Time Series Analysis Project

Irene Cho
2020.10.14

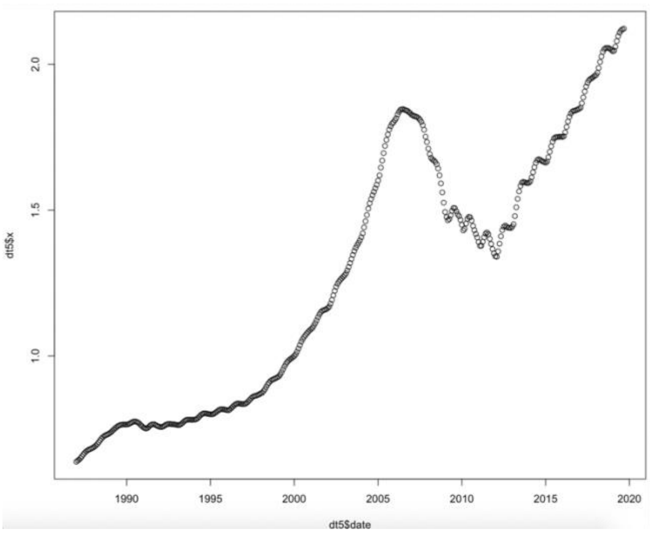
Executive Report

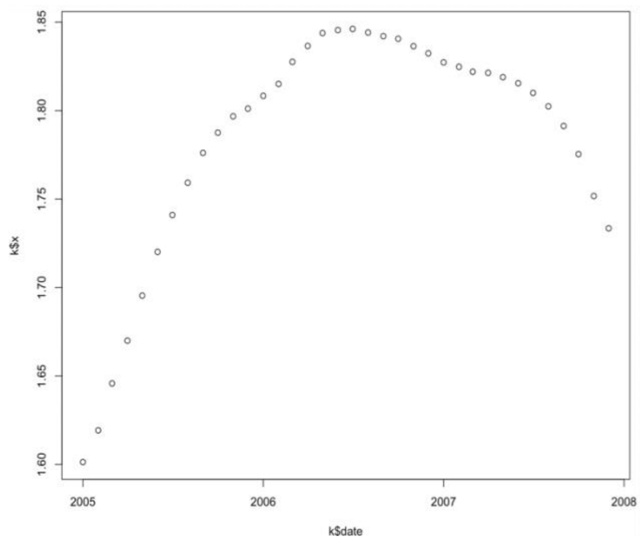
The goal of the following time series analysis is to explore the bubble effect which occurred on 2000 and forecast the future trend of house prices from the given the S&P/Case-Shiller Home Price Index. The dataset includes 393 proportions of housing prices compared to the price in January 200 from January,1,1986 to September 1, 2019. It has 0.6375 for a minimum proportion and 2.112 for a maximum value. If the housing price in January 2000 was $500,000, the maximum proportion of 2.112 in September 2019 implies that the price was increased 212.203 percent which were $500,000*2.122=$1,601,015. There was a housing bubble in 2000 that the housing prices increased dramatically and dropped extremely at the second phase. Therefore, the data were not stationary, and they required differentiating. The first differentiated data was not still stable, so I differentiated the data 2 times. Afterwards, I compared the ACF and PACF values. The ACF values showed sinusoidal decay but they decreased slowly, so the data might include moving average (MA) components. The sample PACF values were cutoff at lag 12 which implied a AR (12) model. I used a periodogram plot to decide a final model that there were 2 peaks and 1 dip. One of the peaks was existed at the 0 frequency, thus, the possible models were ARIMA (3,1,1) and ARIMA (3,1,2). ARIMA (3,1,1) had relatively small AIC values, -3392.98 and more ACF of residuals were in borderlines than AIRMA (3,1, 2). Therefore, the appropriate model for this dataset would be ARIMA ((3,1,1). The estimated coefficients of the model looked like $X'' = 0.7459\, X_{t\text{-}1} + 0.5275\, X_{t\text{-}2} + -0.5018\, X_{t\text{-}3} + Z_0 + 0.7069 Z_{t\text{-}1}$ The standard errors of the coefficients of the model were all small enough to prove that the variables had significant effects on Xt. The residuals did not have good normal assumptions by QQplot, but they had constant variance by a residual plot.

