Time series forecast project

Presented to

Dr. Crunk

Department of Mathematics and Statistics

San José State University

In Partial Fulfillment

Of the Requirements for the Class

MATH 265

By

Jung-a Kim

December, 2019

< Dataset 1 – Executive summary >

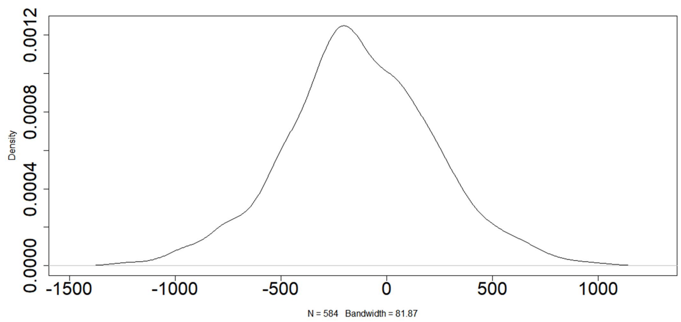

Figure 1. Density of dataset 1

The first dataset appears approximately normally distributed and centered around -126 (figure 1). The final model chosen after comparison with three methods was auto-regressive model with order of 5 which needs at least 5 past values to predict one future value. The order of the model was chosen based on the assumption that data follows normal distribution since the data appears approximately normal. The final model was created using Burg's method since it had the smallest prediction error for the 10% of the data that was held out before training. In the model, the future values are weighted sum of the past five values and the predictions heavily depend on the immediate past two values as the weights are 2.7, -2.6, 0.2, 1.1, and -0.6.

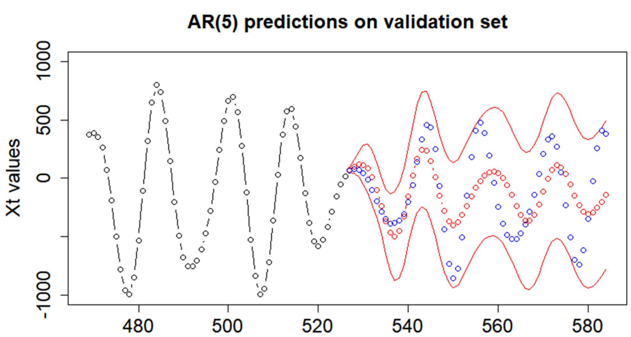

Figure 2. The last 58 values of training data and 58 predictions(red points) on validation set and true values(blue points) using Burg method

Before the analysis of the data, it was split into portions of 90:10 and 90% of the data was used as training data and 10% of the data as validation set to measure the prediction error. In figure 2, the model fitted to the training data seems to predict the nearest 17 future values fairly well, but after then the predictions are way off from the true values. However, the prediction intervals for the next 58 values captures all the true values and they are as narrow as they can be.

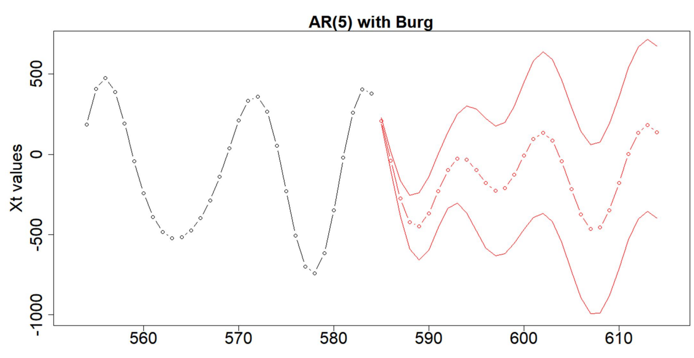

Figure 3. The last 30 values of data and 30 predictions

The final model fitted to the whole data set is shown in figure 3 and 30 future values were forecasted . The first three prediction intervals have widths less than the standard error of the data which may imply the nearest three predicted values may be reasonable to use like other past values. After the fourth prediction, the interval becomes too wide and much greater than the standard error, they may not be useful.

The fitted model is

$$\mathrm{X_t} = -13.9666 + 2.7323\mathrm{X_{t-1}} - 2.6074X_{t-2} + 0.2186X_{t-3} + 1.1387X_{t-4} - 0.5925X_{t-5} + Z_t$$

< Dataset 1 - Appendix >

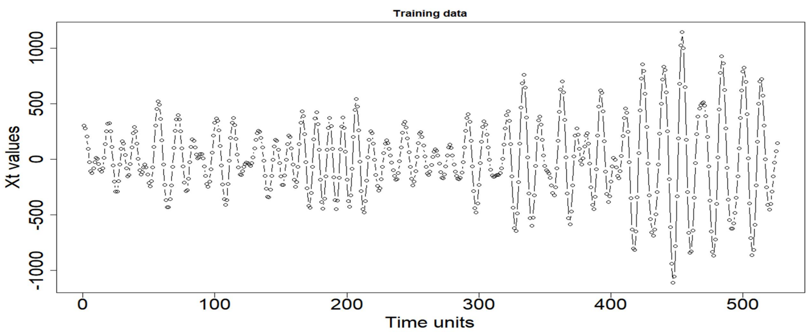

Figure 4. Time series plot of training data

From figure 4, the data seems stationary and it seems that there are two periods. One guess is that there are 48 short peaks and 35 tall peaks which lead to periods of 12 and 16.

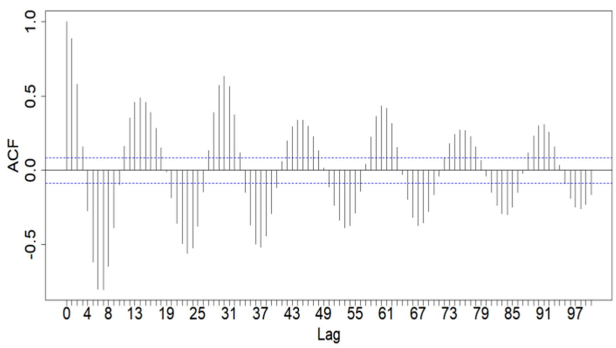

Figure 5. ACF plot of training data

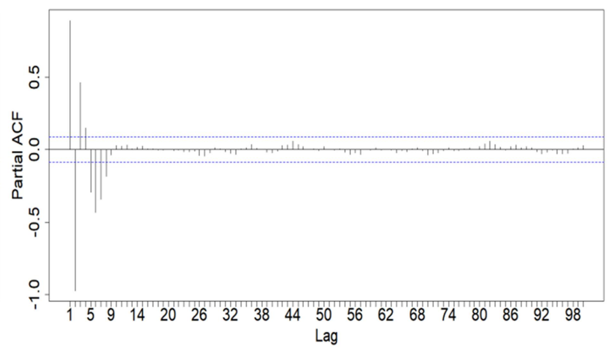

Figure 6. P.ACF plot of training data

There is a clear sinusoidal decay in sample autocorrelation function starting from lag zero through lag 100 which means MA component is not likely to be in the model. The sample partial autocorerlations from lag zero to lag 7 are outside the 95% confidence interval and lag 8 through lag 100 is within the interval. Since there can be 5 out of 100 sample p.acf's outside the confidence interval, the true p.acf(4) through p.acf(8) may be zero or less than these 5 true p.acf's may be zero. Thus, there can be 6 candidate models from AR(3) through AR(8) based on the acf and p.acf plots.

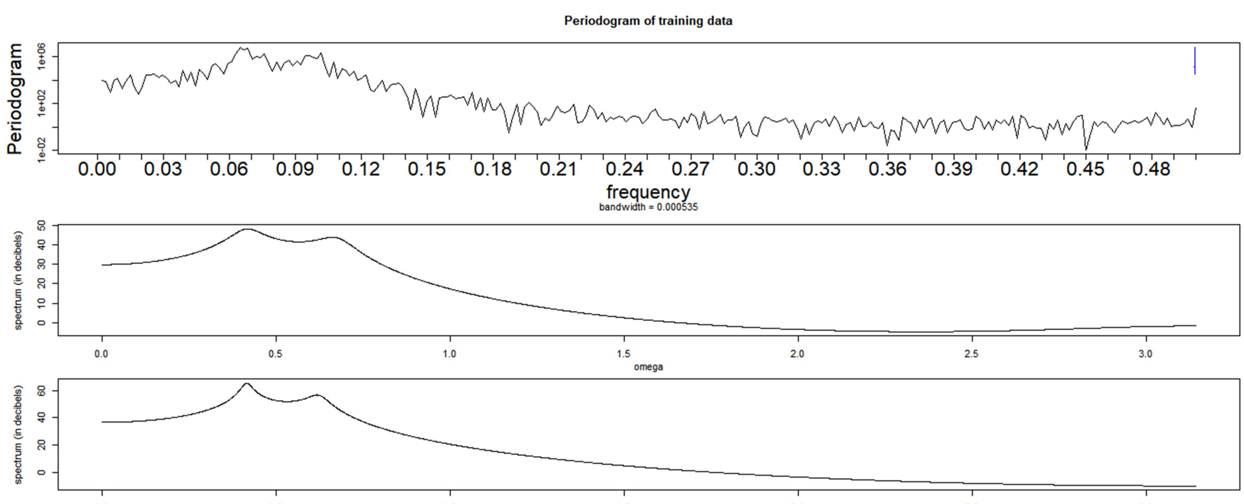

Figure 7. Periodogram(top), spectrum of AR(5), and spectrum of ARMA(4,1)

At one glance, the periodogram in figure 7 shows two peaks of periodograms at the frequencies which mainly contribute to the variability of the data. The estimated frequencies from scale 0 to 0.5 are 0.065 and 0.1 which converts to the periods of 15 and 10 which were somewhat close to the periods of 16 and 12 that had been guessed from the time series plot. The remaining frequencies form 0.20 to 0.5 seem like noise without dips or peaks. But it is worthwhile to try some models to make more comparisons. The candidate models are AR(4), AR(5), AR(6), ARMA(4,1), ARMA(4,2), ARMA(5,1), and ARMA(5,2).

