Utah Housing Market

Danielle Yoseloff

I am interested in property investment in the Salt Lake City area. My goal is to predict where the Utah housing market is headed and at what rate. The most conventional way to research this topic is a measurement called housing price index (HPI). I reviewed the Federal Housing Finance Agency's webpage and used their Master HPI Data set which I specialized to contain only the information I was interested in. I found the best model to predict the market is a SARIMA(1,1,4)x(1,1,0)[4]. With 95% confidence, the housing price index in Utah is expected to be between 395 and 481 points by the end of 2017 (figure11). Most of the points in this range are an increase from the current HPI, indicating that property is a valuable investment.

Table of Contents

1) THE DATA

i. Finding Data

I am interested in the Housing Price Index (HIP) of Utah. The Federal Housing Finance Agency meticulously tracks the US housing market and makes the information available under the Housing Price Index Datasets page online (Resource1). The dataset that best suits my needs is the first one on the page titled "Data" under the category "Master HPI Data (Appends Quarterly and Monthly Data)" (Resource2).

ii. Cleaning Data

The original dataset includes 10 variables: HPI type, HPI flavor, Frequency, Level, Place Name, Place ID, Year, Period, and Index of Seasonally adjusted, Index not seasonally adjusted. To clean the data, I took the following steps:

1. Set the place Id variable to UT
2. Set the HPI flavor variable to all-transactions
3. Set the HIP type variable to traditional
4. Remove Index of Seasonally adjusted variable
5. Remove redundant variable columns: HPI type, HPI flavor, Frequency, Level, Place Name, Place ID, Year, Period.

iii. Description

The clean dataset is quarterly unadjusted Housing Price Indexes from the year 1975-2016 for Utah, where HPIs are traditional and all-transactions. There are 168 observations, one HPI value per quarter for 42 years. The data is sufficiently long to draw conclusions about the next few periods.

2) MODEL SELECTION

i. First Look

The plot of the time series (figure1) shows a positive linear trend, so the mean is non-constant. This is the first indication of non-stationarity. There is no real change in the variance across time. In the few years following 2008 there is a dip in the HPI which is an abnormality in the trend. However, we expect this discontinuity as a result of The Great Recession when the housing market took a great hit nationwide. The ACF (figure2) decays very slowly, never reaching below significance. The PACF (figure2) is almost 1 at lag 1 and extremely insignificant elsewhere. The features of the ACF and PACF also indicate non-stationarity.

ii. Transformations

As we just saw, the variance of the time series looks constant, so a Box-Cox transformation is not needed.

iii. Testing Stationarity

To test stationarity is essentially testing for the presence of a unit root. For an Augmented Dickey-Fuller test the null hypothesis is nonstationarity and the alternative hypothesis is stationarity. In this case the p-value is .04848, which is less than .05 so, we conclude the data is stationary. This contradicts the behavior of the graphs. For a second opinion, I ran a KPSS test. For this test the null hypothesis is stationarity and the alternative is non-stationarity. Here, the p-value is less than .01 so we confirm the presence of a unit root. Since the results of these tests conflict, I will pursue both the instance of stationarity and non-stationarity to carry out model selection.

iv. Stationarity

A. Classical Decomposition

A classical decomposition is a way to estimate the trend and seasonality of data. Selecting models based on the ACF and PACF of the random component of a classical decomposition is an alternative to model selection from the ACF and PACF of the regular data. In this case an additive seasonal decomposition is most appropriate because the amplitude of the variance is not increasing in the time series plot and there is a clear seasonal pattern (figure3). The ACF (figure4) seems to cut off after lag 3 suggesting MA(3). The PACF either cuts off after lag 1. So, we look at ARMA(1,3) as a candidate. See Table1, Model7. Its' AICc value is very high indicating it is not a good model to purse. Goodness of fit can also be seen in an analysis of the residuals. Figure5 shows a patterns in the residual plot, which is problematic and significant lags on the ACF residual plot which is also an indicator of trouble observations. The p-values of the Ljung-Box statistic cannot be calculated due to the size of the psi weights. So far all methods of analysis say that this model is not a good fit. One more way to check is to run the automated fitting function on the residuals of the fitted data. This should result in an ARMA (0,0,0) however in this case is does not (Table1, Model7). So after fitting this model to the data there is still information to be fitted. Nothing indicates model 7 should be pursed further.

B. Automated

In a second attempt to find a model, assuming the data is stationary, I will try an automated method on the data. This suggests the best model is an SARIMA(1,1,0)x(2,0,2)[4], its information can be found in Table1, Model8. The AICc model criteria is large so it will probably not compare to the goodness of fit of other models. The residual plot for this model (figure6) forms a clear pattern around 2004-2010 indicating an issue with the model. An issue is also present because there are significant values of the ACF of the Residuals. The p-values for Ljung-Box statistic, furthermore shows that the model is not a good fit to the data because the p values are almost significant.

v. Non-stationarity Approaches

A. Differencing

To remove the unit root that is present we must difference the data. In this case a seasonal differencing of frequency 4 for the quarterly data is most appropriate. The times series plots of the raw and differenced data (figure7) shows improvement of nonstationarity because now the mean is constant at zero. Comparing the ACP and PACFs (figure 8) also shows improvement because the plots now have the form where models can be suggested by analysis of the plots.

B. Models via Plots

The ACF and PACF of the seasonally differenced data will implicate the best models to fit (figure9). First, I'll establish the seasonal component of the model. The ACF is significant before the first seasonal lag but at the first seasonal lag it appears insignificant. Therefore, I've decided a seasonal MA(0) component is appropriate. The PACF shows significance at lag 4 but not at the seasonal lags beyond so seasonally the model is AR(1). Now for the non-seasonal component of the model. The ACF appears to cut off at lag 4, although lag 5 is almost significant so we should consider MA(4)and MA(5) for the non-seasonal portion of the model. Another interpretation is that the ACF exponentially decays to zero so we should also consider MA(0). The PACF can be interpreted a few ways. It could be seen that lag 1 is the only significant lag, or that lags cut off after lag 4, or that lags are significant up to lag 7. So, we should consider AR(1), AR(4), AR(7) for non-seasonal components. At a glance of the plot that is most probable model seems to be (4,1,4)x(1,1,0)[4]. In total we will fit these 6 models to the data and compare them (Table1 Models1-6).

The first method of comparing the models is via the AICc criteria, the lower the better. Based on this Model3 and Model4 are the best. However, a comparison of the 6 time series diagnostics (figure10) shows Model1 as the best fit. All residual plots appear to follow the normality assumption then follow a tight sine wave around 1990 on. This indicates a lack of fitness but because it occurs in all models it is reasonable to assume this is due to a lack of adjustment for the recession. The ACF of the residuals for all models look good because there are no obvious outliers, although the lags are most moderate for Model1. The p-values of the Ljung-Box statistic are the clearest indicator that Model1 has the best fit to the data. None of the p-values of Model 1 are significant whereas all other models have at least one significant p-value. This means, only the residuals of Model1 are following a white noise.

The automated modeling function should produce an ARIMA(0,0,0) when the residual data that is passed is already well fit. Therefore, another way to check the goodness of fit for our 6 models is to run the automated function on the residuals for the fitted model in question. The results of this can be seen in "auto.arima" column of Table1. This method of fitness claims the best models are 3,4,5 and 6.

In all, Model1 and Model3 seem to be the best contenders because they both have criteria that support them to be a good fit. By convention it is better to go with a simpler model, so we will pursue forecasting of Model1 later on.

