ZILLOW TWO BEDROOM HOUSING REPORT

A Time Series Analysis

Lewis, Yolanda

Non-Technical Summary

Zillow is a well-known online real estate company that provides consumers with the ability to search, buy, sell, rent, or finance housing. Zillow maintains housing data in which they provide a seasonally adjusted measure of a typical home value based on region and housing type. Zillow refers to this measure as the "Zillow Home Value Index (ZHVI). In this report, I focus on 2-bedroom housing price data for large and small housing market analysis. More specifically, I will focus on Chicago, IL and Minneapolis, MN. It should be noted that based on the smoothed seasonally adjusted data collected from Zillow, second order differencing was required to achieve stationarity.

According to this study, between 1996 to 2020, the Chicago housing market has not yet recovered from the 2008 housing market crash. The Chicago 2-bedroom housing prices had a lot of volatility between 2008 and 2015 and resembled an up and down pattern from 1996 to 2020. The model used to predict Chicago housing prices shows a decrease in housing prices followed by an increase later in the forecast. The housing prices did not appear to re-bound to the housing prices prior to 2008 and continued to level off around 2018.

Technical Summary

To begin exploratory analysis on the Chicago data, a times series object was created using the adjusted housing price (price was adjusted using the customer price index). Refer to figure 1.

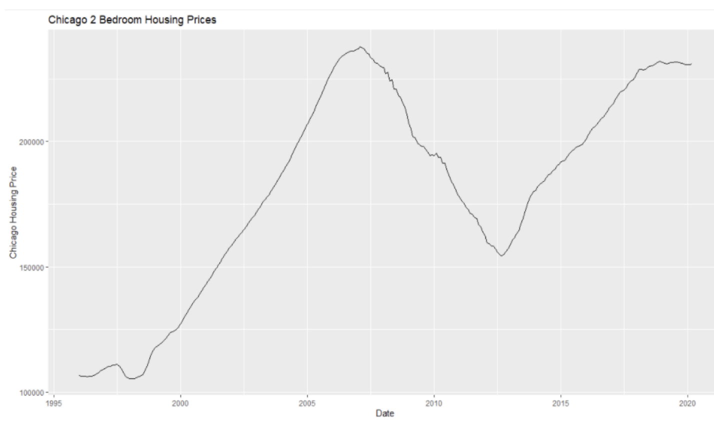
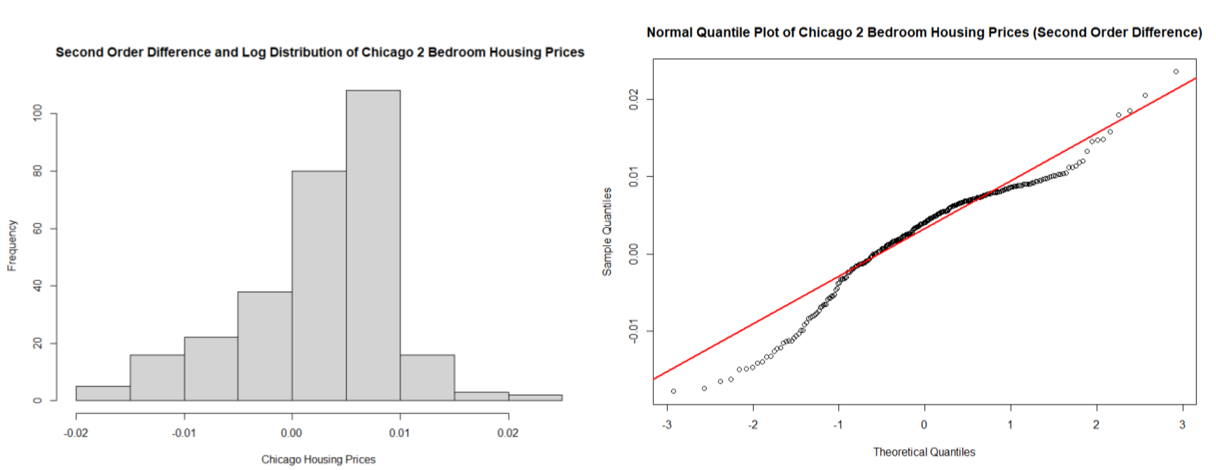