AR(4) did not satisfy the assumption that residuals are white noise. The correlation of residuals in AR(4) at lag 1 and lag 2 were far away from the 95% confidence interval. The residuals in AR(5) and AR(6) showed strong sign of white noise distribution in scatter plots and correlogram.

ARMA processes had insignificant MA coefficients and the spectrum seemed to be overfitting the periodogram as observed in the spectrum plot of ARMA(4,1) in figure 7.

For the parameter estimation, three methods were used: MLE (initialized with OLS estimates), OLS, and Burg. Several criteria were chosen for model selection including diagnostics on the residual assumptions, prediction error for validation set, and t-test for the coefficients for each pair of the model and the method.

For MLE method, AR(5) was chosen for the lowest PSSE on the validation set. AIC showed no particular difference ranging from 3929 to 3935. The qq-plots for all four candidates showed strong sign of normality. For OLS method, AR(6) was chosen for the lowest PSSE and four out of six coefficients were considered to be significant while there were three out of five significant coefficients in AR(5) model created by MLE method. For Burg method, AR(5) had the lowest PSSE and all the coefficients were significant compared to the other models.

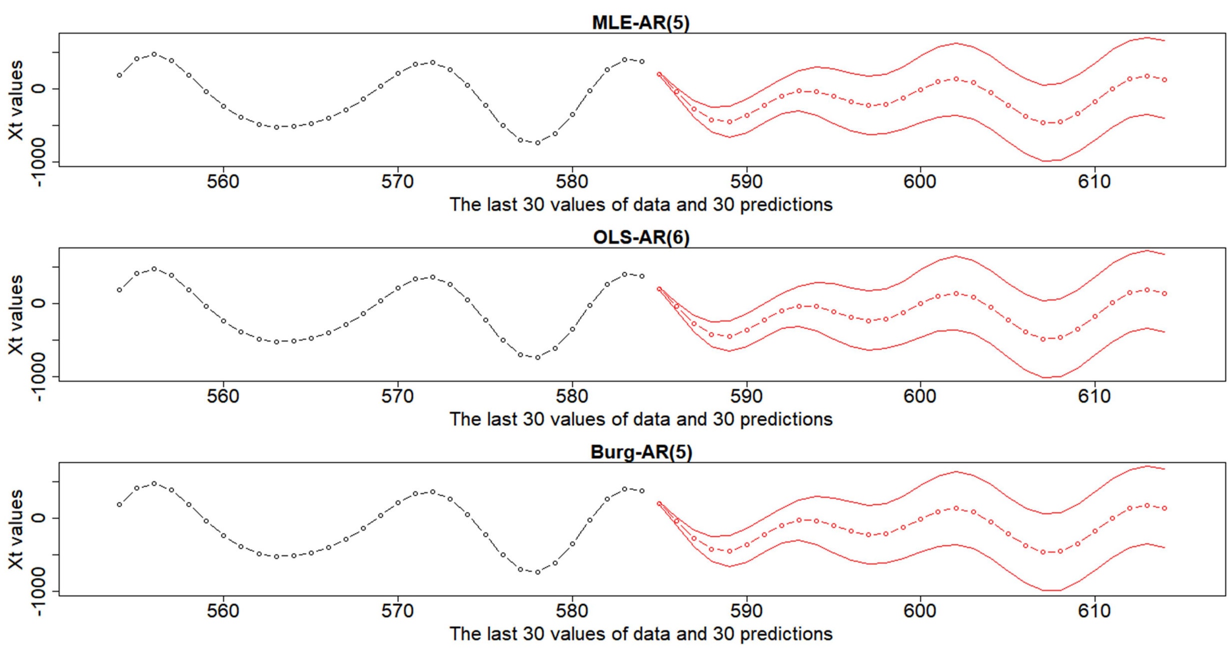

Figure 8. Prediction intervals for the next 30 values using MLE, OLS, and Burg

Figure 8 shows the prediction intervals for the next 30 values by fitting the whole data to the model using the three methods. The coefficients in all three models are the same up to second decimals. Thus, the prediction intervals and the point estimates are very similar. Particularly, the absolute mean difference between OLS-AR(6) and Burg-AR(5) predictions is only 1.0 which is very small considering that the standard deviation of the data is 351. As a final decision, Burg-AR(5) was chosen for these reasons: 1) Burg-AR(5) had the slightly lower prediction error compared to the other models. 2) Only Burg-AR(5) model had all the coefficients that are significant. 3) it has one less parameter to estimate than OLS-AR(6) which reduces the potential error of estimating one extra parameter.

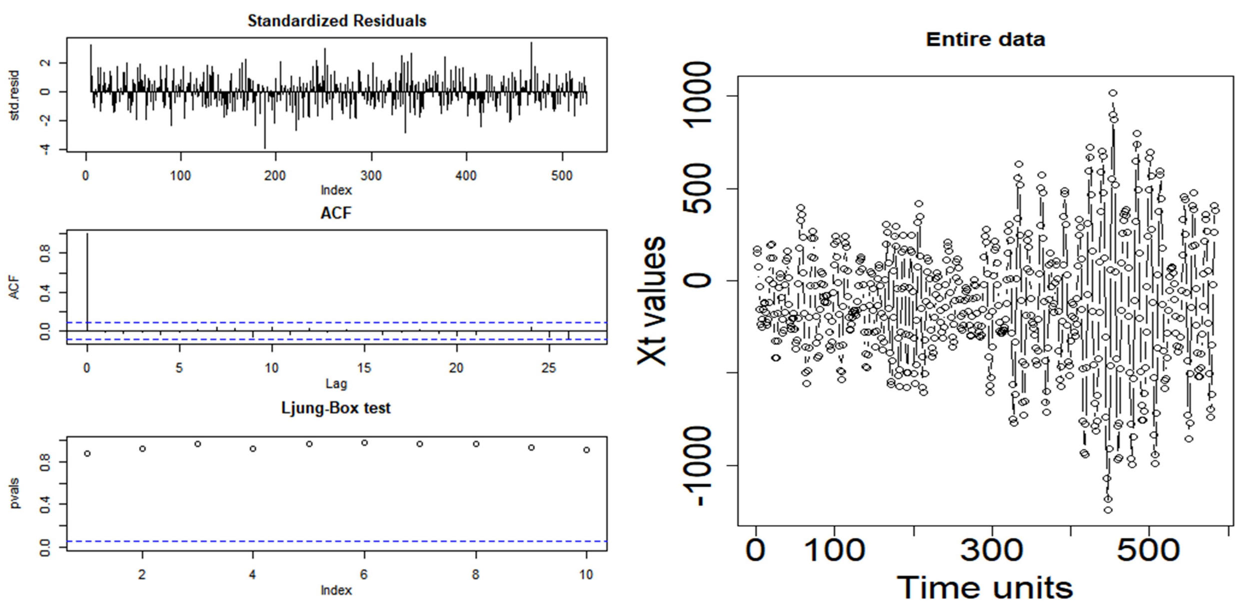

Figure 9. Residual plots and time series plot of the whole data

Adding back the last 58 observations does not change the residual assumptions or stationarity or periods of the data. The standardized residuals seem like white noise without particular pattern. Sample autocorrrelations are cut off after lag zero and the true acf's after lag zero may be zero as well since p-value of Ljung-Box test at lag 10 is very high as 0.8 which tests whether all acf's up to lag 10 are zero. The last 58 observations seem to be part of the stationary data since the time series plot does not seem to show irregular gaps between time units or non-constant mean that could indicate non-stationarity. Thus, Burg-AR(5) can be the optimal model for this type of dataset.

< Dataset 2 – Executive summary >

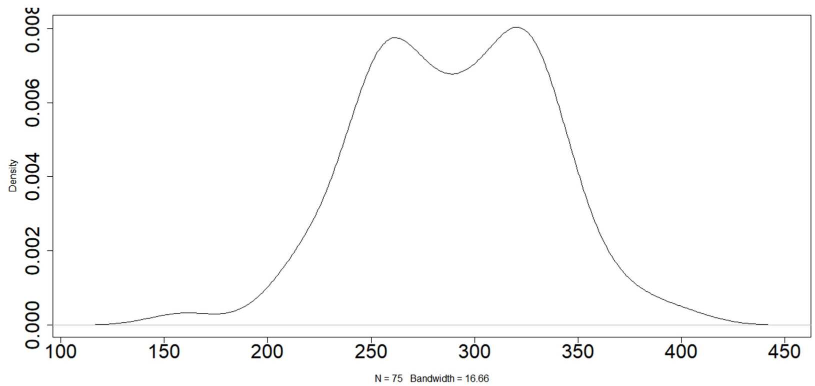

Figure 1. Density of dataset 2

There are 75 observations in this data. The data looks bimodal which strongly indicates that the data is not normal, so the method of maximum likelihood estimator was ruled out from model selection which assumes normality in the data.

The conclusion on this analysis was that the data is simply noise without any corrleation with one another, thus it is impossible to predict the future values based on this data. Several models were examined through model adequecy check, but most of the times there appeared strong sign of white noise.