C. Subset Modeling

In this approach to finding a model for the data, I will search through subsets of the seasonally differenced data then fit and assess the model. I cap the max p and q values at 13 because the ACF and PACF (figure4) leads me to believe no more than a 3 years of lags need to be considered. This method suggests a (5,1,12)x(0,1,0)[4] model (Table1 Model9) with significant components at MA(1), MA(3), MA(5), AR(2), AR(3), AR(4), AR(8), AR(12). (figure13). The AICc indicates it is a poor model and the model diagnostics show some problems too. The residuals are not normal (figure14), there is a pattern in residuals plot although this issue has been noted and may not be resolved. The ACF of the residuals has problem lags at period 2 and 3 (lags 8 and 12). Also, the residuals do not follow a white noise according to the Ljung-Box statistic. Another way to analyze the behavior of the residuals is via a qqplot (figure15). The residuals do follow a straight line as they should for normal behavior but the tails stray a lot. This is cohesive with our other findings; the residuals cannot confirm goodness of fit. However, the automated modeling function does suggest a null model for the residuals. So while the model passes the R 'tests', the graphs indicates something deeper is not being captured.

D. Automated modeling

The final method of selecting a model assuming the data needed unit root adjustment, is automated modeling. This method produces (3,1,0)x(2,1,0)[4] as the best model(Table1 Model10). The model selection criteria (AICc) is poor compared to our best model thus far Model1. The model diagnostic (figure12) shows that the model is a very poor fit. All Ljung-Box p-values contradict that the residuals are white noise. In addition, there are several significant lags on the residual ACF. Checking validation via fitting the residuals further disproves any goodness of fit because the automated function indicated the residuals should be fit with a (0,0,0)x(2,0,2)[4] model when we hoped it would suggest a null model as the best fit.

3) FORECASTING

i. Model 1

Above, I found the best model to be Model 1. I forecasted the next 3 years and see a positive trend (figure11). The error of the confidence intervals gets wide very quickly though. At a 95% confidence, Utah HPI is expected to change anywhere from a decrease of 6 points to an increase of 24 points next quarter (the first quarter of 2017). While the trend is expected to be overall positive the wide standard error predicted that by the last quarter of 2019 the HPI could change from its current standing to between a drop of 58 points and an increase of 316 points. It is also important to keep in mind that this analysis is done without adjusting for a recession. Without taking this abnormality into account it is difficult to produce models with more focused predictions. While this kind of a prediction is not accurate enough to place bets, it still gives an idea of where the market is headed. The HPI is a complicated measurement with many implications but a to make it simple knowing the forecast predicts increase in leads to trends in the market like increased home prices and decreased mortgage interest.

ii. Other Forecasting

Here, I will showcase comparing forecasts via root mean square error and mean average percent error criteria. It is convention to set aside 10-15% of the latest data as test data. In this case I will use 12 observations, 3 years of data. The rest of the data is the training data.

A. Training Data Subset

Using subset selection on the training data, an ARMA (5,1,12)x(0,1,0)[4] with significant values as MA(1), MA(4), MA(5), AR(3), AR(5), AR(8), AR(12) is suggested as the model that best fits the first 156 observations (figure18). This model is similar to the subset model that was previously fitted on the differenced version of the full data as expected, but varies on a few of the significant components. A forecast of this model for the next 6 years can be seen in figure19. The error of prediction is very wide. The model is a fair fit because the 95% forecast intervals of the model do contain the observed values that were set aside as test values (figure20). However, the observed values are not in the middle of the confidence interval which indicated the model will be good at predicting the trend of the HPI but will not be very accurate the further out predictions are.

B. Training Data Holt Winters

Holt Winters is a type of smoothing to the data. It is similar to a simple exponential smoothing with large values of lambda (figure16). The holt winters is better in this case because it better factors into account the non-constant mean. It can be seen that the Holt winters forecast has a sharper fit. In this case we let gamma be true because there is a seasonal trend. The seasonal type is additive because amplitude of variance does not increase over time. The Holt winters forecast for the next 6 years has a much steeper increase than the subset model forecast (figure17). Comparing the 95% forecast interval with the observed values of test data (figure20) all forecast fall in the middle of the interval. This indicated the holt winters method of forecasting is accurate.

C. Comparing

Comparing the subset model forecasting and the holt winters forecasting can be done using error criteria, the smaller the error the better. RMSE is the root mean square error which measures the accuracy of the model to the test data. MAPE is the mean average percent error. This method of describing the accuracy of the model is produced as a percentage. In both cases, the smaller the errors the better the model is at predicting the values that were observed. Above we observed that the Holt winters forecast was slightly better at prediction, however a comparison of the errors favors the subset model (Table2). There are a few reasons this could be the case. The method of MAPE always favors forecasts that are low. We observed that both methods contained the observed values but the subset model confidence intervals forecasted lower observations. So, we could expect the MAPE value for the subset model to be lower. The most concerning issue is the great discrepancy in values between the two forecasting methods. Overall, we expect the RMSE to be lower for the holt winters. This is most likely not the case because of an issue in the coding.

APPENDIX

RESOURCE 1:

https://www.fhfa.gov/DataTools/Downloads/Pages/House-Price-Index-Datasets.aspx#mpo

RESOURCE 2:

http://www.fhfa.gov/HPI_master.csv

TABLE 1:

ID	MODEL	AICc	auto.arima
1	(1,1,4)x(1,1,0)[4]	-810.33	(3,0,1)x(2,0,2)[4]
2	(1,1,5)x(1,1,0)[4]	-808.25	(3,0,1)x(2,0,2)[4]
3	(4,1,4)x(1,1,0)[4]	-815.79	(0,0,0)
4	(4,1,5)x(1,1,0)[4]	-813.8	(0,0,0)
5	(7,1,4)x(1,1,0)[4]	-809.44	(0,0,0)
6	(7,1,5)x(1,1,0)[4]	-798.07	(0,0,0)
7	(1,0,3)x(0,0,0)	926.88	(0,0,0)x(2,0,1)[4]
8	(1,1,0)x(2,0,2)[4]	854.72	(1,0,0)x(1,0,1)[4]
9	(5,1,12)x(0,1,0)[4]	841.44	(0,0,0)
10	(3,1,0)x(2,1,0)[4]	887.16	(0,0,0)x(2,0,2)[4]

TABLE 2:

	Holt Winters	Subset (5,1,12)x(0,1,0)[4]
RMSE	44.66117	23.44326
MAPE	8.89687	4.990669