Figure 1: Housing Price - Chicago

The series appeared to be multiplicative and possibly a random walk. The normality of the series was then explored using a histogram, qqplot and the jarque-bera test. The series did not appear to be normally distributed, so the difference and the log of the series were taken to improve approximation to normal distribution. Based on the normality observation, the null hypothesis that the data is normally distributed is rejected. The data was transformed using second order differencing and a log transformation. The normality was increased, but the data was not fully normalized. Stationarity was checked by performing the Dickey-Fuller and KPSS tests. Second order differencing was required to get the series to become stationary. Based on the ADF test (p-value = 0.01) the null hypothesis that the series is non-stationary is rejected. Also, based on the KPSS test (p-value = 0.1) we would fail to reject stationarity. See figure 1.




```
Title:
 Augmented Dickey-Fuller Test

Test Results:
  PARAMETER:
    Lag Order: 1
  STATISTIC:
    Dickey-Fuller: -10.4832
  P VALUE:
    0.01
```

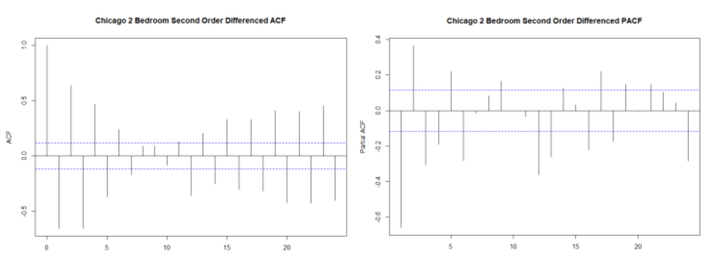
```
        KPSS Test for Level Stationarity

data:  transformed_CHI_Housing_2
KPSS Level = 0.033174, Truncation lag parameter = 5, p-value = 0.1
```

Figure 1. Normality and Stationarity Checks

<u>Model Fitting</u>

To determine an appropriate model for the city of Chicago, the ACF, PACF and the EACF of the series was observed. The ACF suggested that the series exhibited MA behavior with a negative autocorrelation at lag -1. The PACF suggested a negative correlation at lag-1 and AR behavior throughout. The EACF suggested a possible ARMA (1,3) or ARMA(2, 4) for the Chicago series. See figure 2 below.




```
AR/MA
  0 1 2 3 4 5 6 7 8 9 10 11 12 13
0 x x x x x x x o o o x  x  x  x
1 x o x o o o o o x o o  x  x  o
2 x o x x o o o o x o o  x  x  o
3 x x x o o o o x o o o  x  o  o
4 x o x x o o o o o o o  x  x  o
5 x o x o o x o o o o o  x  o  o
6 o x x o x x o o o o o  x  x  x
7 x o x x o x o o o o o  x  x  o
```

Figure 2: ACF, PACF, and EACF of Second Order Differenced Chicago Series

To build a model, several combinations of AR and MA behavior were explored. Three candidate models were selected based on performance. The final model (model 1) selected was SARIMA (1,0,3)(0,0,1)[12] which contained the lowest AIC, BIC, mean absolute percentage error, and was the most parsimonious model. Model 2 presented below was less parsimonious and had a larger mean absolute percentage error. Model 3 was obtained using auto arima with “bic” . Model 3 was not considered due to the number of parameters and the mean absolute percentage error of the model. The models are presented below in figures 3 and 4.

Model No.	Model	No. of Parameters	AIC	BIC	Sigma^2	MAPE	Residual Analysis
1	Sarima(1,0,3)(0,0,1)^{12}	5	-2760.28	-2734.61	3.94e-06	1.1%	White Noise
2	Sarima(2,0,4)(0,0,1)^{12}	6	-2761.87	-2732.54	3.47e-06	6.56%	White Noise
3	Sarima (3,0,2)(2,0,2)^{12}	9	-2766.41	-2729.76	3.55e-06	1.4%	White Noise

Figure 3: Candidate Models for Chicago Housing Market

```
Series: transformed_CHI_Housing_2
ARIMA(1,0,3)(0,0,1)[12] with non-zero mean

Coefficients:
          ar1     ma1     ma2      ma3     sma1   mean
      -0.7318  0.5342  0.4254  -0.3555  -1.0000  0e+00
s.e.   0.0500  0.0634  0.0592   0.0636   0.0702  1e-04

sigma^2 estimated as 3.494e-06:  log likelihood=1387.14
AIC=-2760.28   AICc=-2759.88   BIC=-2734.61
```