< Dataset 2 – Appendix >

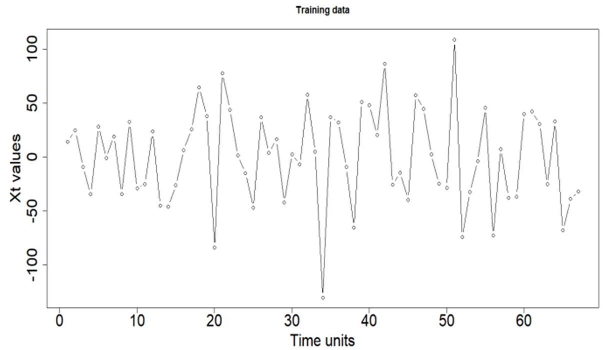

Figure 2. Time series plot of training data

Figure 3. Periodogram of training data

In figure 2, at first glance the data looked like it had periods between 2 and 4, but in certain time intervals the length of period was irregular. The data seems stationary in the mean and correlations. In figure 3, there seems to be too many local frequencies that contribute to the variability of data. Since the training data has only 67 points, having too many frequencies within this small dataset may just indicate that the data is noise. If there happens to exist any period, then the first frequency to consider would be 0.208, then the period would be 4.8, but it is hard to identify such period in the time series plot.

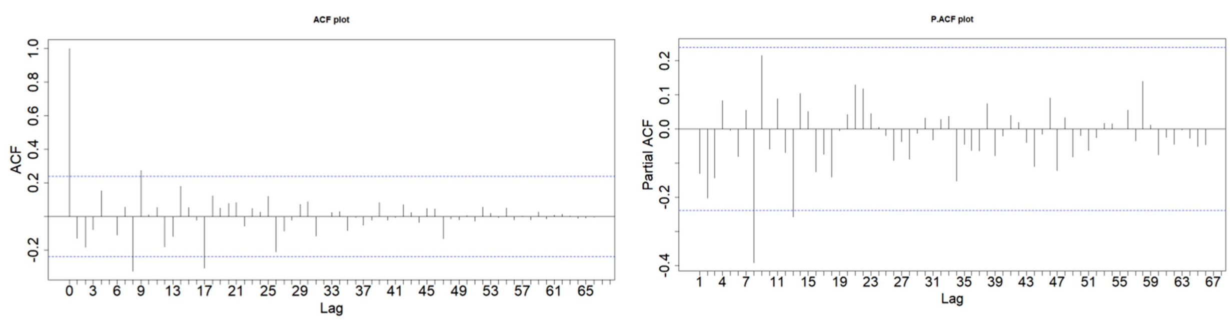

Figure 4. Sample autocorrelations(left) and Sample partial autocorrelations(right)

Another indication that the data is white noise can be shown by acf and p.acf plots. ACF could die off after lag zero, lag 8, or lag 17 which are out of the confidence interval, but the acf at lag 8, 9, and 17 are just a little bit outside the interval. There doesn't appear a sign of sinusoidal decay in acf plot.

P.ACF could die off after lag 8, but only two out of 67 points are outside the confidence interval, so they may be considered as sampling errors and true p.acf may be zero after lag zero. Also, it is unlikely to see all sample p.acf's at lag 1 through lag 7 inside the interval if the true p.acf at lag 8 was nonzero. One suspicion remaining is that the sample p.acf at lag 8 is quite far away from the interval. Thus, the second candidate model other than ARMA(0,0) was ARMA(8,0).

ARMA(8,0) was fitted with both OLS and Burg methods and it seemed to satisfy the residual assumptions. But all 8 coefficients in the model were insignificant based on t-tests. ARMA(1,0) was also fitted with OLS method in case there may be at least one order of AR in the model, but the coefficient was not significant.

Thus, the conclusion is that the data is white noise.

< Dataset 3 – Executive summary >

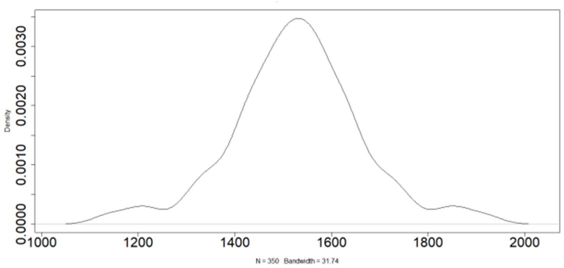

Figure 1. Density of dataset 3

There are 350 observations in the dataset 3. The chosen model is ARMA(5,1). The method to find this model was MLE which assumes the normality in the data. The data seems to follow normal distribution in figure 1 so it satisfies the assumption of the method. The last 35 observations were held out as validation set for model selection.

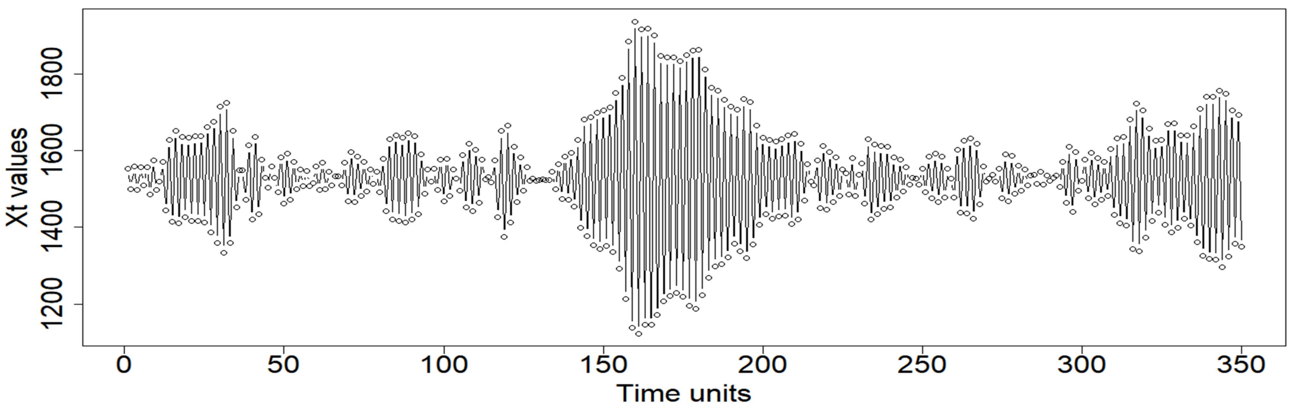

Figure 2. Time series plot of the data

From figure 2, it is noticeable that there is some seasonality in the data. The difference between two adjacent values depend on which time they were observed. For example, this type of trend could happen when the time unit is day of a year and the observations are temperatures of day and night throughout the year such that during the summer, the difference can be large, but in spring or fall, the difference may be relatively small.

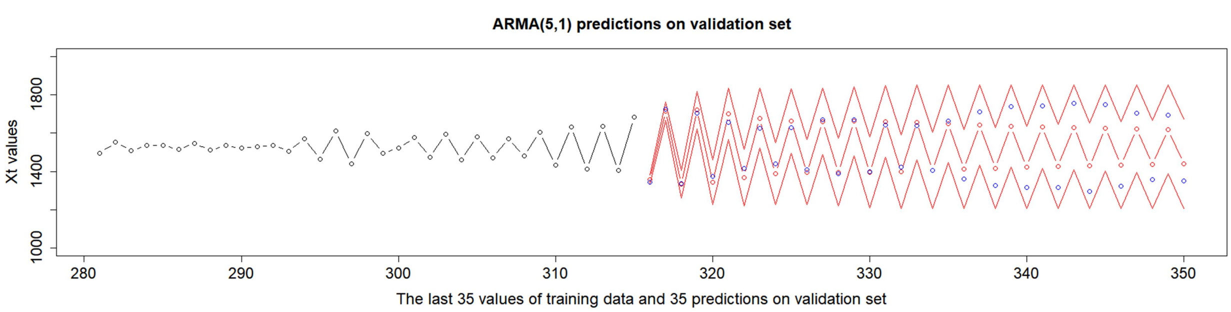

Figure 3. The last 35 values of training data and 35 predictions(red points) on validation set and true values(blue points) using Burg method

In figure 3, the predictions on the nearest 20 values in the validation set seem to be fairly accurate, but then the predictions become closer to the sample mean. The prediction intervals capture the true validation points and yet they are not too wide.

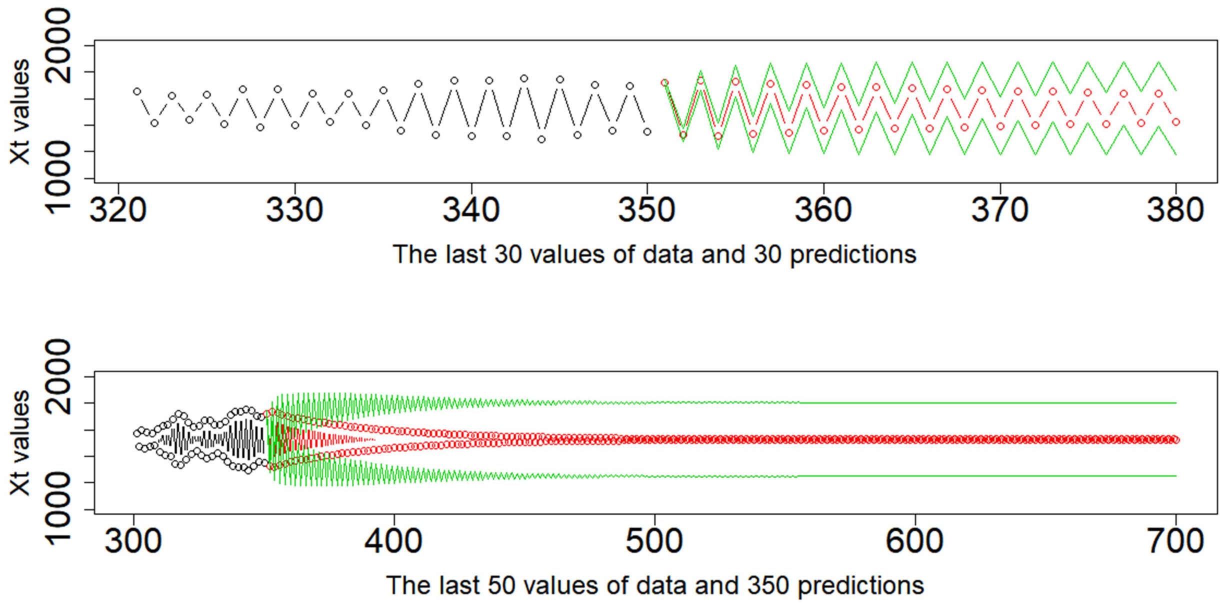

Figure 4. Prediction intervals for the next 30 values and the next 350 values using MLE

Considering the seasonality in the data, the next 30 values seem to be reasonable as they reflect the trend of increase and decrease in the amplitudes of data over time. But once the predictions are made further to the next 350 values, they don't reflect the trend of the data and converge to the sample mean. Also, the length of prediction intervals become as wide as they could. Thus, the predictions more than 30 time units ahead may not be reliable.