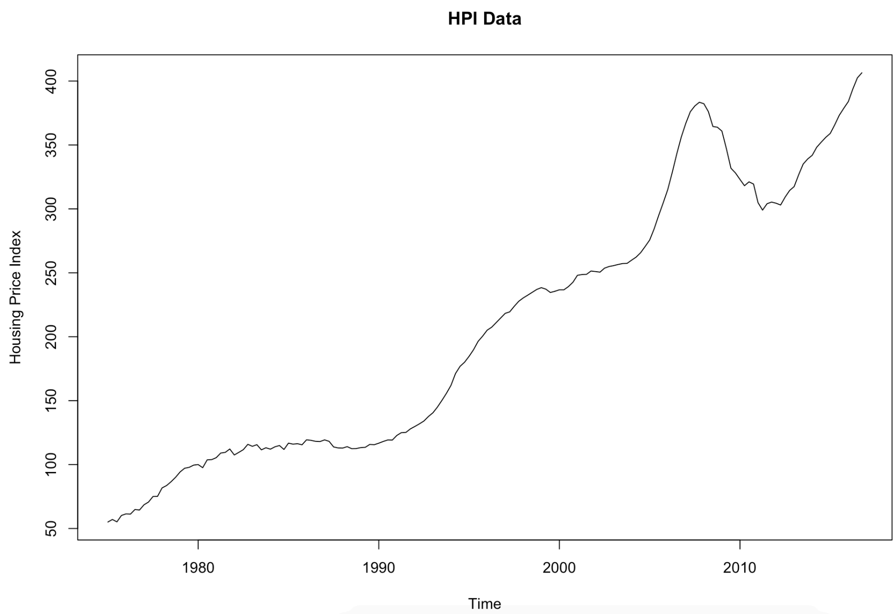


FIGURE 1

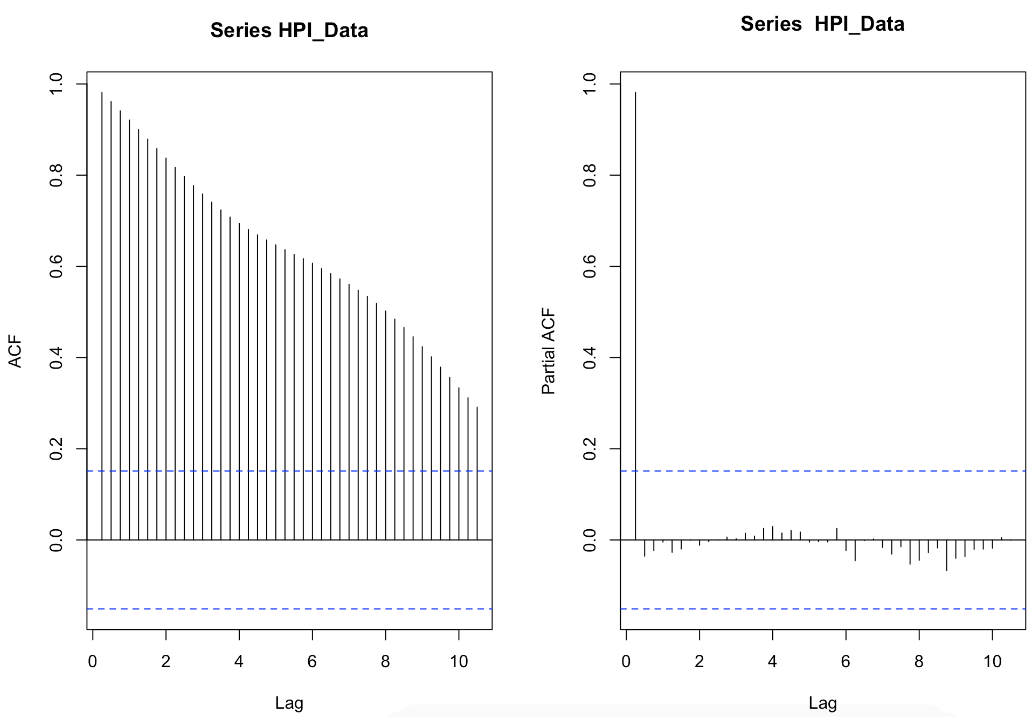


FIGURE 2

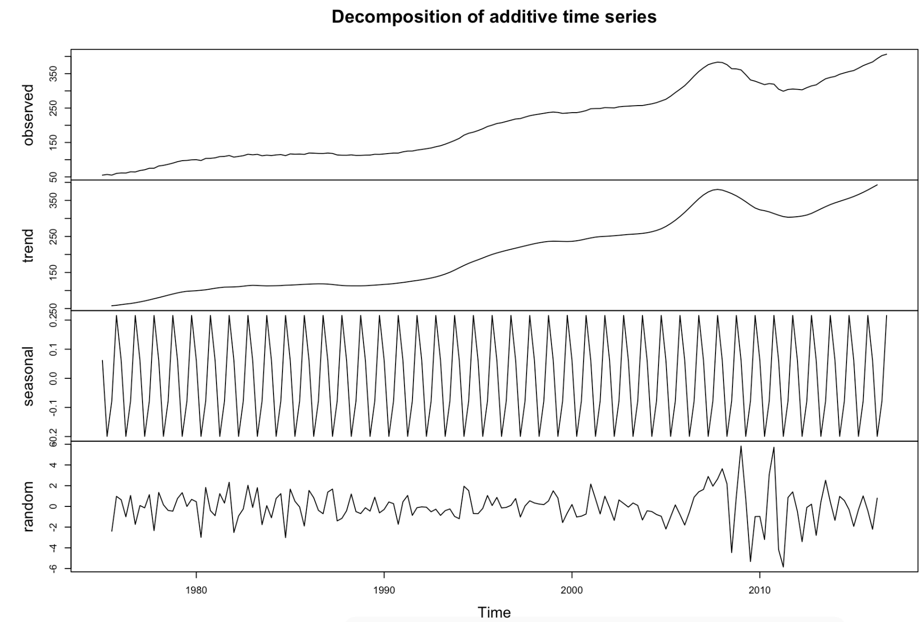


FIGURE 3

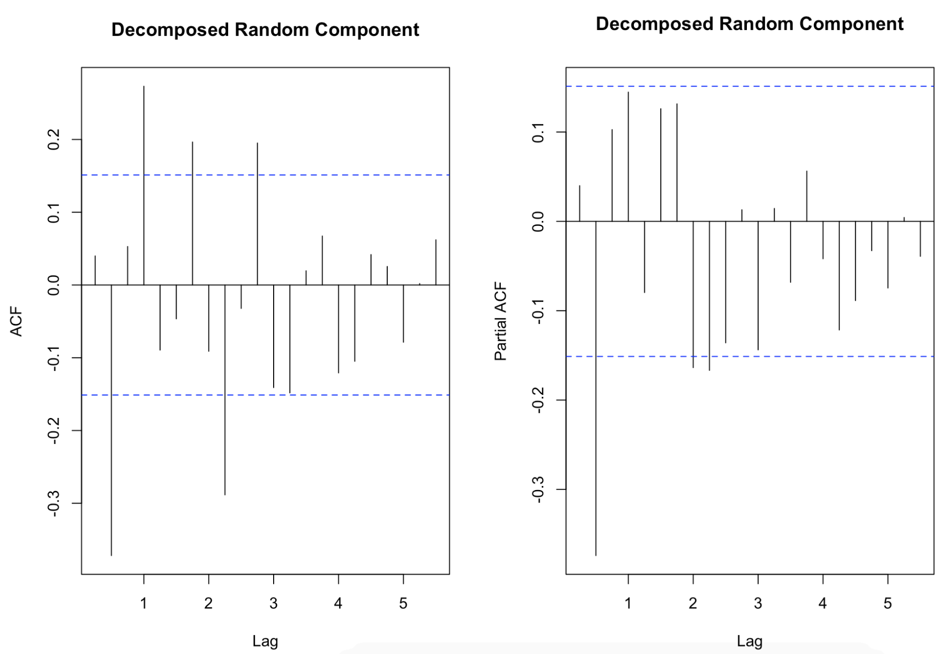


FIGURE 4

