Национальный исследовательский университет

«Высшая школа экономики»

Факультет мировой экономики и мировой политики

Образовательная программа «Мировая экономика»

Отчет о самостоятельной работе

по дисциплине «Эконометрика» на тему

«**A Study of the Real Estate Market in the United States of America**»

Выполнена:

Фам Тху Чанг, БМЭ204

Москва 2023

Table of Contents

Introduction

All around the world, the real estate market has always been busy with potential buyers seeking for available houses while owners offer their listings with corresponding prices. In the US, except for the largest commercial cities such as New York, Los Angeles, or Chicago, it is not rare for citizens to look for a house in other cities of the US that are convenient for them to go to work or for their kids to go to school. However, since global financial crisis (GFC) in 2007 – 2009, the real estate market in the US has become a topic for people around the world, considering that it was the root of the crisis that turned the US and whole world to a new direction.

What now has become common knowledge is that Great Recession from December 2007 to June 2009, which was the consequence of the GFC, originated from its housing market bubble. Normally, it is understandable for banks and lending institutions to offer lower mortgage rates to incentivize potential homebuyers to borrow money from them. However, at the start of the crisis, the fact that mortgage rates were lowered to unrealistic levels to attract borrowers that would not be able to pay their debts led banks to the financial crisis that affected millions of businesses in the US and hence, around the world.

In this research, we aim to study the US real estate market from as long as the figures were recorded before the GFC (1987) up till the end of 2021 so as to see how house prices and mortgages rates behaved in both normal times and during chaotic times. The scope of our study, though smaller, remains relevant as the US has just witnessed another crisis that also made an impact on its real estate market – the COVID-19 pandemic. In our view, it is vitally important to look back on the history of the market and identify how the house market's indicators, in our cases: house prices and mortgage rates, were affected by such crises. In a bigger picture, we aim to derive from our study useful conclusions for both inventory owners and house seekers to stay aware of potential risks on the market and prepare themselves to get the best deals.

In order to achieve the goal, in our study, we set to accomplish the following objectives:

- analyze the chosen time series of house price and mortgage rate based on the common characteristics (autocorrelation, stationarity, trend, seasonality);
- detect and remove trend and seasonality of the times series;
- choose an optimal model to forecast future values;
- formulate hypotheses and testify them based on the results of each part.

The main body of our study includes 4 parts which corresponding majorly to the objectives above: part 1 describes the data in general; part 2 analyses each chosen time series; part 3 decomposes the time series into trend, seasonality, and residuals; and part 4 focuses on building the optimal forecast model.

Part 1: Data description

1. Data description

No.	Object's characteristic	Name of variable	Periodicity and Number of time points	Units of measurement
1	Interest charged on mortgages in the US	mortgage	The data is **taken monthly** (from January to December) in the span of 35 years from 1987 to 2021. Hence, there are 35 * 12 = **420 time points** in the series.	percentage
2	The US house price calculated according to the base period (01.01.2000)	house		index*

* *House price on January 01, 2000 is the base value (100%), the rest are calculated as: (current value / base value) * 100%*

2. Source of dataset:

"U.S. Housing Market Factors" from Kaggle.

URL: https://www.kaggle.com/datasets/faryarmemon/usa-housing-market-factors?select=Monthly_Macroeconomic_Factors.csv

3. Meaning of the chosen time series

On average, Americans who wish to buy a new house would have to borrow a certain amount of money of the bank in order to purchase for it. The banks would charge borrowers a certain amount of money based on the mortgage rates. In simplest terms, because mortgage rates affect people's demand for housing, they would also be related to house prices.

4. Main hypotheses

Hypothesis 1: There is a seasonal pattern for house prices within a period of 12 months, as they tend to decrease during winter months and increase during summer months.

It is usually more convenient for households to move in summer because of the following reasons:

- convenient weather conditions (sunny with longer daylight);
- families with children who are having summer vacation can get settled before a new school year begins in autumn;
- particularly in the US, the holiday period from November to January is a busy time for families to consider moving their house.

Therefore, as demands for houses during summer months go up, their prices go up as well.