The fitted model is

$$X_t = 2028.448 - 0.9164X_{t-1} + 0.7516X_{t-2} + 0.6481X_{t-3} - 0.4240X_{t-4} - 0.3867X_{t-5} + Z_t - 0.9274Z_{t-1}$$

which depends on the past five values.

< Dataset 3 – Appendix >

From figure 2, it is clear that one of the periods is 2, but it is hard to see other periods because of the amplitudes.

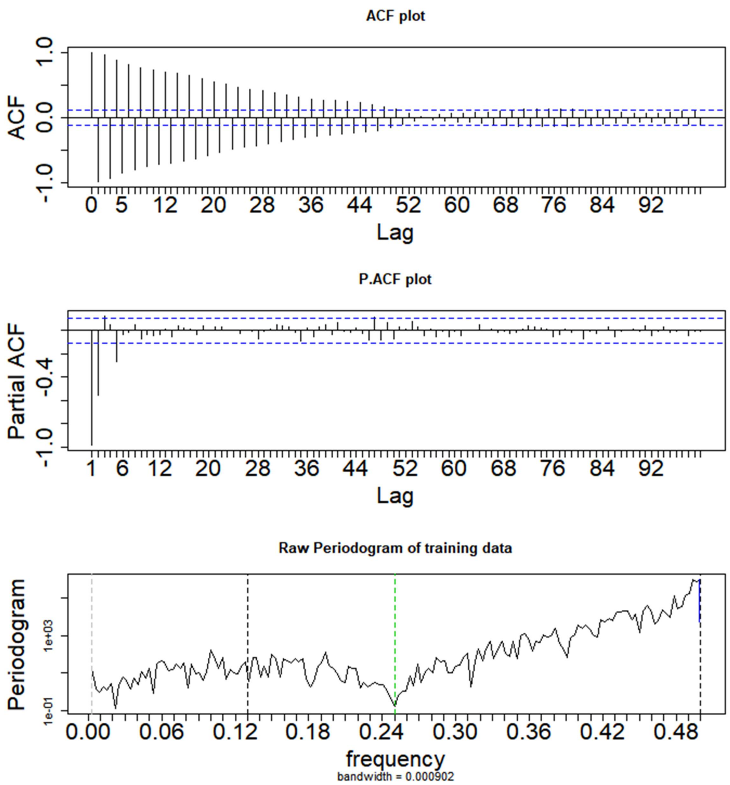

Figure 4. ACF, PACF, and Periodogram of the training data

ACF is decreasing sinusoidally and PACF may cutoff after lag 2 or lag 5. From the periodogram of frequencies scaled from 0 to 0.5, there is one clear peak at 0.5 which corresponds to the period of 2. There is one smooth peak around 0.13 which corresponds to period of 7.7. There are possible dips around 0 and 0.2, but they may as well be the effect of noise from the two local peaks.

Based on the three plots, the candidate models are AR(5), AR(2), ARMA(2,2), and ARMA(5,2).

AR(5) and ARMA(5,2) seemed to overfit the data in their periodograms and β_2 turned out to be insignificant in ARMA(5,2) model. AR(2) and ARMA(2,2) seemed to underfit the data based on the periodogram and the ACF of the residuals were certainly correlated at some lag. Thus, another model between the overfitting and underfitting candidate models had to be chosen. ARMA with p = 3 and p = 4 underfit the data and q = 2 results in insignificant β_2. Thus, ARMA(5,1) was chosen as the final model which also meets the residual assumptions and has all significant coefficients.

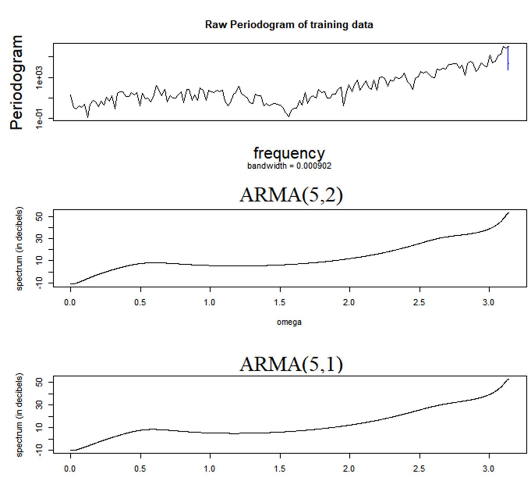

Figure 5. Periodogram, spectrum of ARMA(5,2) and ARMA(5,1)

AR(5) had quite higher AIC and MSE than the other two models and ARMA(5,1) and ARMA(5,2) had similar AIC and MSE. The three models showed some bumpy spectrum around frequency 2.7(scaled from 0 to π) while the periodogram doesn't show a local peak around that frequency. ARMA(5,1) showed the mildest bump at 2.7 thereof showing the most resembling spectrum to the periodogram. The roots of the fitted ARMA(5,1) are two complex conjugate pairs and one negative root with corresponding peaks of the spectrum at frequencies 0.526, 2.658, and π.

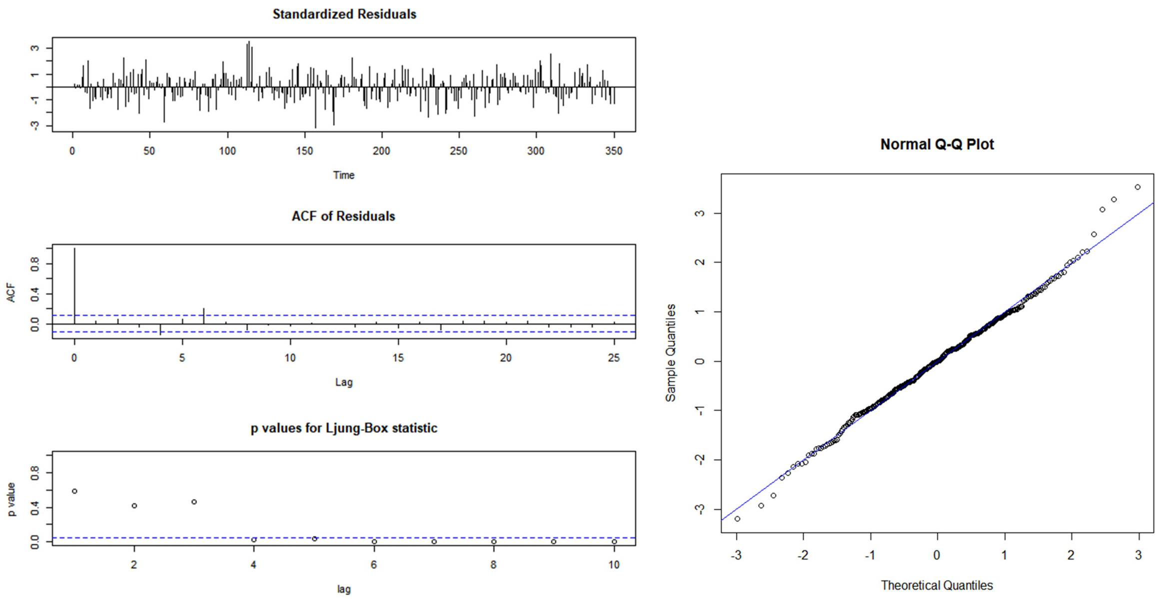

Figure 6. Residuals, ACF, Ljung-Box p-values, and QQ plot of ARMA(5,1)

The final model ARMA(5,1) seems to satisfy the residual assumptions with ACF zero after lag 0 although the p-values of Ljung-Box test shows that at least one of the first four acf's is nonzero, it seems like the acf at lag 4 is only slightly out of the confidence interval. The MLE method used to fit the model assumes normality of the data and residuals and the residuals seem to follow normal distribution based on the qq plot. All the coefficients in the model are significant. AIC is 2718.99 and MSE is 129.5.

< Dataset 4 – Executive summary >

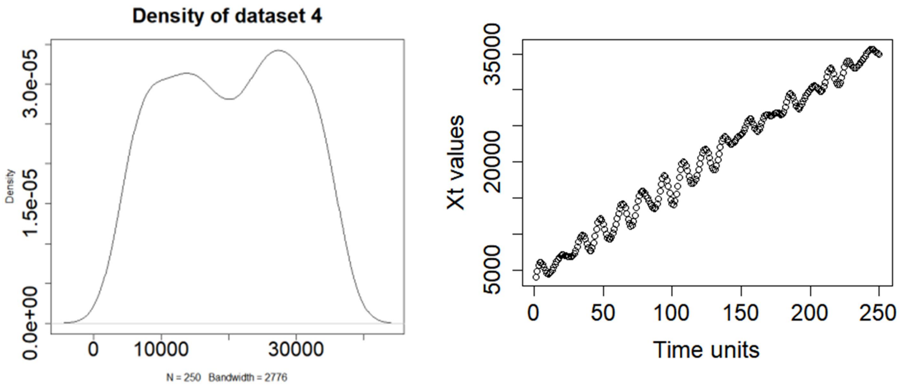

Figure 1. Density(left) and time series(right) of dataset 4

The dataset is bimodal and least-squares method was used to fit the model. The data is sinusoidally increasing throughout the time.