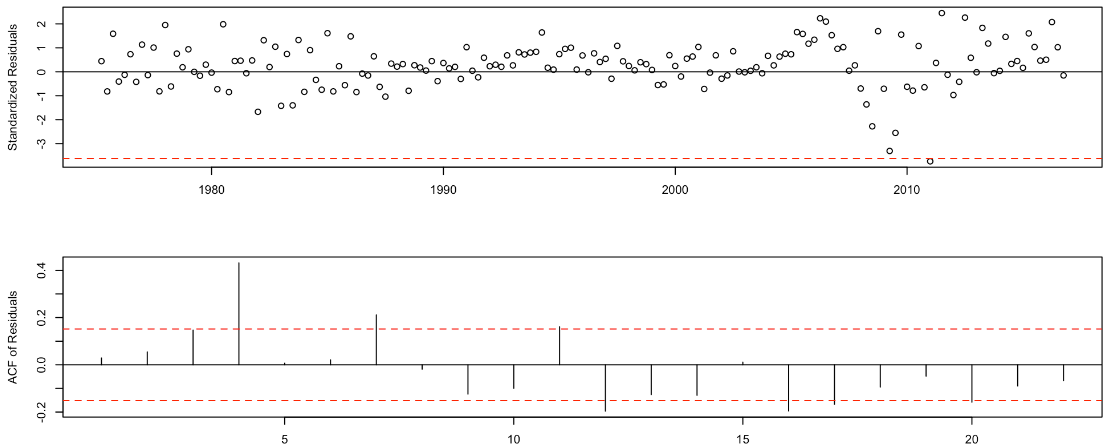


FIGURE 5

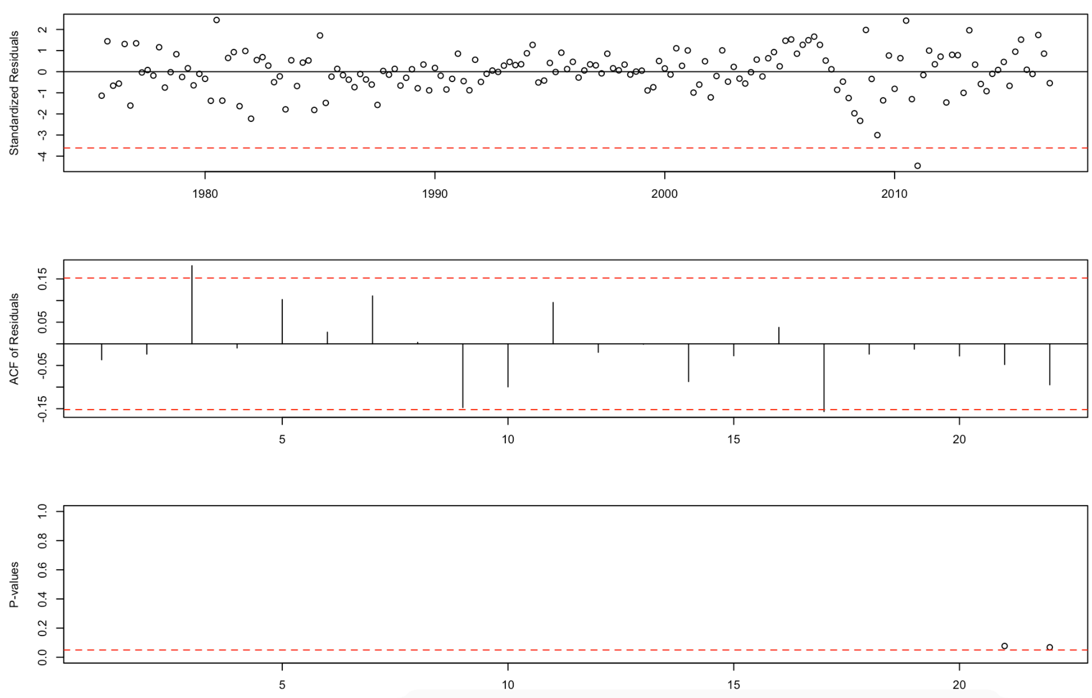


FIGURE 6

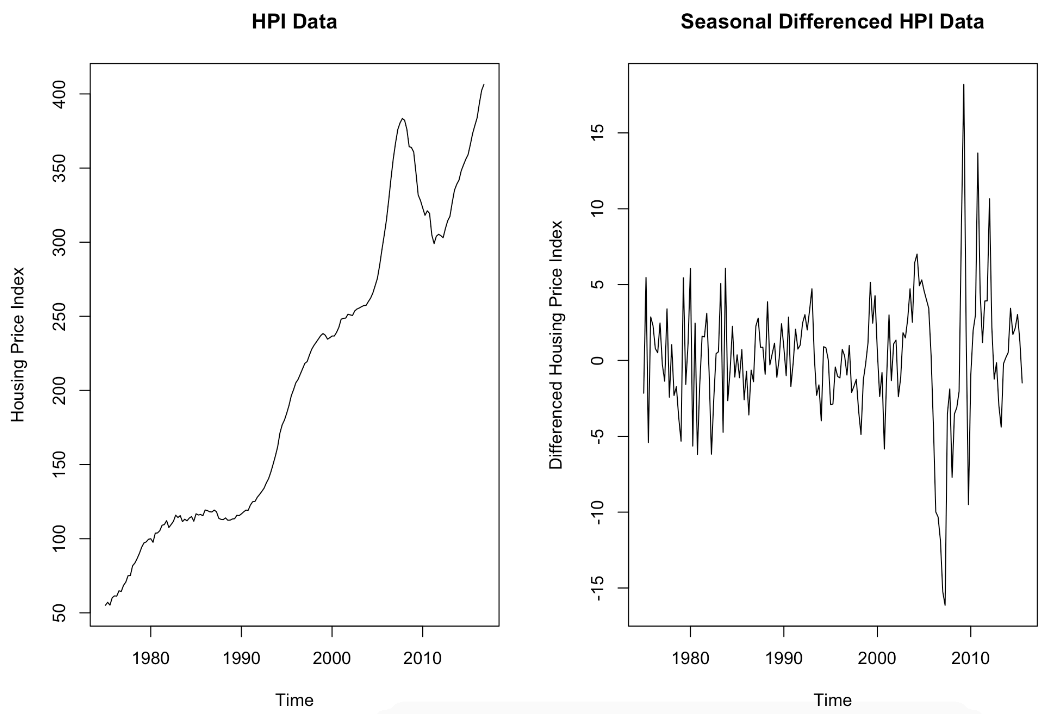


FIGURE 7

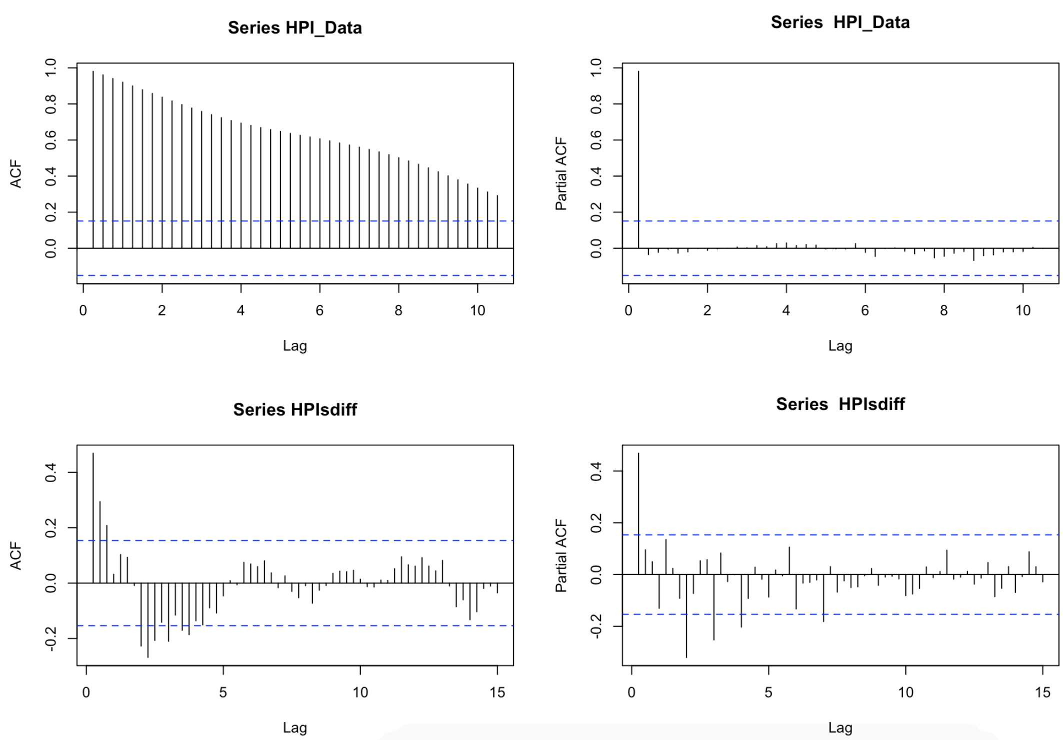
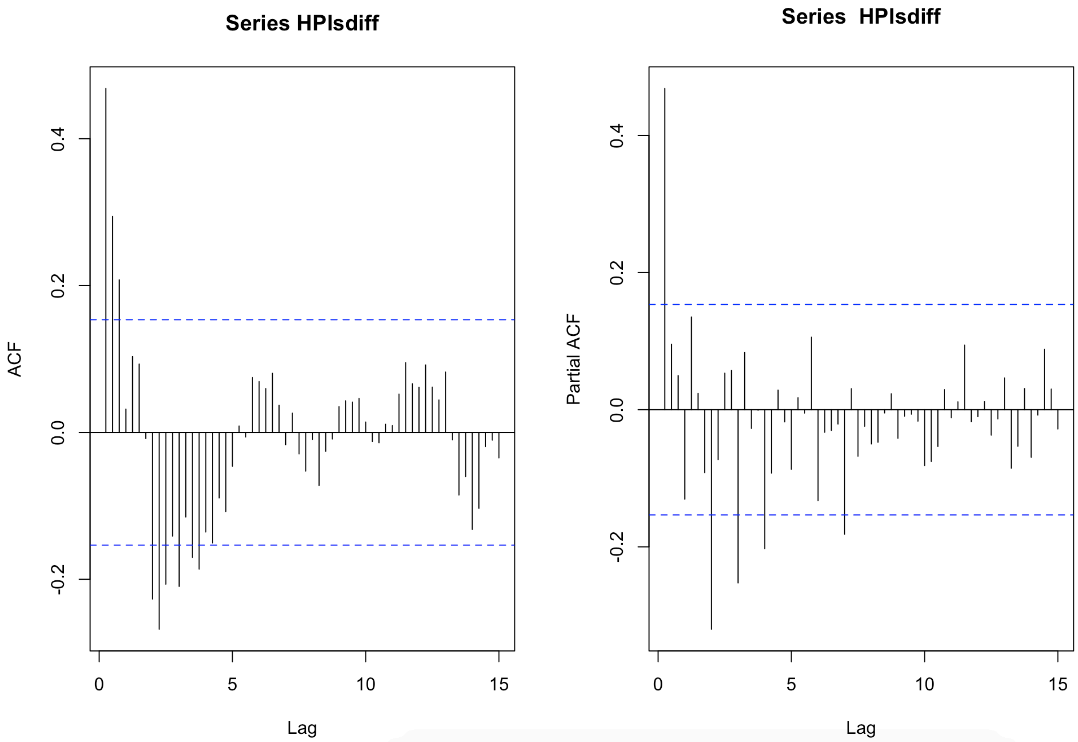