I forecasted future housing values depending on bubble effects. I trained 90 percent of original data and predicted the remaining 10 percent and 9 more future values. If there were no bubble effects, most of the proportion of housing prices were almost zeros that the price would not increase or decrease in the future. However, the prices were kept increasing when there were bubble effects. I trained all observations and predicted 9 future values which were 2.124, 2.128, 2.13, 2.14, 2.15, 2.16, 2.170, 2.178 and 2.185. I presented them on the original scatter plot, and they were well connected with previous lag values. However, the predicted values in 2 differential model is not stable. Therefore, we have to consider environmental problems for the accurate values.

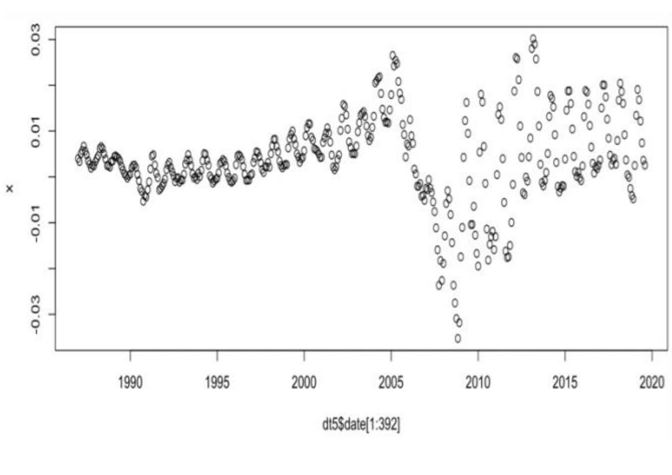
Technical Report

The dataset includes the proportion of housing price comparing to the price on 2000 when the housing bubble occurred. The data had 393 observations in which the minimum value was 0.6375 and the maximum value was 2.112. I checked the distribution of the data through two different scatter plots

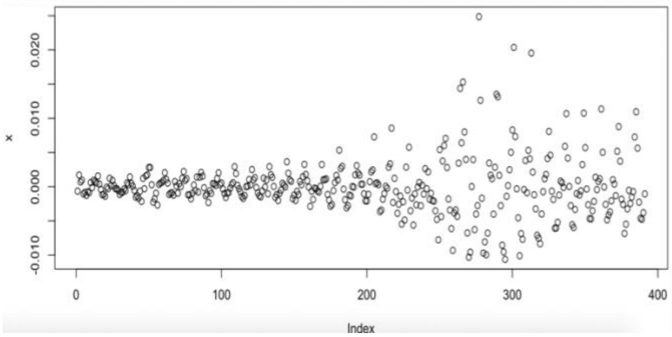




The scatter plot of this data showed that housing bubble would occur in 2000 when the housing prices abruptly increased. The burst point of bubble where the prices decrease dramatically was July 2006 that the proportion of the price was 1.8365. The housing prices decreased until March 2011 which was 134 percent and then they started increasing monotonically. This increasing phenomenon could be occurred by the effect of the bubble. I differentiated the data to see the trend of housing prices without bubble effect.