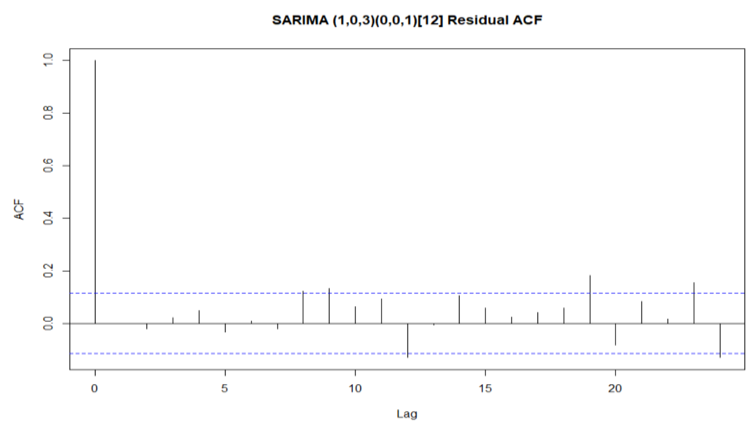
```
z test of coefficients:

             Estimate  Std. Error  z value  Pr(>|z|)
ar1       -7.3176e-01  4.9963e-02 -14.6459 < 2.2e-16 ***
ma1        5.3420e-01  6.3439e-02   8.4207 < 2.2e-16 ***
ma2        4.2540e-01  5.9241e-02   7.1809 6.923e-13 ***
ma3       -3.5549e-01  6.3624e-02  -5.5874 2.305e-08 ***
sma1      -9.9998e-01  7.0242e-02 -14.2362 < 2.2e-16 ***
intercept -1.4060e-05  6.0394e-05  -0.2328    0.8159
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 4: SARIMA (1,0,3)(0,0,1)[12] Model Coefficients

Residual analysis and model diagnostics

Residual Analysis was also performed for the Chicago SARIMA (1,0,3)(0,0,1) [12] model. The residual analysis was performed using the Ljung box test, the dickey fuller test, and the KPSS test for verification of stationarity. The Ljung box test contained a p-value of 0.9354, which indicated that white noise is present. The Dickey full test contained a p-value of 0.01 which rejects non-stationarity. The KPSS test contained a p-value of 0.1 which fails to reject stationarity. The ACF of the model residuals was observed with some seasonality or autocorrelation still present in the residuals. Some volatility appeared to still be present in the residuals especially during the housing market crash in 2008 until roughly 2015. See figure 5.



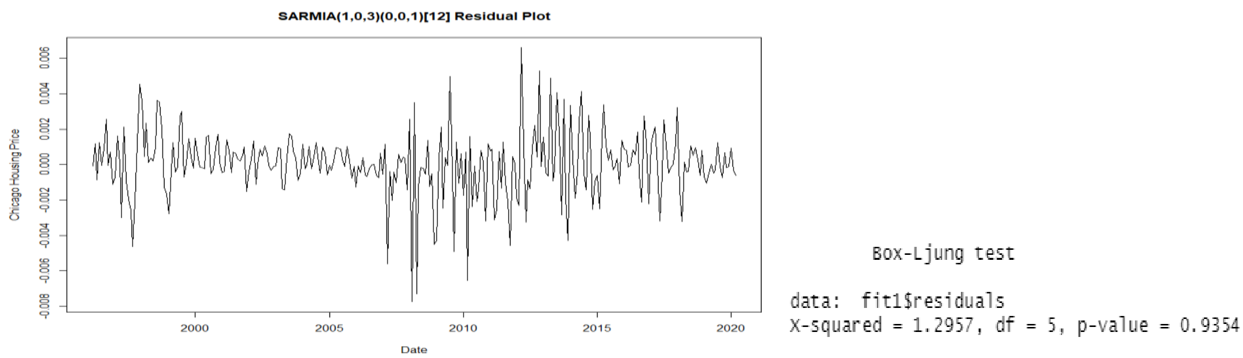


Figure 5: Residual ACF Plot, Time Series and Ljung Box Test

Forecast analysis