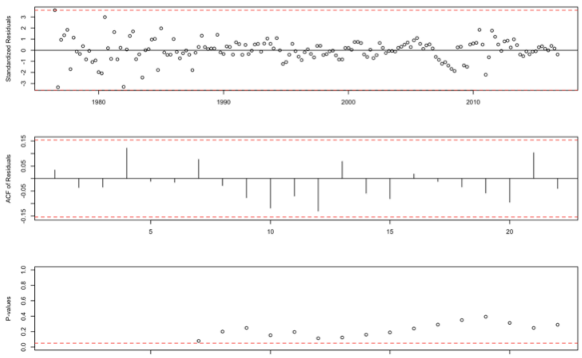
FIGURE 8



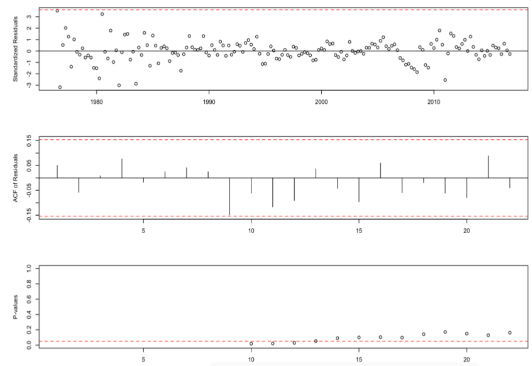
Series HPIsdiff
ACF
0.4
0.2
0.0
-0.2
0
5
10
15
Lag
Series HPIsdiff
Partial ACF
0.4
0.2
0.0
-0.2
0
5
10
15
Lag