When selecting the model, the last 25 points were held out to compare how well each candidate model generalizes with the new data.

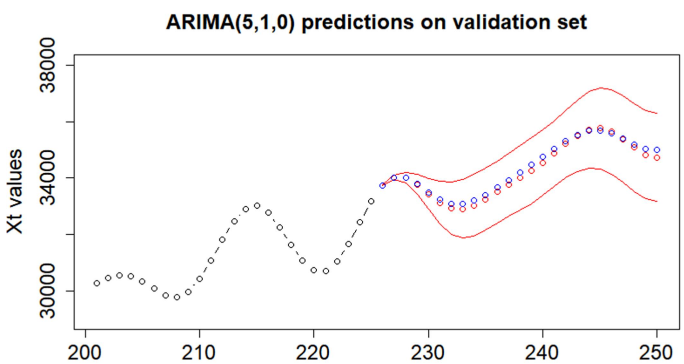

Figure 2. The last 25 values of training data and 25 predictions(red points) on validation set and true values(blue points)

The chosen model's prediction on the validation set is not bad as observed in figure 2. The point estimates are pretty accurate themselves and the prediction interval is not too wide. $\sqrt{\text{PSSE}}$ in the 25 validation points is 137.509 which means the model was roughly wrong in prediction by 137 for the next 25 new data points which is comparatively less than the standard deviation of data which is 9304.

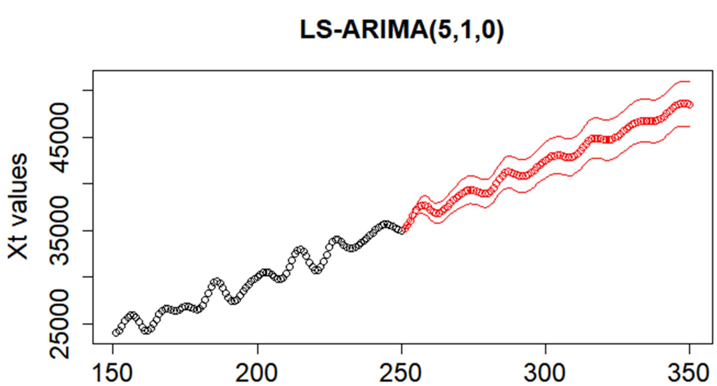

Figure 3. The last 100 values of data and 100 predictions

Figure 3 shows the model fitted with the whole data and the predictions in the next 100 points. The prediction intervals may or may not contain the true values, but it seems to reflect the trend of the data. The width of prediction intervals of the next 80 values is less than a half of the standard deviation of the data and this narrow interval may be desirable since since the model seems to generalize well.

The fitted model is

$$\mathrm{X_t} = 14.24725 + 3.7164X_{t-1} - 5.2463X_{t-2} + 2.6071X_{t-3} + 1.1804X_{t-4} - 1.8912X_{t-5} + .6336X_{t-6} + Z_t$$

which depends on the data$_{t-2}$ more than data$_{t-1}$.

< Dataset 4 – Appendix >

The original data is not stationary with increasing trend but the variance seems to be constant in figure 1.

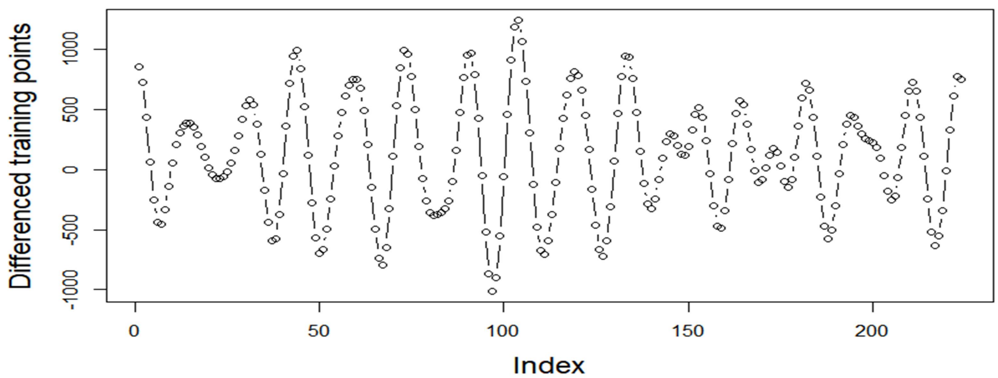

Figure 4. The training data after differencing by 1

After differencing by 1, the data seems second-order stationary. There seems to be small peaks and tall peaks with regular periods. The total number of peaks may be 19 and the number of tall peaks may be about 13, thus there may be periods of 13 and 25 in this dataset.

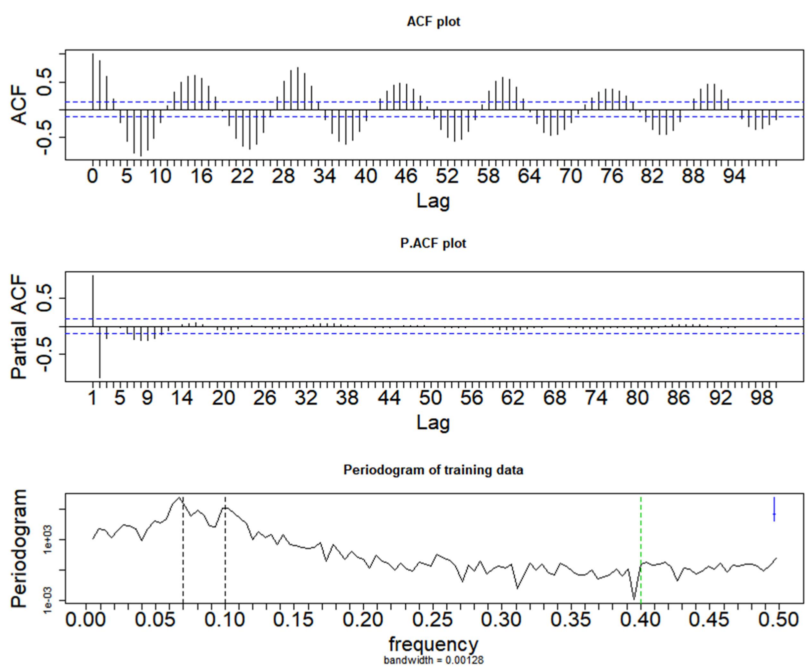

Figure 5. ACF, PACF, and Periodogram of the differenced training data

There is a clear sinusoidal decay in ACF plot and PACF cuts off after lag 2 then starts sinusoidal decay. In the periodogram, there are two local peaks and possibly one dip around 0.4. The two peaks correspond to the periods of 14 and 10 which may be the periods observed in figure 4. Since there is some sinusoidal decay after PACF at lag 2 and a possible dip in the periodogram, one could say there is an MA component. Thus, the initial candidate models were ARIMA(4,1,0), ARIMA(4,1,1), and ARIMA(4,1,2). But it turns out that these models do not satisfy the residual assumptions that they are uncorrelated and the spectrum of these models has much sharper peaks compared to the periodogram although all the coefficients are significant. Thus, higher ordered models ARIMA(5,1,0), ARIMA(5,1,1), ARIMA(5,1,2), and ARIMA(6,1,0) were compared. The ARIMA models with p = 5 satisfies the residual assumptions and ARIMA with q ≥ 1 show insignificant MA coefficients. ARIMA with p = 6 has insignificant 6^{th} AR coefficient, so p = 5 seems to be the proper order of AR and there may not be an MA component. The spectrum of all these models still seem to have sharper peaks than the periodogram.

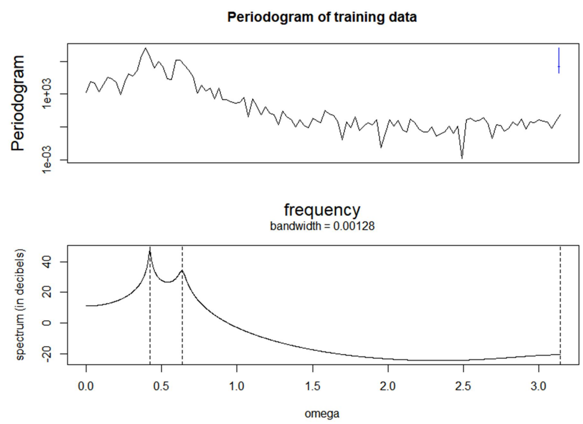

Figure 6. Periodogram of the differenced training data and Spectrum of the ARIMA(5,1,0)

The final model is ARI(5,1) and its roots are two conjugate pairs and one negative real root which correspond to the frequencies .422, .637, and π. The coefficients are significant except for the third one, but following the principle of hierarchy, all 5 coefficients are included in the model.

