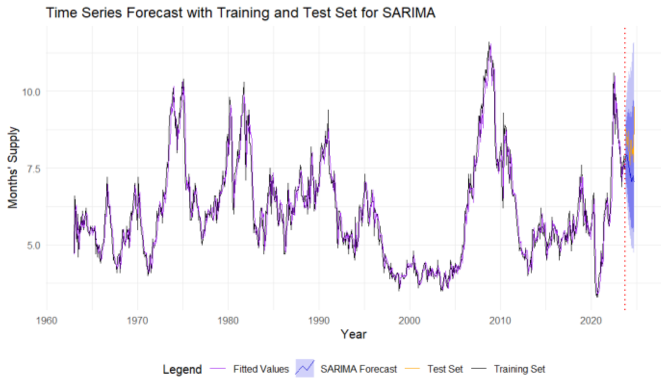
	ME	RMSE	MAE	MPE	MAPE	MASE
SARIMA	0.906	1.065	0.906	10.622	10.622	10.622
ETS	0.502	0.703	0.560	5.712	6.468	0.493
Ad. HW[a]	0.550	0.734	0.598	6.297	6.925	0.527
M. HW[b]	0.507	0.706	0.583	5.780	6.770	0.514
Prophet	1.965	2.047	1.965	23.323	23.323	
Prophet[c]	0.807	0.971	0.810	9.375	9.405	
TBATS	0.912	1.050	0.912	10.663	10.663	0.814
NN	1.529	1.699	1.529	18.079	18.079	1.348

Table 3: Test set accuracy of model forecasts

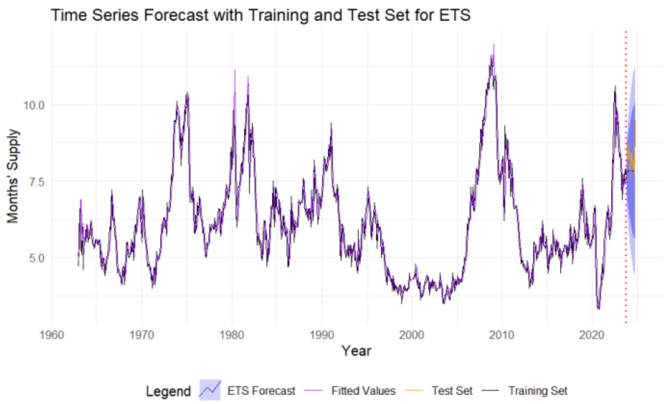
When table 2 is observed, the test set accuracies of the ETS model has the minimum RMSE and MAPE values. Therefore, it has the best forecasting performance compared to other forecasts. Since it is not actually a good idea to fit this due to not capturing seasonality, Holt Winters' Multiplicative model is the second-best forecast.

When table 1 and 2 are compared, overfitting issues can be seen in Prophet, TBATS, and NNTAR since their train set accuracies are considerably smaller than the test set accuracies.

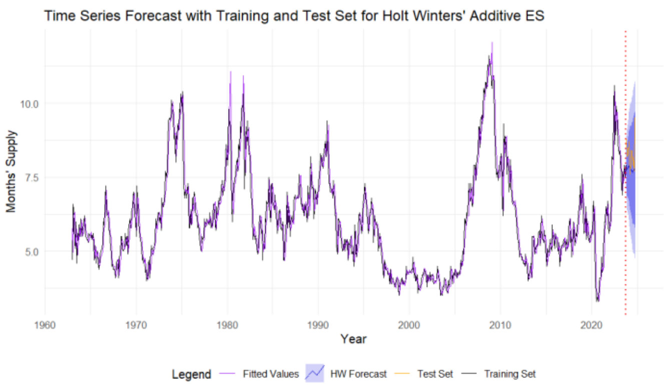
Models can also be visually inspected by the following plots:



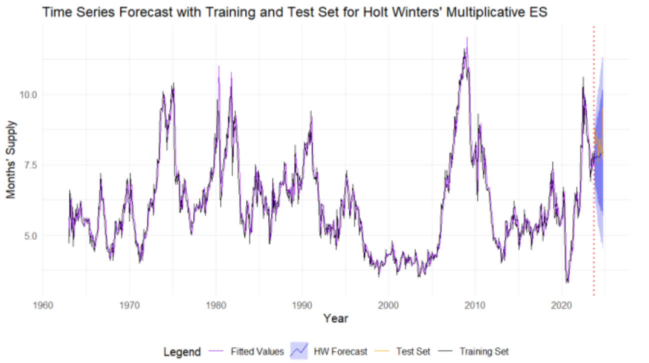
Graph 20 Plot of SARIMA forecast (SARIMA(0,1,1)(2,0,1){12})



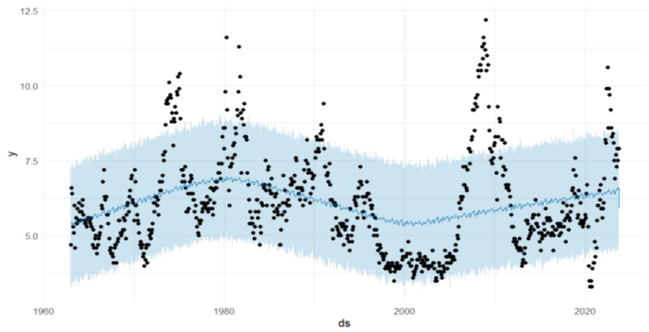
Graph 21 Plot of ETS(M, Ad, N) forecast



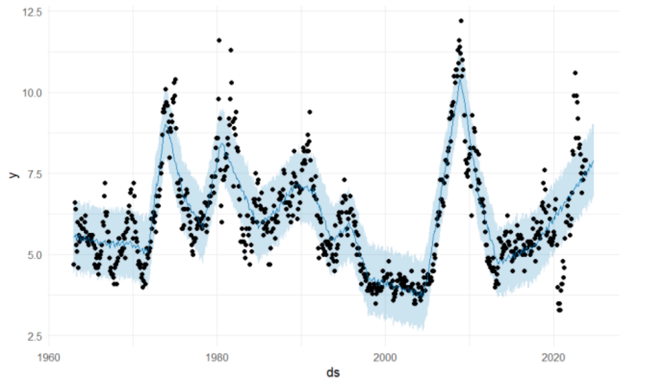
Graph 22 Plot of Holt Winters' Additive exponential smoothing forecast



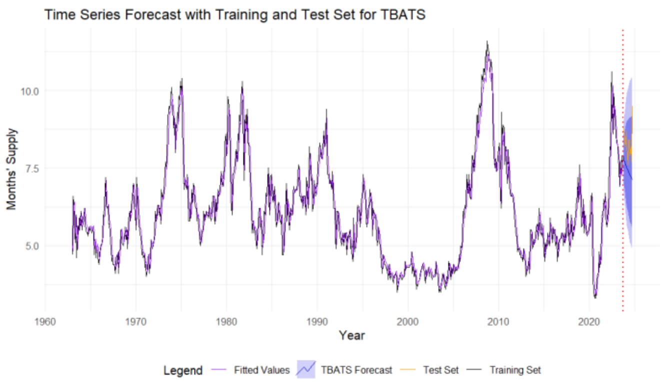
Graph 23 Plot of Holt Winters' Multiplicative exponential smoothing forecast