FIGURE 9

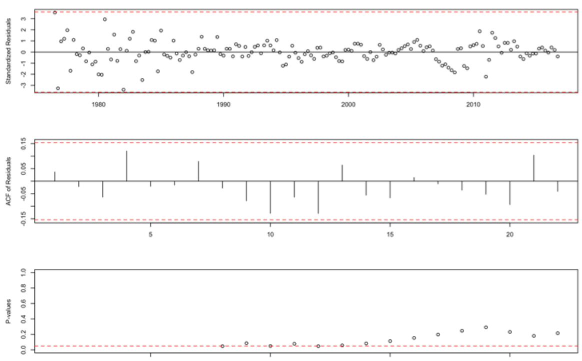
Model 1



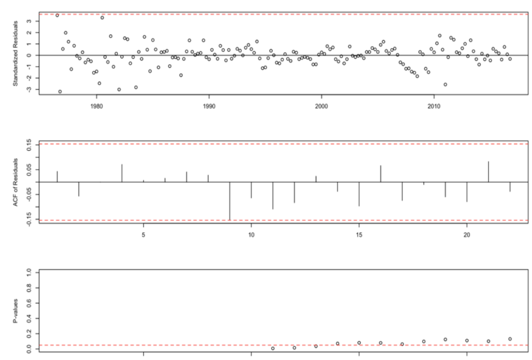
Model 3



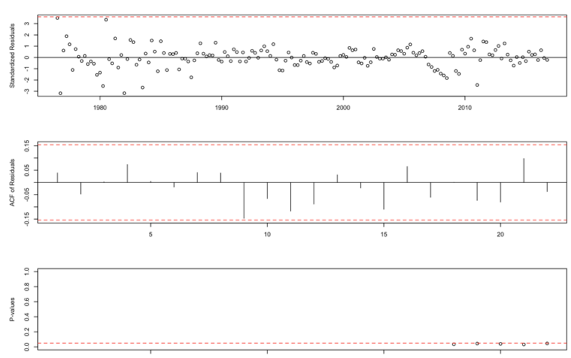
Model 2



Model 4



Model 5



Model 6

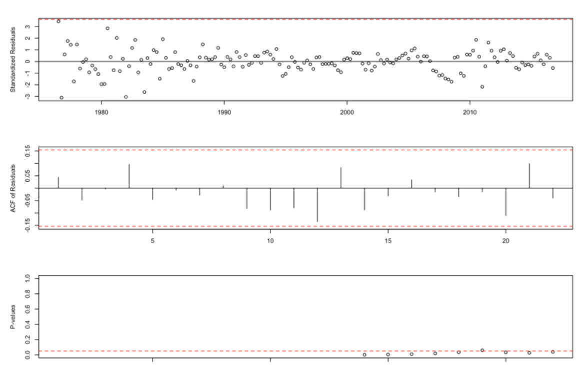
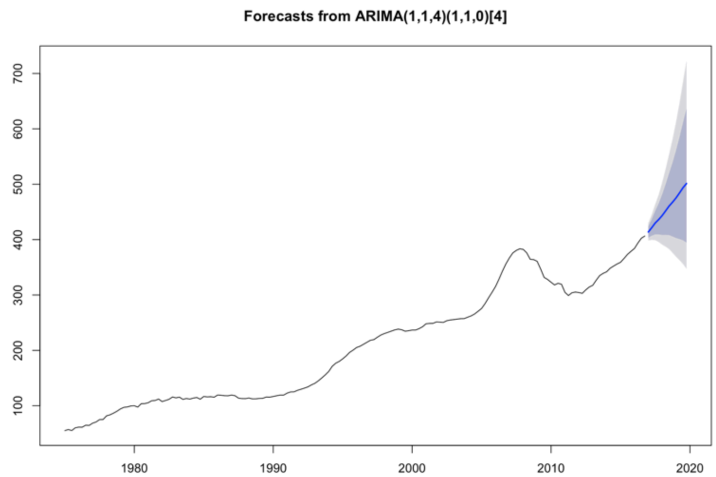


FIGURE 10



	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
2017 Q1	413.6806	403.6143	423.9979	398.3850	429.5633
2017 Q2	421.6905	407.0847	436.8203	399.5588	445.0481
2017 Q3	430.0710	409.4914	451.6850	398.9988	463.5630
2017 Q4	436.2916	409.3230	465.0371	395.7282	481.0130
2018 Q1	443.5088	408.4381	481.5909	391.0098	503.0568
2018 Q2	451.9776	408.5453	500.0272	387.2695	527.4976
2018 Q3	460.7329	408.2122	520.0108	382.8782	554.4185
2018 Q4	467.7224	405.6513	539.2915	376.2003	581.5101
2019 Q1	475.3687	402.8991	560.8734	369.1217	612.1976
2019 Q2	484.4320	401.0050	585.2155	362.8254	646.7969
2019 Q3	493.7806	398.8182	611.3544	356.1815	684.5366
2019 Q4	501.3207	394.6028	636.8996	347.6399	722.9390