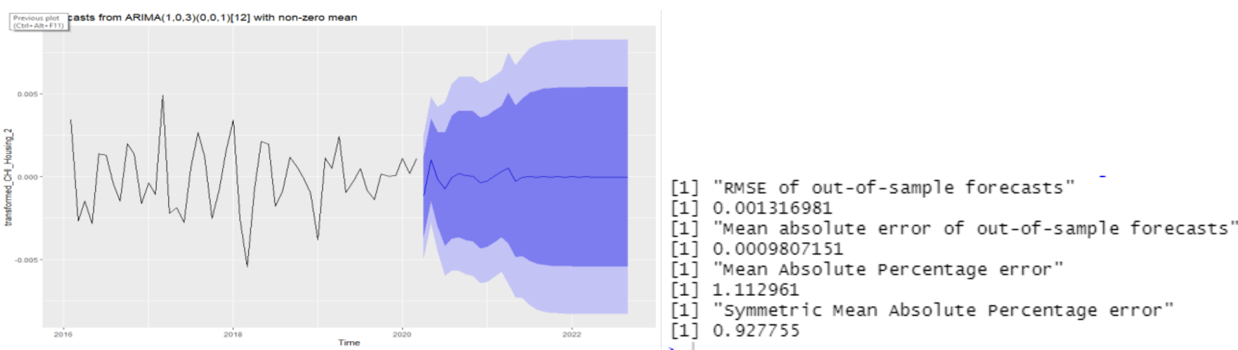


Figure 6: Chicago SARIMA (1,0,3) (0,0,1) [12] Forecast and Back Test

The final Chicago model was used for a 30 step ahead forecast for Chicago 2-bedroom housing prices. The model was validated using a backtest with 80 % of the Chicago housing market data. The root mean squared error of the model was 0.013% and the mean absolute percentage error was 1.11%. The final forecast appears to continue the up and down movement as seen previously in the series and eventually reverts to the mean of the series. The confidence bands are representative of the range of Chicago housing prices and the confidence bands widen as the forecast diverges to the mean.

Small and Large Housing Market Comparison

Chicago and Minneapolis were analyzed to determine if there was any cross-correlation between a larger housing market and a smaller housing market. From the plot of Chicago and Minneapolis prices, it was discovered that the growth patterns of Minneapolis and Chicago are similar. In fact, Minneapolis prices appear to be surpassing Chicago as of about 2018. To explore this pattern, a correlation plot was observed for the two cities. Based on the ACF plot, Minneapolis and Chicago are cointegrated with the highest correlation value 0.982 present at lag 0. Lagged regression was performed to predict Chicago housing prices using Minneapolis housing prices as an independent variable. Minneapolis was significant and the estimate was useful for the lagged