Hypothesis 2: During the two acknowledged periods of economic recessions, mortgage rates experienced turbulences with a generally downward trend.

> During difficult economic times, the government often lowers interest rates, to help people purchase more and spur social spendings, including purchase on the real estate market, which leads to lower mortgage rates. In addition, this can also keep the market from freezing under the pressure of recession.

Hypothesis 3: During the two acknowledged periods of economic recessions, house prices had a downward trend.

> During the Great Recession from December 2007 to June 2009 as a consequence of the global financial crisis (GFC), demand for real estate went down; as a rule, house prices decreased as well. Therefore, house prices during the recession caused by the COVID-19 pandemic from January 2020 also experienced a downward trend.

Part 2: Analysis of time series' properties

Part 2.1: The mortgage rate time series

1. General graph

In general, from the first observation, we can see that the time series has a downward trend.

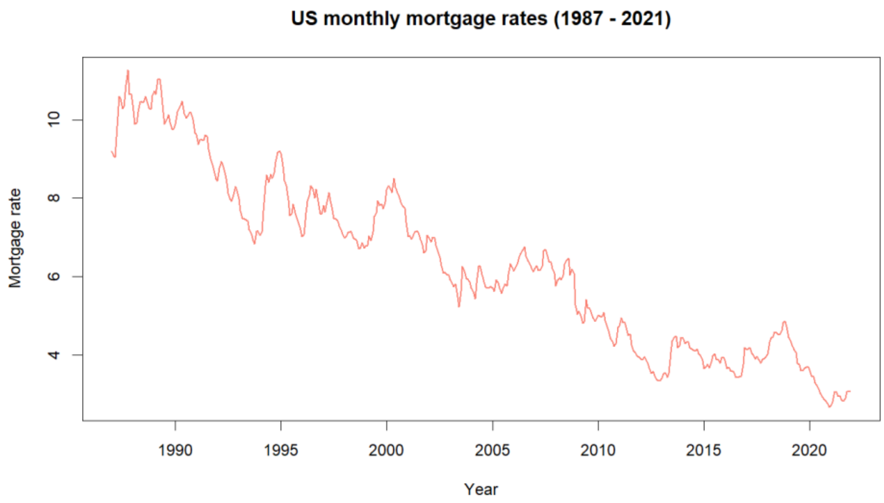

Figure 1: Graph of the US monthly *mortgage rate* changes (1987 – 2021)

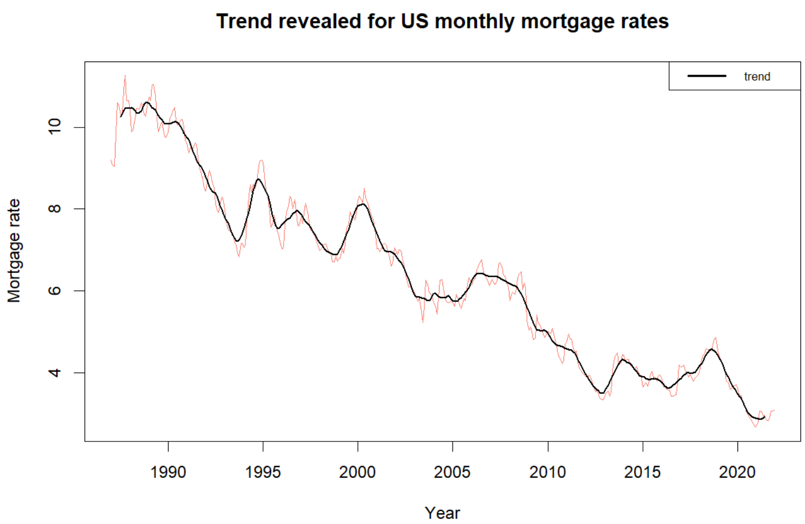

Figure 2: Graph of trend for the US monthly *mortgage rate* changes (1987 – 2021)

Mortgage rate is greatly affected by interest rate, which is one of the instruments of the FED to control the economic growth and fight inflation. It is observable that during 1987 - 1991, mortgage rates almost always stayed over 10%. It can be explained that in 1987, inflation rate in the US rose to 4%, the FED acted by raising fund rates (from 5.8% in October 1986 to 7.2% in October 1987)[1], mortgages also followed. Similar events happened in 1995s and 2000s. At this time, the fast pace of

[1] Jonathan L., 2022. 1987: When mortgage rates last soared this much. *The Orange County Register*. Retrieved from: https://www.ocregister.com/2022/06/18/1987-when-mortgage-rates-last-soared-this-much/

economic growth was followed by high risk of increasing price of goods and services, which led to the rising interest rates by the FED.