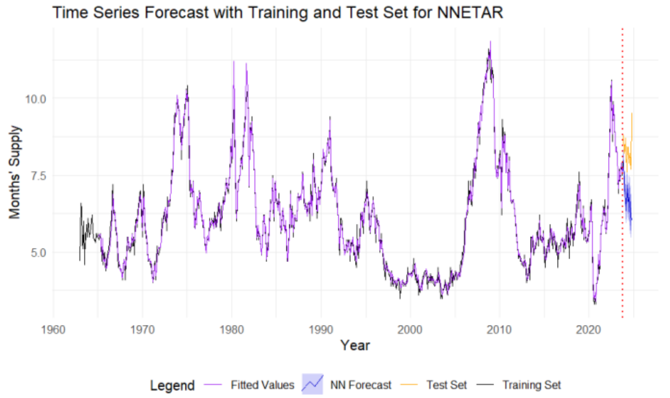
Graph 24 Plot of Prophet forecast



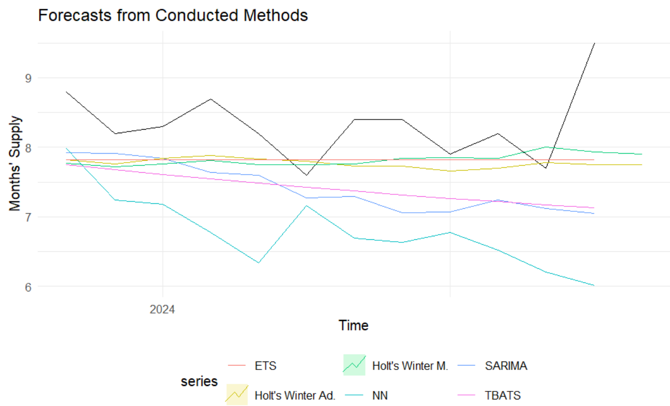
Graph 25 Plot of Prophet forecast with Hyperparameter Tuning



Graph 26 Plot of TBATS forecast (BATS(0.007, {0,0}, 0.962, -))



Graph 27 Plot of NNETAR forecast (NNAR(25,1,15)[12])



Graph 28 Forecast plot of Conducted Methods

Note: Prophet is not included due to formatting

When graph 20 is inspected, the prediction interval (PI) of SARIMA seems wide, its forecast does not seem too bad. Graph 21 shows the ETS forecast, its PI is wide, but the forecast is great. From graphs 22 and 23 it is seen that Holt Winters' models are good at fitting the forecast, but PI still seems quite wide. Prophet can be inspected from graph 24, it does a very bad job since there is no fit and only the general trend is followed. Graph 25 with Hyperparameter Tuning applied Prophet shows a much-improved forecasting performance. When graph 26 is analysed, TBATS is predicting the train well with a narrower PI on the forecast, its forecast is quite far off. Lastly, from graph 27 NNETAR forecast can be seen. It predicts the train set wonderfully, but since it has a visibly bad forecast an overfitting is highly possible.

The plots of TBATS and NNETAR seem like a good fit for the train set, but their performance with forecasting shows that there can be an overfitting problem with these approaches.

Graph 28 shows all the forecasting methods except Prophet (which was a bad fit anyway). From this graph it can be seen that the closest forecast line is again ETS and Holt Winters' Multiplicative method, NNETAR conducts the wors forecast.

IV. Conclusion

This paper examined the MSACSR series to enable opportunities for forecasting to this data. In this detailed analysis, forecasting models such as SARIMA, ETS, Holt Winters', Prophet, TBATS and NNETAR were used. Their accuracy and plots were checked to ensure the best forecasting model.

Both accuracy and plot results supported each other for the best model, which is the ETS method. However, it is better to use the second-best forecast, which is Holt Winters' Multiplicative model as ETS could not capture the seasonality. Even though the prediction of models like TBATS and NNETAR are well fitting to the train set, it is seen that their forecasting performance can be equally worse. That is why a comprehensive analysis is better for choosing the best approach to not fall into overfitting issues.

Future analysis can be conducted to ensure the model and even find a better one as foreseeing the nature of the future values will enable better preparedness for possible disasters.

REFERENCES

[1] Dipasquale, D. (1999). Why Don't We Know More About Housing Supply?. *The Journal of Real Estate Finance and Economics* **18**, 9–23. https://doi.org/10.1023/A:1007729227419

[2] Gyourko, J. (2009). Housing Supply. *Annual Review of Economics*, *1*(1),295-318. https://doi.org/10.1146/annurev.economics.050708.142907

[3] Somerville, C.T. (1999). Residential Construction Costs and the Supply of New Housing: Endogeneity and Bias in Construction Cost Indexes. *The Journal of Real Estate Finance and Economics* **18**, 43–62. https://doi.org/10.1023/A:1007785312398