FIGURE 11

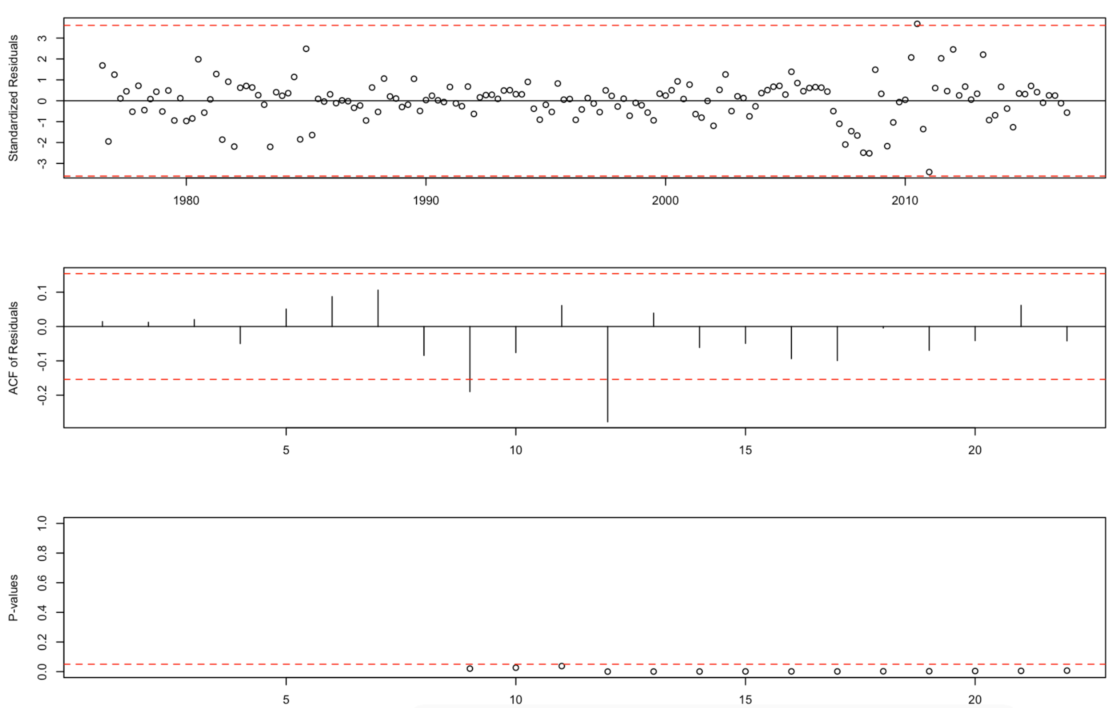


FIGURE 12

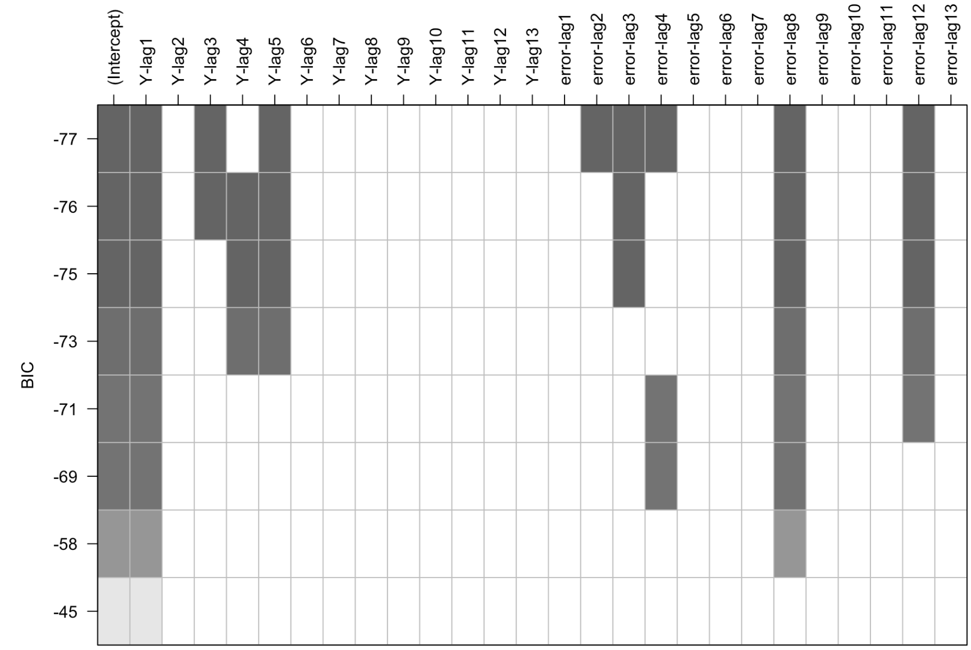


FIGURE 13

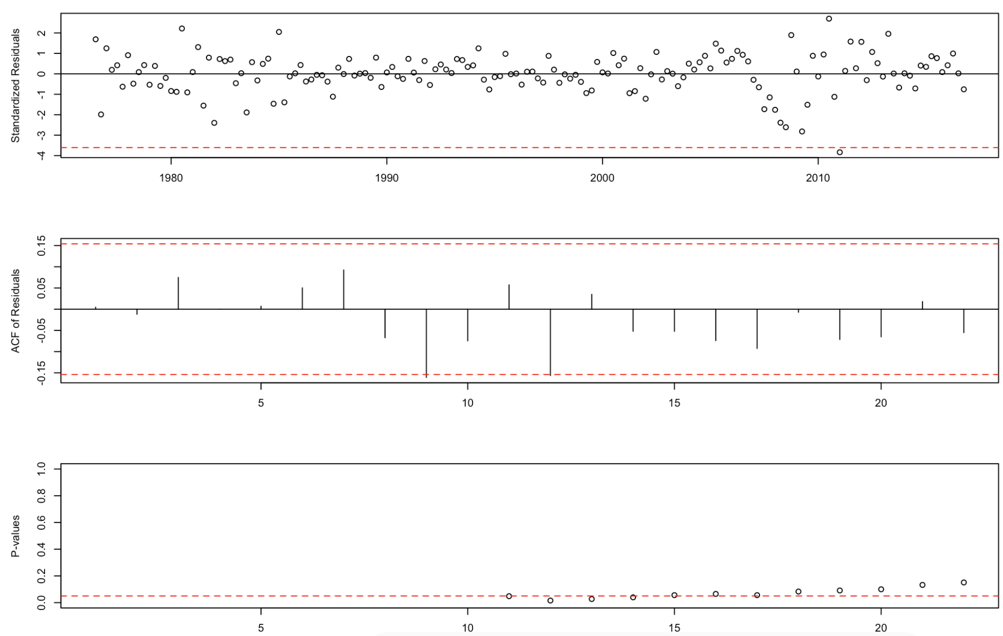


FIGURE 14

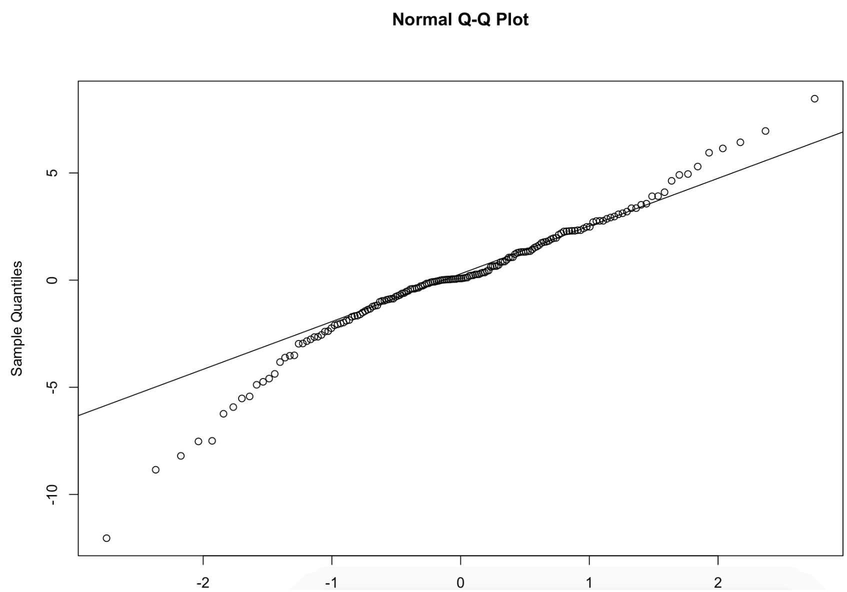


FIGURE 15

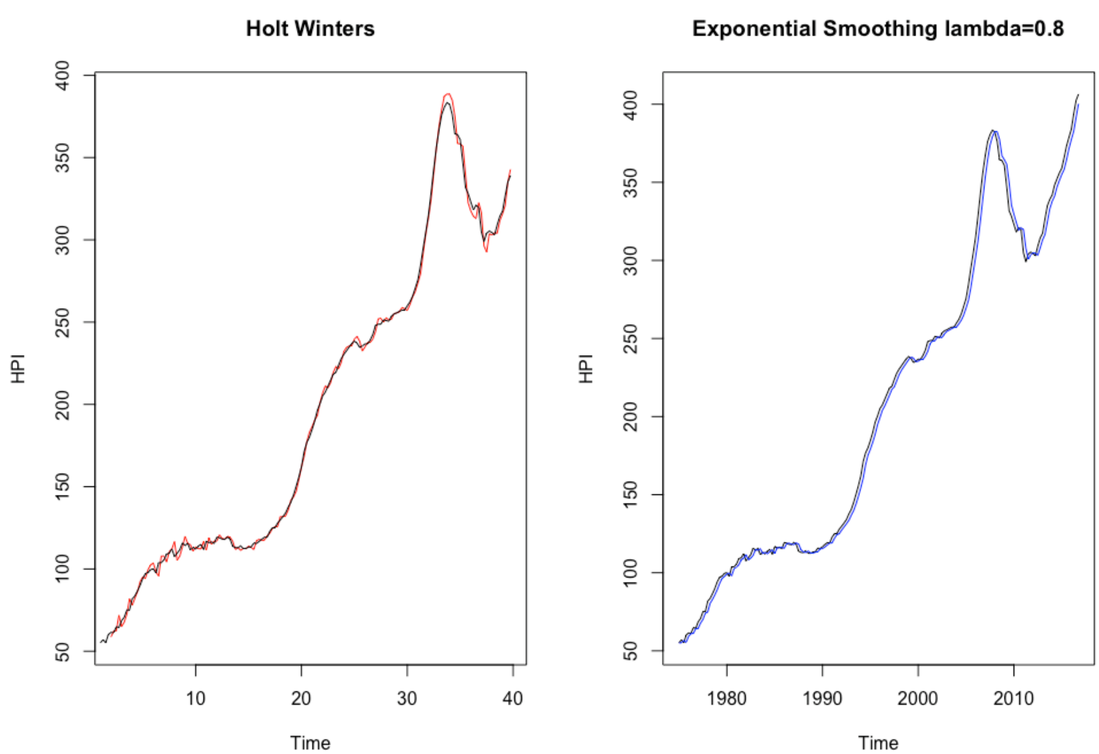
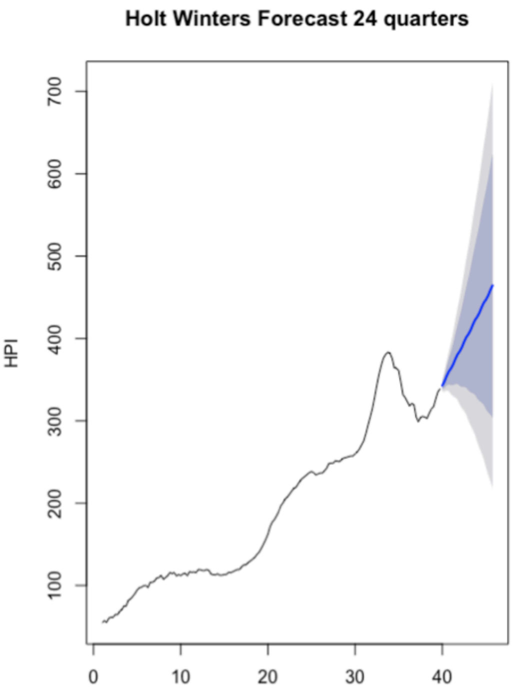