After the 2000s, the mortgage rates saw the signs of great fluctuations. Relatively high mortgage rates caused by subprime mortgages[2] (mortgages issued with no down payment required, and proof of income) triggered the crash of the housing market, as many borrowers found their home values lower than the mortgage and defaulted. The decreasing mortgage rate in the 2008 - 2009 was the attempt of the FED to intervene and ease the situation[3].

After the recession, in 2013, the Fed believed the economy eventually will be strong enough to handle a pullback in stimulus and raised the interest rates[4], followed by rises in mortgage rates. In 2018, the FED raised interest rates again to fight inflation[5]. After 2019, rates decreased in response to the Coronavirus pandemic, however, since 2021, as soon as the economy saw the recovery, mortgage rates started to increase.

Hypothesis 2: During the two acknowledged periods of economic recessions, mortgage rates experienced turbulences with a generally downward trend.

We can see that during the global financial crisis 2007 – 2009 and the COVID-19 pandemic in 2020, mortgage rates, even though experienced some turbulences, but generally had a downward trend. We already explained above, in these periods mortgage rates tend to decrease as the policy of the government. Therefore, we have grounds to confirm our 2nd hypothesis.

Hypothesis 2: confirmed.

From this graph we can see no signs of seasonal pattern or cyclic movement. At the same time, it is visible that the average mortgage rates and mortgage rates' variance are not constant but vary over time. Therefore, we can temporarily assume that the data series is non-stationary.

2. Autocorrelation

[2] John D.V. Subprime Mortgage Crisis. *Federal Reserve History*. Retrieved from: https://www.federalreservehistory.org/essays/subprime-mortgage-crisis

[3] Robert R. The Great Recession. *Federal Reserve History*. Retrieved from: https://www.federalreservehistory.org/essays/great-recession-of-200709

[4] Ylan Q.M., 2013. Interest rate spike emerges as concern for Fed. *Washington Post*. Retrieved from: https://www.washingtonpost.com/business/economy/interest-rate-spike-emerges-as-concern-for-fed/2013/06/26/56c521e0-ddac-11e2-948c-d644453cf169_story.html

[5] Jeff C., 2022. Fed hikes interest rates by 0.75 percentage point for second consecutive time to fight inflation. *CNBC*. Retrieved from: https://www.cnbc.com/2022/07/27/fed-decision-july-2022-.html