< Dataset 5 – Executive summary >

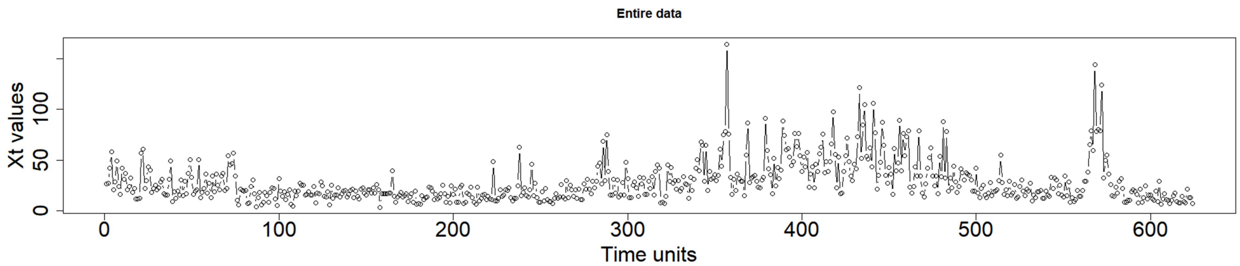

Figure 1. Time series plot of the deposits data

The data is noisy and there are occasional high deposits irregularly. There also seems to be a short period of a repeated pattern in every 5 to 7 years. There is a noticeable outlier(357th observation) which seems to cause the inflated variance in deposits for the next century.

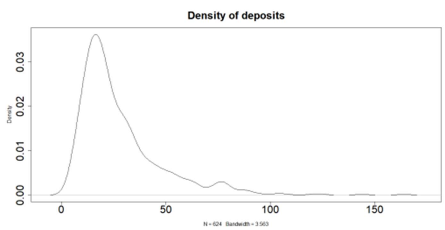

Figure 2. Density of the deposits data

The data is skewed to the high values which explains the noise in the time series plot.

Deposits data have 624 observations and 62 were held out as validation set for model selection.

Based on the final model, the data does not have regular periodicity, thus the predictions are rather constant.

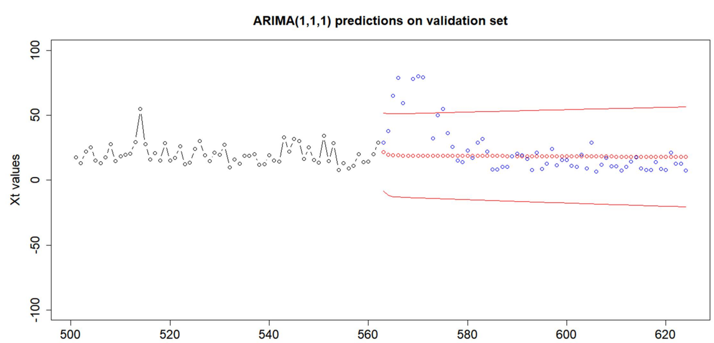

Figure 3. The last 62 values of training data and 62 predictions(red points) on validation set and true values(blue points)

From figure 3, the 62 predictions on validation set are very similar to each other as they range from 15 to 17 except for a couple of points and these are also similar to the sample mean of the last 62 training points in the plot. The prediction intervals are very wide because of the several outliers between 350th and 450th years from the training data, but the lower bounds of the interval may need to be ignored since the data in its nature should be positive values. The model is indicating that data is somewhat close to white noise, but the prediction interval takes into account of the variability of the data such that it manages to capture 90% of the validation set. In the validation set, there is an extreme outlier(568th observation) so the noise caused by this outlier in the next few time units could not be predicted well by the model fit to the training set. Except for the values between 563th and 577th years, $\sqrt{\text{PSSE}}$ is 6.4 which means the predictions are about 6.4 values away from the true values on average. It is comparatively less than the standard deviation of the training data which is 19.4. Except for the noise, the model may be able to predict better than the sample mean.

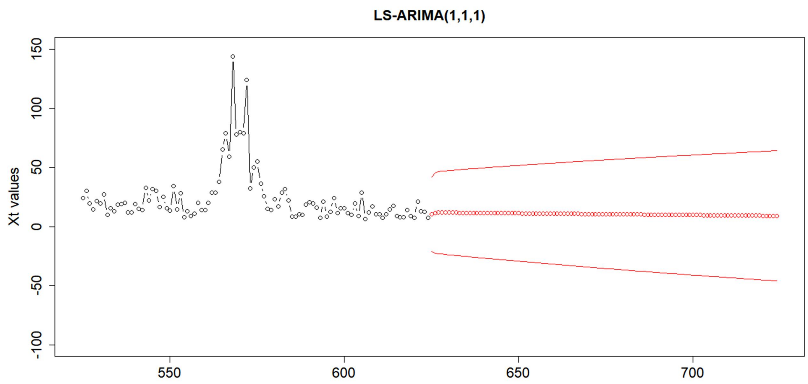

Figure 4. The last 100 values of data and 100 predictions

The future predictions show quite similar results to the predictions on the validation set in terms of wide prediction intervals and somewhat constant predictions.

The fitted model is

$$\mathrm{X_t} = -0.01986768 + 1.3196X_{t-1} - .3196X_{t-2} - .9061Z_t$$

which depends about 4 times more on the previous data than the one before that.

< Dataset 5 – Appendix >

The ACF of the data was decreasing too slowly, thus it was differenced by 1.

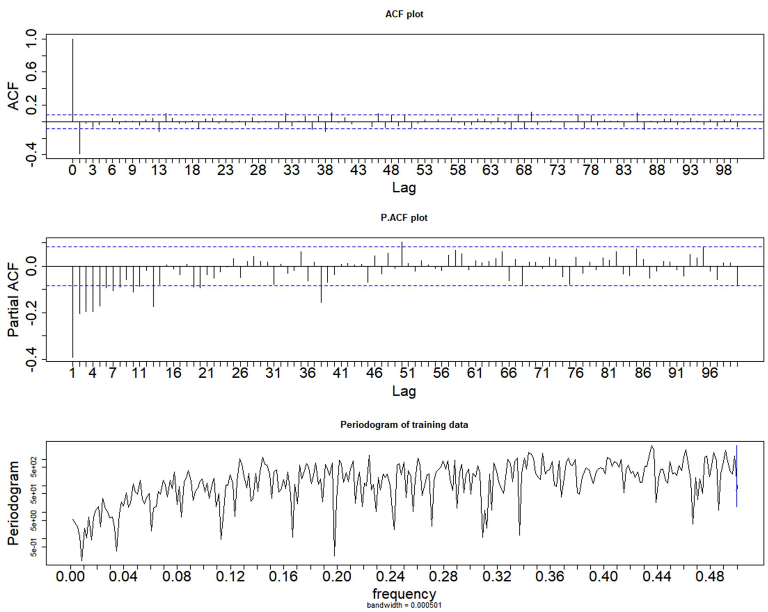

Figure 5. ACF, PACF, and Periodogram of the differenced training data

After differencing by 1, the ACF cuts off clearly after lag 1. PACF may be cut off after lag 1 or lag between 2 and 5. It doesn't seem like there is a sinusoidal decay after then.

In the periodogram, it is hard to tell whether there is any dips or peaks due to the noise in the data. There may be a mild peak between 0 and 0.5 if the true PACF cuts off after lag between 1 and 5. There may be a dip around $\theta = 0$ which means there is one positive real root in MA polynomials and this matches the result from ACF plot.

Based on ACF, PACF, and periodogram, the candidate models are ARIMA(1,1,1), ARIMA(2,1,1), ..., ARIMA(5,1,1). All these five models satisfied the residual assumption of white noise, but only ARIMA(1,1,1) and ARIMA(3,1,1) had significant coefficients in the model. The second coefficient of AR in ARIMA(3,1,1) is insignificant, but the third coefficient of AR is significant and all these coefficients are included in the model following the principle of hierarchy.

ARIMA(1,1,1) and ARIMA(3,1,1) have similar predictions and prediction intervals in the validation set. ARIMA(1,1,1) has lower PSSE and MSE than ARIMA(3,1,1), thus it was chosen as the final model. The AR coefficient of ARIMA(1,1,1) indicates that there is no periodicity in the data which means that seemingly weak periodicity observed in the time series plot may have been just noise.

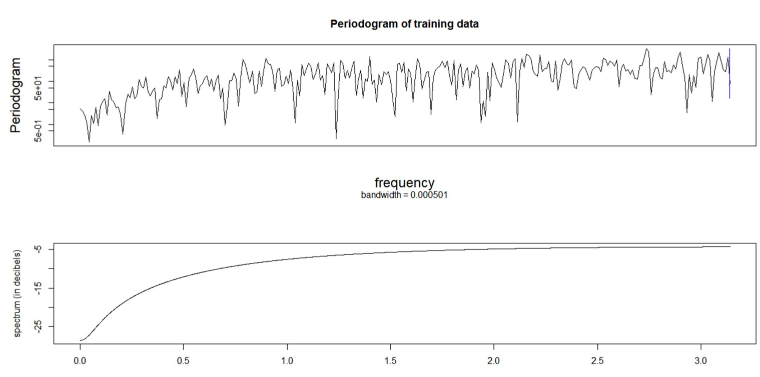

Figure 6. Periodogram of the differenced training data and Spectrum of the ARIMA(1,1,1)

The MA component of ARIMA(1,1,1) shows that there is a dip at $\theta = 0$ which is reflected in the spectrum in figure 6. Comparing the periodogram to the spectrum of the model, it is hard to tell whether the model is reasonable due to a lot of noise reflected in the periodogram.