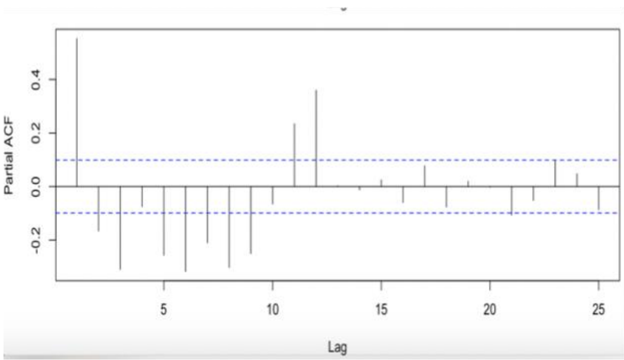


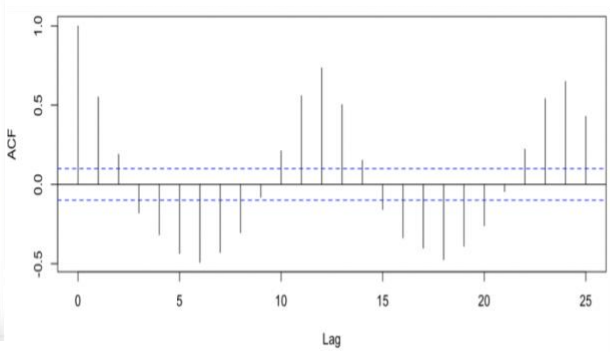
This is the distribution of first differentiated data. Even though I transformed the data, it still had a bubble housing effect.



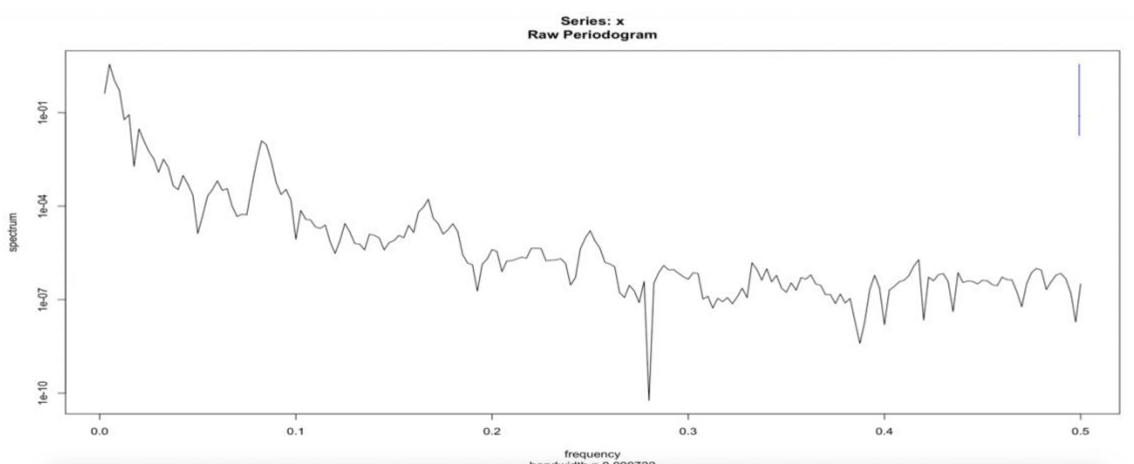
I differentiated the data again and it had stable distribution. Using this updated data, I estimated an appropriated model.

I compared the ACF and PACF values.





The ACF values showed sinusoidal decay but they decreased slowly, so the data might include moving average (MA) components. The sample PACF values were cutoff at lag 12 which implied a AR(12) model. I used a periodogram to find an accurate model.

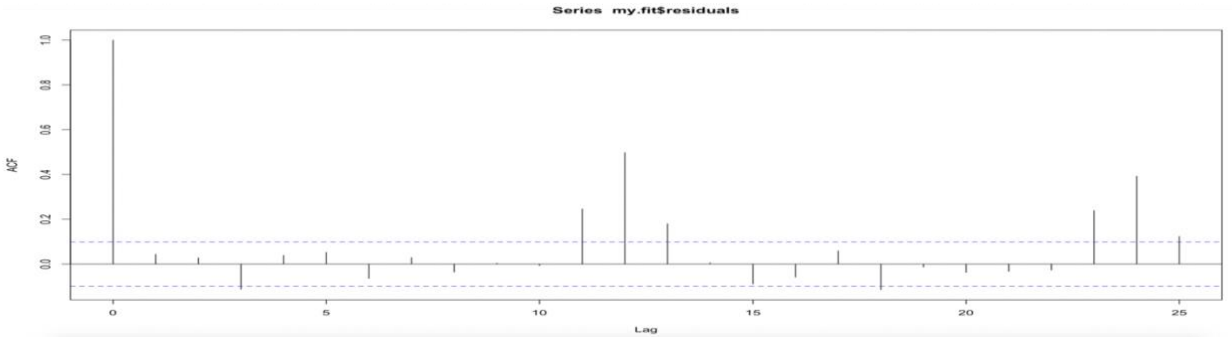


In this plot, there were 2 peaks and 1 dip. One of the peaks existed at the 0 frequency. Therefore, the possible models were ARIMA (3,2,1) and ARIMA(3,2,2).