FIGURE 16



	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
40 Q1	342.8001	338.2173	347.3829	335.7913	349.8089
40 Q2	347.8089	340.0269	355.5908	335.9074	359.7104
40 Q3	354.3880	342.8949	365.8811	336.8108	371.9652
40 Q4	359.9433	344.2907	375.5959	336.0047	383.8819
41 Q1	363.6934	343.1818	384.2049	332.3236	395.0631
41 Q2	368.7022	343.3131	394.0912	329.8730	407.5313
41 Q3	375.2813	344.6534	405.9092	328.4400	422.1226
41 Q4	380.8366	344.6354	417.0378	325.4717	436.2015
42 Q1	384.5867	342.2616	426.9117	319.8561	449.3172
42 Q2	389.5955	341.1181	438.0728	315.4557	463.7352
42 Q3	396.1746	341.2574	451.0918	312.1859	480.1633
42 Q4	401.7299	340.0991	463.3606	307.4738	495.9860
43 Q1	405.4799	336.6697	474.2902	300.2438	510.7161
43 Q2	410.4888	334.4709	486.5066	294.2295	526.7480
43 Q3	417.0679	333.5971	500.5387	289.4103	544.7255
43 Q4	422.6232	331.4625	513.7839	283.2050	562.0414
44 Q1	426.3732	327.1119	525.6346	274.5661	578.1804
44 Q2	431.3820	323.9933	538.7708	267.1451	595.6190
44 Q3	437.9612	322.2275	553.6949	260.9618	614.9606
44 Q4	443.5165	319.2262	567.8067	253.4310	633.6020
45 Q1	447.2665	314.0482	580.4849	243.5267	651.0064
45 Q2	452.2753	310.1038	594.4469	234.8427	669.7080
45 Q3	458.8545	307.5322	610.1767	227.4271	690.2819
45 Q4	464.4098	303.7439	625.0757	218.6925	710.1270

FIGURE 17

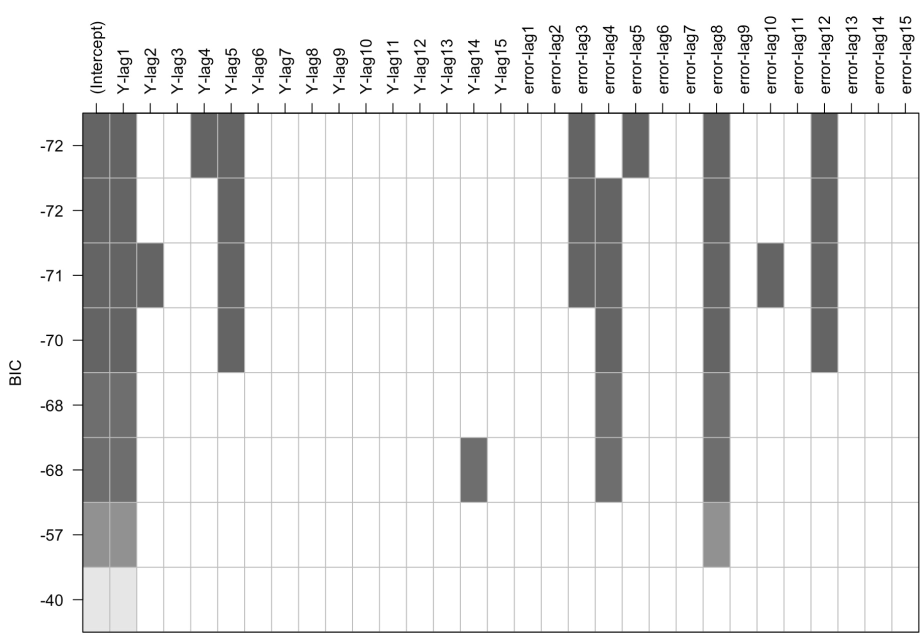
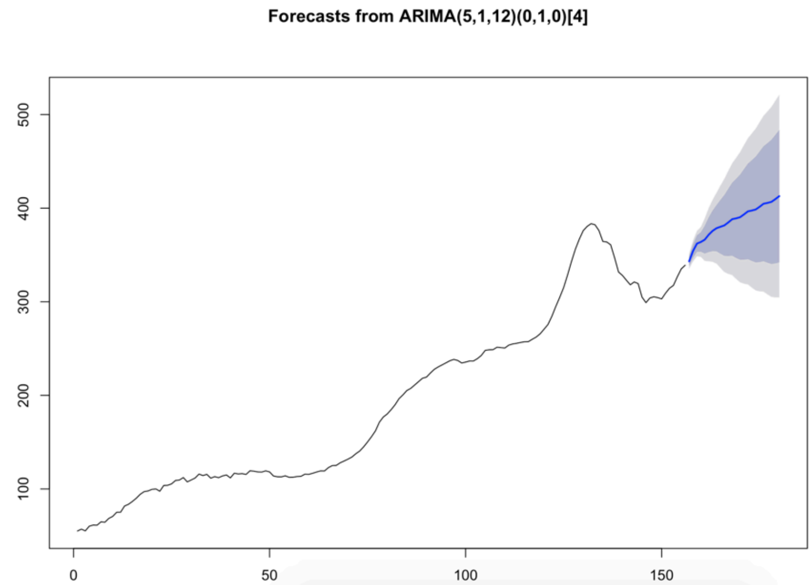


FIGURE 18



	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
157	343.2116	338.1330	348.2902	335.4446	350.9787
158	354.2261	347.0333	361.4190	343.2257	365.2266
159	362.1467	353.3247	370.9686	348.6547	375.6386
160	363.8744	353.4651	374.2836	347.9548	379.7939
161	366.3890	351.5182	381.2598	343.6461	389.1319
162	371.6488	353.0996	390.1980	343.2803	400.0173
163	375.7688	354.1036	397.4340	342.6347	408.9029
164	378.6046	353.8569	403.3522	340.7563	416.4528
165	380.1040	351.3410	408.8671	336.1148	424.0933
166	381.5957	349.0866	414.1048	331.8774	431.3141
167	384.7996	348.8412	420.7580	329.8060	439.7933
168	388.2359	349.1925	427.2792	328.5242	447.9475
169	389.1149	346.8099	431.4200	324.4149	453.8149
170	390.3453	344.9018	435.7889	320.8454	459.8453
171	393.3393	344.9455	441.7330	319.3274	467.3511
172	396.5789	345.5515	447.6062	318.5393	474.6185
173	397.4600	343.6297	451.2903	315.1336	479.7863
174	398.5971	342.0623	455.1319	312.1346	485.0596
175	401.5312	342.4244	460.6380	311.1352	491.9273
176	404.7532	343.3190	466.1873	310.7978	498.7086
177	405.5997	341.6796	469.5197	307.8424	503.3569
178	406.7256	340.3839	473.0674	305.2647	508.1865
179	409.6484	340.9842	478.3126	304.6355	514.6612
180	412.8586	342.0753	483.6419	304.6049	521.1123

FIGURE 19

2014	1	341.93
2014	2	348.24
2014	3	352.21
2014	4	355.97
2015	1	359.04
2015	2	365.87
2015	3	373.28
2015	4	378.75
2016	1	383.95
2016	2	393.81
2016	3	402.47
2016	4	406.46

FIGURE 20