< Dataset 6 – Executive summary >

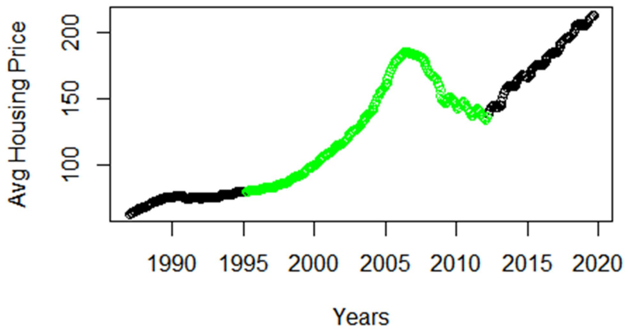

Figure 1. Time series of housing dataset

In figure 1, the housing bubble in U.S. started to form around January 2000 as the average housing price started to increase rapidly and reached its peak during the second quarter of the year 2006. Then it plunged until February 2012. After it reached the lowest point, the housing price started to increase rather more quickly than the pre-bubble era. A model fit to the data before the bubble formation shows that there is clearly after-effect.

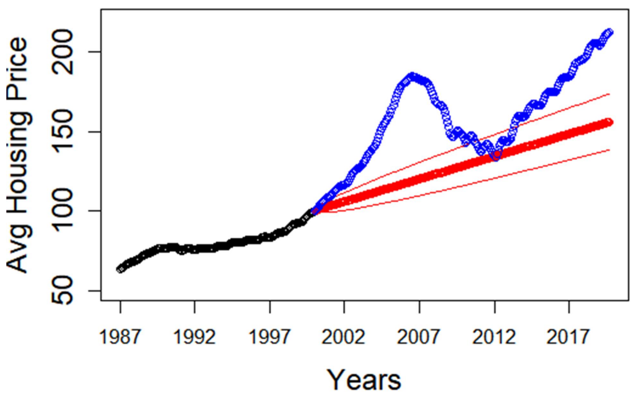

Figure 2. Estimated housing price without the bubble(red)

Figure 2 shows a clear gap between the slope of the actual price after 2012 and the slope of the predicted price after 2012 by the model assuming the bubble did not exist. The price returned to the expected level for only four months between December 2011 and March 2012, then it jumped back very quickly at the almost same rate as the rate of the bubble formation between 2000 and 2006. This indicates that even after the bubble burst, the housing price is still pulled up by its after-effects.

It may not be surprising that the after-effects usually remain in real-estate market. Using the data until March 2012 which is the time before the price increased rapidly again, a model was fit to see if the after-effects were also what could have been expected.

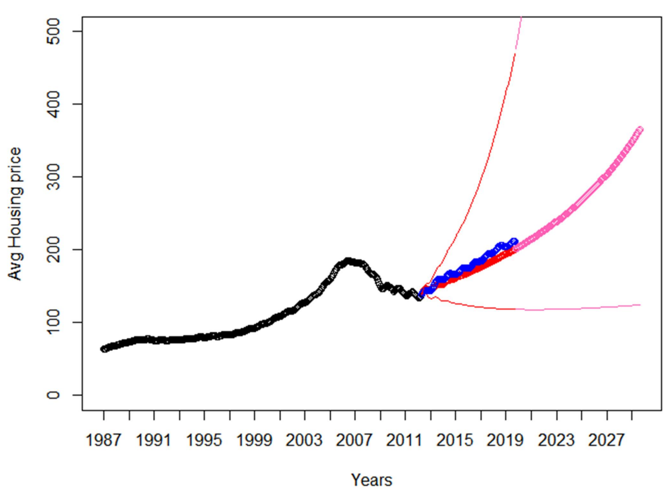

Figure 3. Estimated housing price after March 2012(red) and predicted price after September 2019(pink)

The extremely wide 95% prediction intervals of the model indicates high uncertainty in the fluctuation of the prices, but its point estimates are close to the true trend of the price after March 2012 with $\sqrt{PSSE} = 8.8$ which means the estimated price between April 2012 and September 2019 is only 8.8 units away from the actual price on average. This shows the after-effects of the bubble could have been expected to some extent. The pink dots and lines are the predicted prices of this model for the next 10 years which show that the prices are likely to increase exponentially but with uncertainty growing extremely large as well.

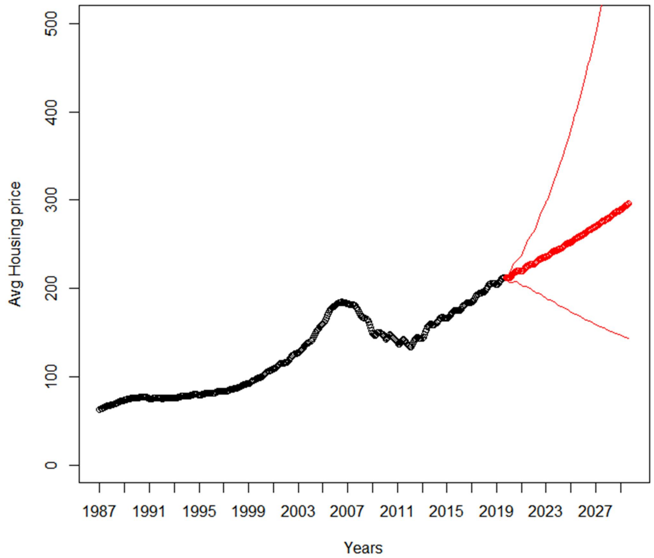

Figure 4. Predicted price after September 2019(red)

Final model using the entire data shows a similar results as before but with more gradual change in the price. There is a chance that the price will increase exponentially as years go by, but there is also a chance it can plummet like the period between 2006 and 2012.

The fitted model is

$$\mathrm{X_t} = (.999993448 - .4431X_{t-1} + .3109X_{t-2} - .00986X_{t-3} - .1185X_{t-4} + .07778X_{t-5} + .0152X_{t-6} - .07592X_{t-7} + .0672X_{t-8} + .02994X_{t-9} - .05712X_{t-10} + .05088X_{t-11} - .2Z_t)^{-5}$$

which depends highly on the past two immediate prices but requires about a year-long data to predict the next month's average housing price.

< Dataset 6 – Appendix >

From figure 1, the sharp increase in the price after the bubble burst left suspicion that the after-effect of the bubble was driving the price up very quickly until now and the future. Thus, a model was fit to the pre-bubble data before the year 2000 to measure the after-effects as of the year 2012 by comparing its estimated prices to the actual prices.

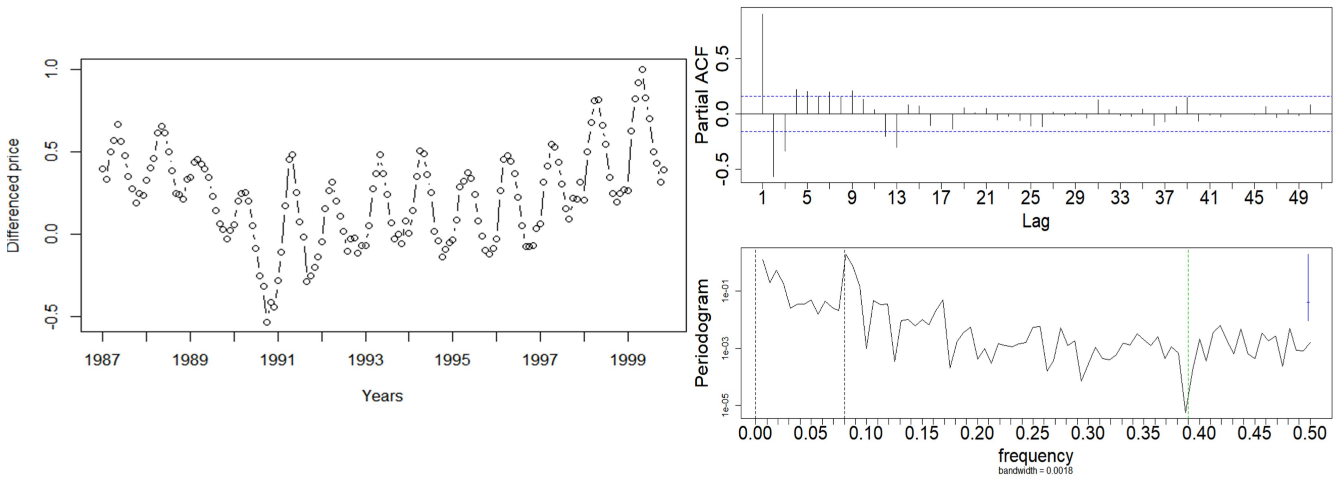

Figure 5. Time series plot, PACF, Periodogram of the Differenced monthly average housing price from 1987 to 1999

The 156 monthly average housing prices until the year 1999 showed linearly increasing trend, thus it had been differenced. After differencing, there seems to be a regular period of about 10. The ACF showed a sinusoidal decaying pattern and PACF seems to cut off after lag 2 or lag 3. The periodogram showed two possible peaks around 0 and 0.08 which correspond to no period and period of 12 respectively. There may be an annual seasonality involved in the differenced data with noise. There is a possible dip around 0.39.