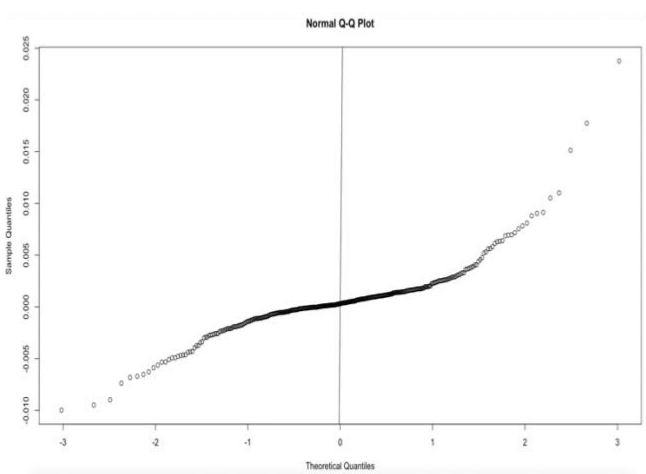
ARIMA (3,2,1) had relatively small AIC values, -3382.38 and more ACF of residuals were in borderlines than AIRMA (3,2, 2). Therefore, the appropriate model for this dataset would be ARIMA (3,2,1). The estimated coefficients of the model looked like

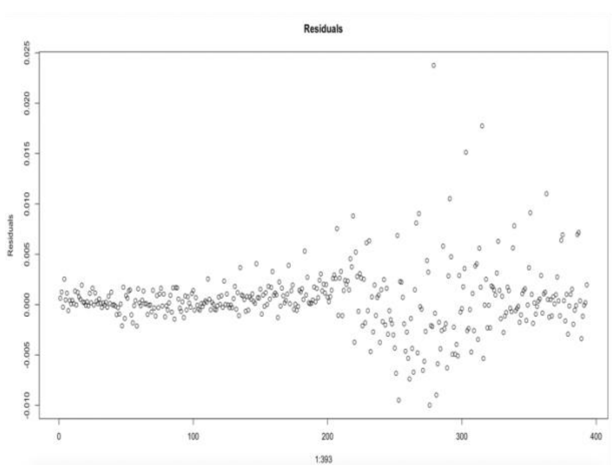
$X_t = 0.7459\ X_{t-1} + 0.5275\ X_{t-2} - 0.5018\ X_{t-3} + Z_t + 0.7069 Z_{t-1}$.

The standard errors of the coefficients of the model were all small enough to prove that the variables had significant effects on Xt. I validated the model using various residual plots.



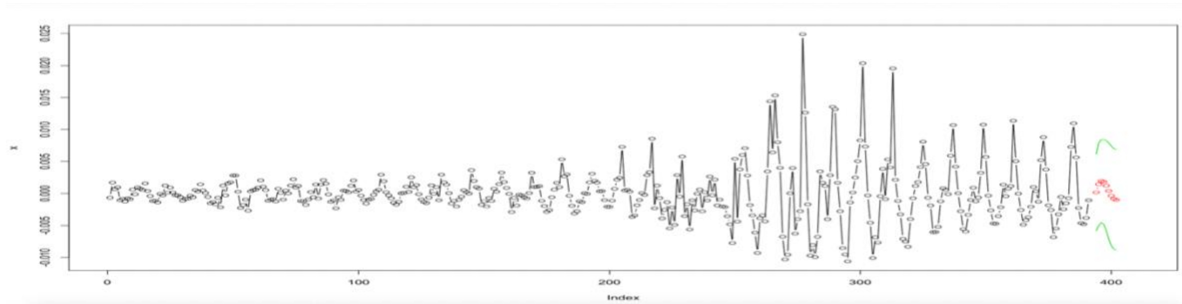
This is the plot of sample ACF of residuals and most of the values were within two borderlines. I assessed the normality of the residuals from the estimated model using QQplot and checked the pattern of residuals using a residual plot.