regression model. Residual analysis was conducted using the Dickey- Fuller and KPSS tests which indicated residuals of the model were stationary. The residuals exhibited white noise. The lagged regression model indicated that Chicago and Minneapolis are cointegrated. Refer to figures 7 and 8 below.

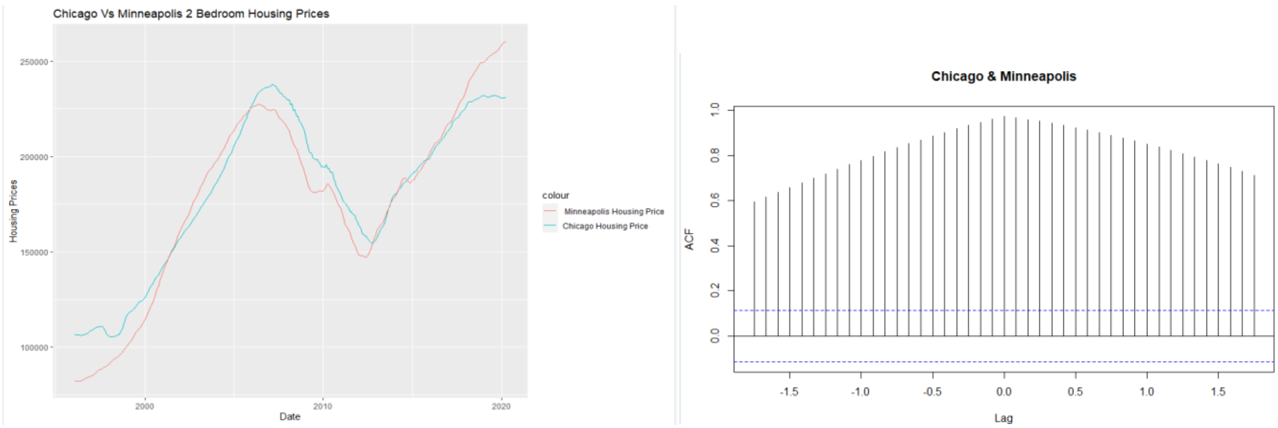


Figure 7: Chicago and Minneapolis Comparison and Correlation Plot

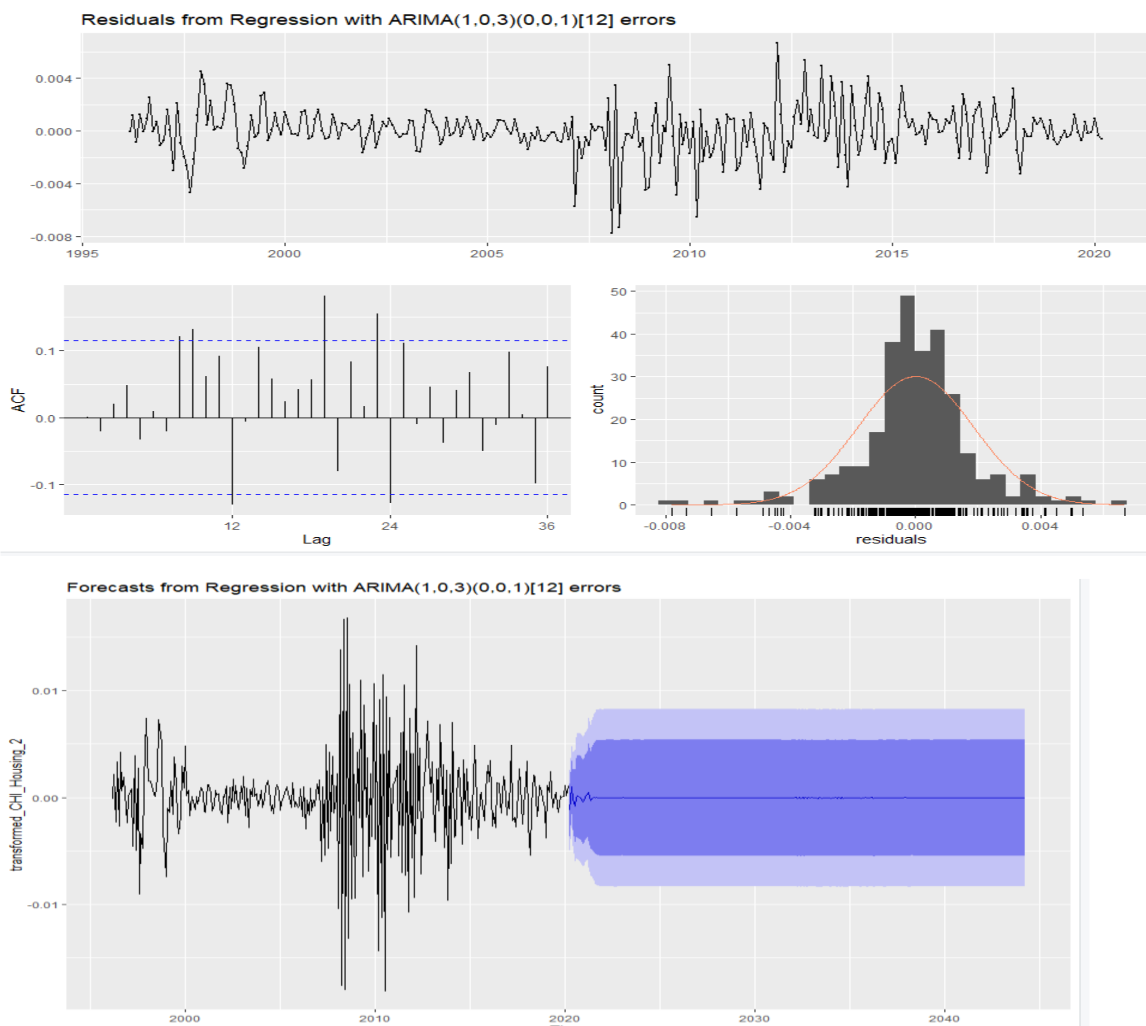


Figure 8: Chicago and Minneapolis Regression Residual Plots

Analysis of Results and Discussion

The time series required second order differencing to achieve stationarity due to the smoothed data provided by Zillow. The Chicago housing market has not fully returned to the 2-bedroom housing prices seen prior to 2008. The modeling performed for the Chicago market indicates that the housing market will continue to have an increasing and decreasing pattern in the housing prices. The coefficients of the final SARIMA model explain that the final housing prices were mainly influenced by the prior shock (random effects) since most of the coefficients were explained by ma behaviors. Lagged regression appeared to be the best modeling technique when comparing Chicago and Minneapolis housing markets. Chicago and Minneapolis Housing markets were cointegrated and showed a similar growth pattern despite Minneapolis being smaller housing market. Minneapolis housing prices were found to exceed Chicago housing prices after 2018.