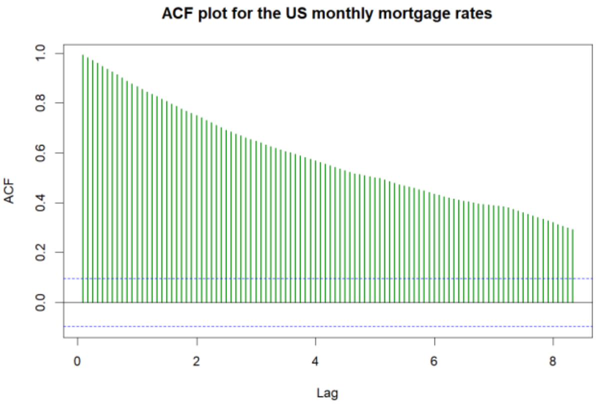

Figure 3: ACF plot for *mortgage rates*

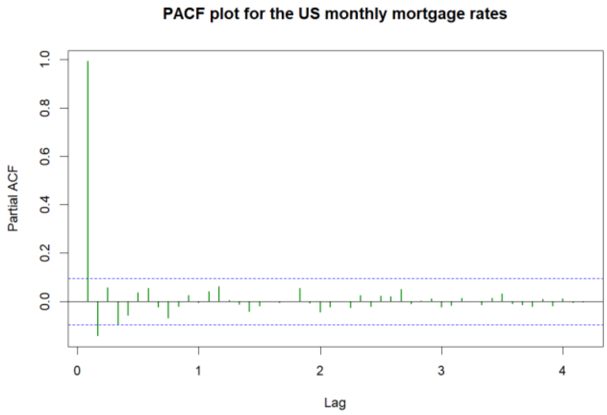

Figure 4: PACF plot for *mortgage rates*

The autocorrelation graph decreases gradually with all spikes above the zero bound of significance, which shows that house prices have high autocorrelation with the previous time period and therefore, it is not white noise. The PACF graph indicates that we should take into consideration only those first 2 lags which have spikes beyond the significance line, because after lag 2 the PACF tapers to 0.

In general, there exists a trend as the ACF graph decreases gradually and there is no sign of seasonality as there is no visible fluctuation in any period of time.

3. Periodogram

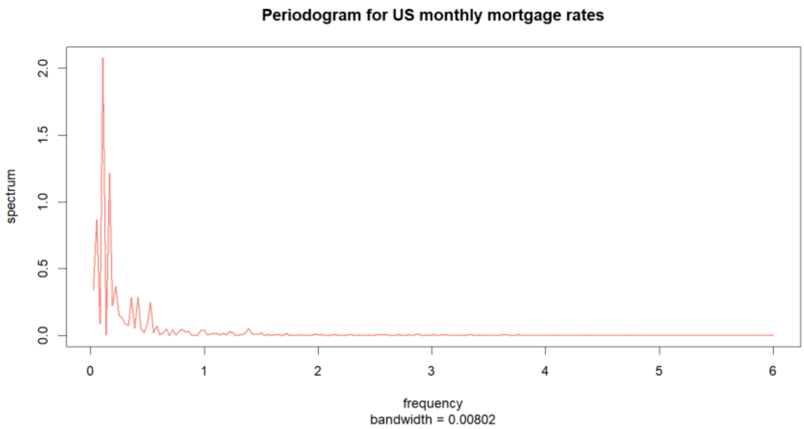

Figure 5: Periodogram for *mortgage rates*

From the periodogram, it is observable that the dominant peak is near the zero frequency. The graph shows some high waves around the peak, which indicates that there is not a clear trend. The unclear trend is also reflected the fluctuation of time series from the general graph. In addition, the line becomes almost completely flat after intensive waves near 0 with extremely small waves.

4. Augmented Dickey-Fuller (ADF) test

H_0: Time series is non-stationary.

H_1: Time series is stationary.

Result:

Dickey-Fuller statistic:	-3.5918
p-value:	0.03379

Thus, we reject the null hypothesis, which means the time series of mortgage rates is stationary.

Stationarity in the time series of mortgage rates means that the average mortgage rate as well as the spread of mortgage rates do not change over time. However, because we can still clearly observe the downward trend, we consider this is because mortgage rates have small values from 1% to 13%, and therefore, ADF test considers the changes of mortgage rates insignificant.

5. Conclusion

Mortgage rates are highly autocorrelated by month, which means that the mortgage rate of a month is influenced by the mortgage rate(s) of the previous month(s).

In general, mortgage rates in the US have a downward trend from the beginning till the end of the period. Because house prices tend to increase over time, mortgage rates were offered lower to encourage people to buy houses. During tough economic times, mortgage rates tend to decrease even further to incentivize people to borrow and spend more. However, right after the economy started to pick up, such as the post-pandemic period, mortgage rates showed signs of increasing again.

Part 2.2: The house price time series

1. General graph

In general, from the first observation, we can see that the time series has an upward trend.