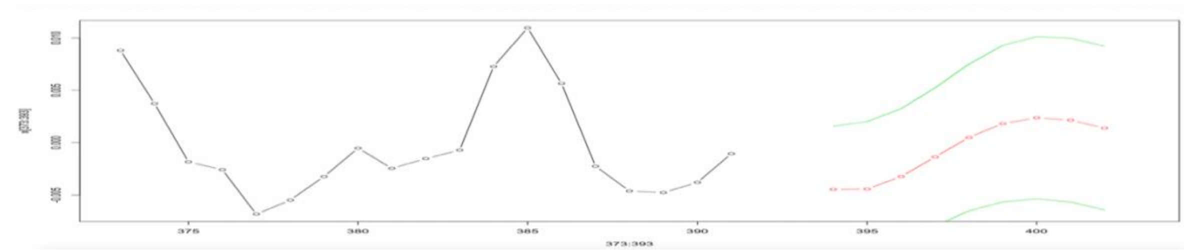




The residuals of the model were not really on the line, so the model might not represent all patterns of the data. The residuals in the residual plot did not have a specific shape. Thus, it seems that the residuals had constant variance.
I first forecasted future housing prices if there was no bubble effect on 2000 using estimated ARIMA (3,2,1) model. When there was no bubble effect, the data repeated the patterns of getting bigger and smaller over time, but they did not have a monotonically decreasing or increasing

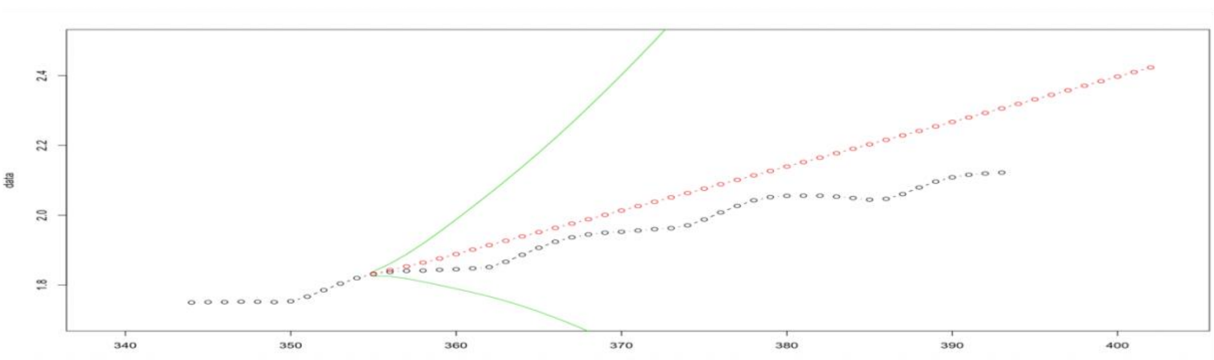
trend, which means that they were stationary. The distance between peaks seems pretty regular, so periodicity existed.





The first plot has predicted values with an entire dataset, and the second plot has the same predicted values with the partial data from January 2018 to September 2019.

The predicted values are 0.000166, 0.00143, 0.0019, 0.0018, 0.0012, 0.00039, -0.000359, -0.0008502914 and -0.000997. Most of the proportion of housing prices were almost zero which means that the price would not increase or decrease in the future.
When I considered the bubble effect, the predicted values were different. I forecasted future data points using an estimated ARIMA (3,2,1) model. I first trained the model with 90 percent of the observations, which were 291 values, and predicted the remaining 10 percent of observations, which were 32 values. In addition, I predicted the next 9 values from the original data and represented them on the following plot.



The housing prices kept increasing because of bubble housing effect. I used the whole data set and forecasted the next nine values again.

When I used all 584 observations, the forecasted 9 values were more smoothly connected to previous values. They were 2.124, 2.128, 2.13, 2.14, 2.15, 2.16, 2.170, 2.178 and 2.185.
I presented them on the original scatter plot, and they were well connected with previous lag values. However, the predicted values in 2 differential model are not stable. We have to consider environmental problems for the accurate values.

s