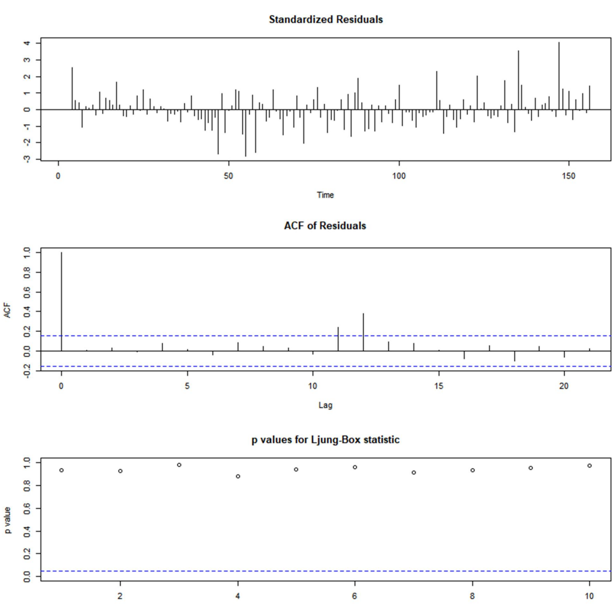

Figure 6. Standardized residuals, ACF, Ljung-Box p-values of the residuals of ARIMA(2,1,2)

Based on ACF, PACF, and periodogram, the candidate models are ARI(2,1), ARI(3,1), ARIMA(2,1,2), and ARIMA(3,1,2). The models were fit using least-squares method instead of MLE method since the data is skewed to the low prices. The residuals of ARI(2,1) were unlikely to be white noise and ARIMA(3,1,2) had insignificant coefficients. ARIMA(2,1,2) had seemingly white-noise residuals as in figure 6 and less prediction errors than ARI(3,1).

The ARIMA(2,1,2) predictions were compared to the actual prices after 2000 in figure 2. If there wasn't any after-effects, they should have been very close after March 2012. But the only time that the price has returned to its place was between December 2011 and March 2012. It is clear that the after-effects exist based on this model, thus another model to measure the effect was created. The model is trained on the data until March 2012 to compare the prices affected by the estimated after-effects and the actual after-effects. If the estimation is quite accurate, then the model may be reliable to predict the future prices as of September 2019.

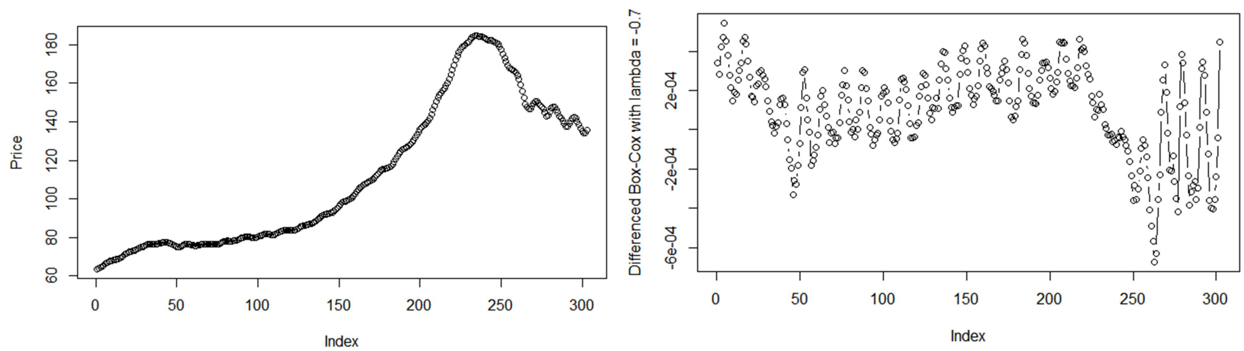

Figure 7. Time series of pre-transformation(left) and transformed data(right) until March 2012

The data including the bubble has polynomial trend which requires rather complex transformation to detrend. The transformation considered was differenced Box-Cox power transformation where the MLE power λ is -0.7. Based on ACF, PACF, and periodogram, ARI(1,1), ARI(2,1), ARI(7,1), and ARI(9,1) were considered to be candidate models.

ARI(1,1) and ARI(2,1) have a sinusoidal decay in residuals and ARI(7,1) and ARI(9,1) approximately satisfies the residual assumptions of white noise and have significant coefficients. There seems to be no difference between ARI(7,1) and ARI(9,1) in terms of PSSE and MSE. Thus, ARI(7,1) was chosen as the model as in figure 3.

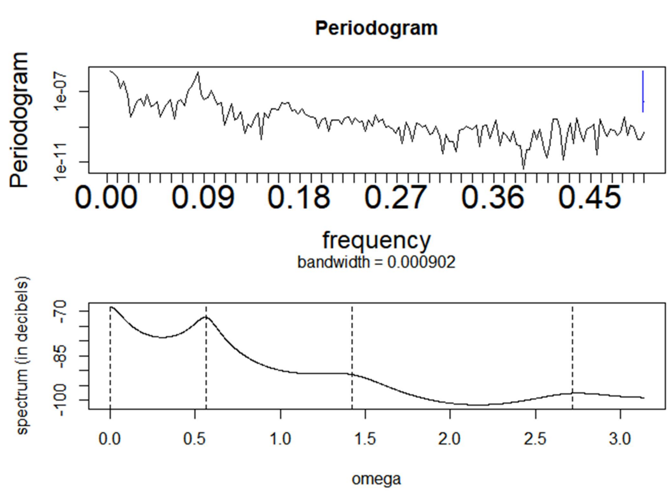

Figure 8. Periodogram of transformed data until March 2012 and Spectrum of ARI(7,1)

The spectrum of ARI(7,1) seems to match the periodogram with the model's frequencies which correspond to the periods of none, 11, 2.3, and 4.4 respectively. The frequncies corresponding to the periods of 11 and none seem to be stronger than the others which verifies again that there is annual seasonality in the differenced housing prices and also a lot of noise. Since the model produced quite good estimated prices between April 2012 and September 2019, its point estimates may be reliable for predicting the prices as of September 2019 as well.

The final model with the whole dataset using the same Box-Jenkins procedure with the same type of transformation as did with ARI(7,1) may as well be reliable for predicting the future housing price.

Based on ACF, PACF, and periodogram, the candidate models were ARI(1,1), ARI(2,1), ARI(4,1), ARI(7,1), ARI(9,1), ARI(10,1), and ARI(13,1).

ARI with $p \leq 9$ do not have uncorrelated residuals. Final model chosen is ARI(10,1) with fewer coefficients approximately satisfying the residual assumptions.

As the data contains more complex structure, the order of AR polynomial also seem to increase.

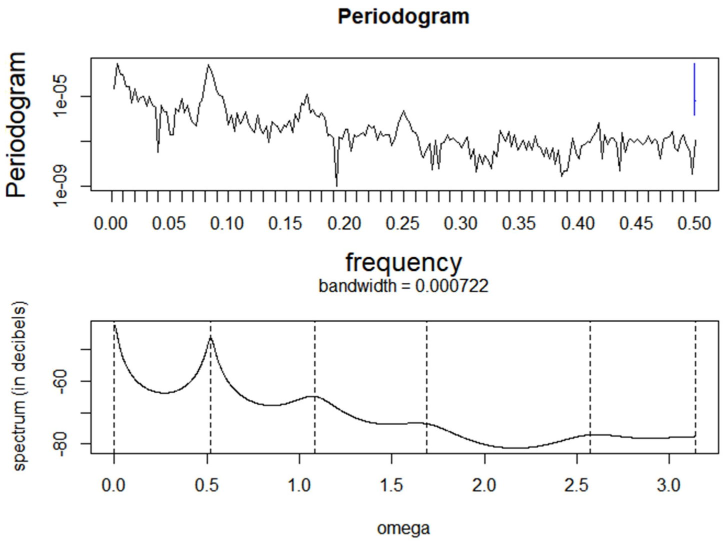

Figure 9. Periodogram of transformed entire data and Spectrum of ARI(10,1)

The spectrum of ARI(10,1) matches well with the periodogram of the data. The frequencies of the model correspond to the periods of none, 12, 5.8, 3.7, 2.4, and 2.

Fitting ARI(10,1) to the whole data shows similar results as the model fit to the data before March 2012 as shown in figure 3 and 4 in a sense that the predicted prices as of September 2019 are increasing with high uncertainty. Both models tell us that while there's a chance that the price will increase exponentially, there is also a chance that it will drop to the price of March 2012 after 10 years from now. The difference between these models is that the increasing rate estimated by the final model tin the next 10 years is somewhat close to the rate in the previous 10 years.

There was clearly an after-effect of the housing price bubble happened in mid 2000's which made the prices increase even faster than they used to in the pre-bubble era of 90's and 80's. The current housing prices are still in effect of the bubble and it may be possible that there will be another bubble in the next 10 years based on the models although the point estimates only go up.

The model fit to the data before March 2012 was quite reliable since it predicted prices very close to the actual prices for 7 years between April 2012 and September 2019. In that sense, this model could be useful for people who plan to buy their house in a couple of years. It is noteworthy that while the lower bound of the prediction interval is constant and similar to the price in the year 2012, the upper bound is increasing dramatically. For example, if a person buys a house right now, the maximum loss is about 100 units after 10 years. If s/he waits another 10 years, s/he might have to pay at least 200 more than now along with the risk of losing up to 250 if the price plunges.

It is reasonable to be in fear of purchasing the house at a bubble's peak since it may take 10 or more years for the price to be back to the price of purchase. If the person is sure that s/he will not sell the house for more than 15 years or so, s/he will be able to retrieve the market value of the house. In practice, this may be the case of a person looking for settlement in an area with good economy and plans to pay mortgage for the next 30 years or so.

In local areas with good economy with high employment rate, the housing price keeps going up as long as mortgage rate is affordable for many people. If the price increases gradually, people would not suddenly stop buying houses, but if the economy cannot keep up with the price, then there will be a sudden drop soon. Thus, one would have to check whether the increasing rate is consistent with the local economic growth before deciding to purchase a house.

< Reference >

Chatfield, C. (2004). The analysis of time series: an introduction. Florida, US: CRC Press.

R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Venables, W. N. & Ripley, B. D. (2002) Modern Applied Statistics with S. Fourth Edition. Springer, New York. ISBN 0-387-95457-0

United States housing bubble. (2006). In *Wikipedia*. Retrieved December 17, 2019. from https://en.wikipedia.org/wiki/United_States_housing_bubble