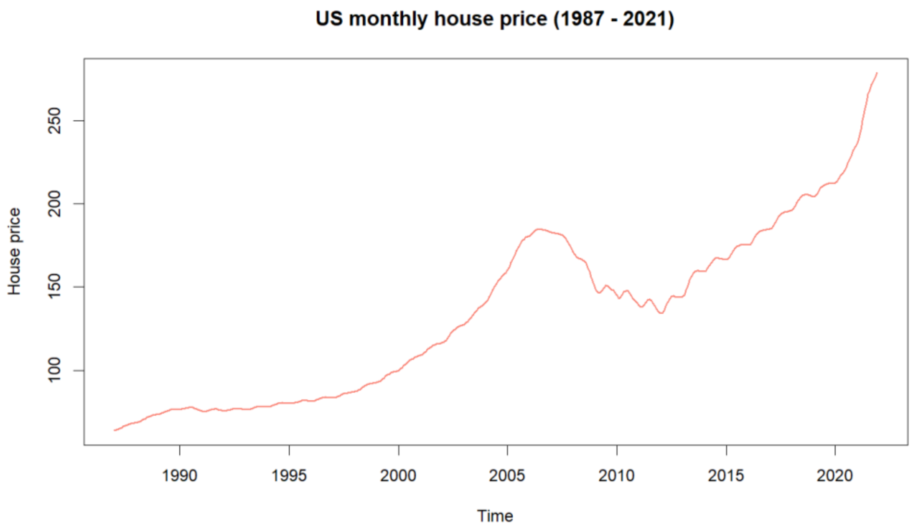

Figure 6: Graph of the US monthly *house price* changes (1987 – 2021)

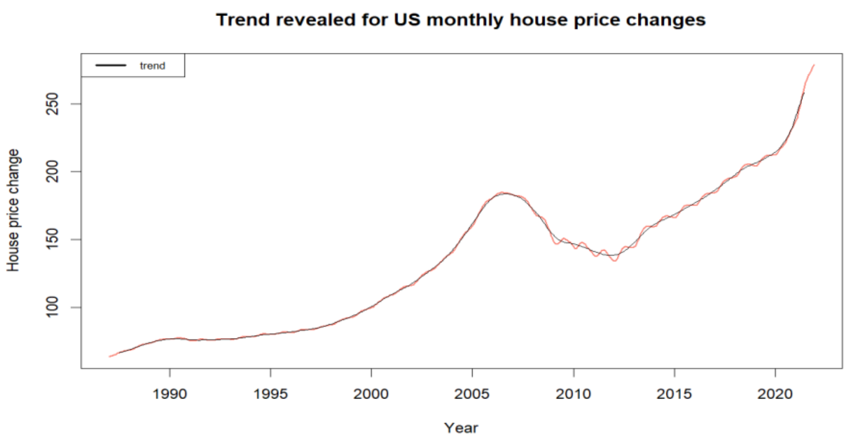

Figure 7: Graph of trend for the US monthly *house price* changes (1987 – 2021)

From 1987 to 2000, house prices rose gradually. From 2000 they continued to grow but at a faster pace until they reached a peak in 2006. Because of the global financial crisis during 2007 – 2009, house prices began to decline as an effect of recession and only returned to the previous positive trend about 4 years after the crisis. From 2009 to 2020, the upward trend in house prices in the US was not as “smooth” (with more fluctuations) as it was before. It can also be seen that from 2020, house prices increased faster than its fluctuating trend, which can be explained by the inflation caused by the COVID-19 pandemic.

From this graph we can see no signs of seasonal pattern or cyclic movement, which is a sign for us to reject our first hypothesis about potential seasonality of house prices.

At the same time, it is visible that the average house price and house price's variance were not constant but varied over time. Therefore, we can temporarily assume that the time series is not stationary.

Moreover, it is not difficult to notice that house price behaviors are different during the two acknowledged recession periods. Whereas the house prices dropped in the 2007 – 2009 recession, in the COVID-19 recession, they actually went up, but not down as we expected while formulating our 3rd hypothesis. Therefore, we have grounds to reject the 3rd hypothesis.

Hypothesis 3: During the two acknowledged periods of economic recessions, house prices had a downward trend.

While looking for an answer to our false assumptions, we managed to find three main reasons:

- First, mortgage rates in the COVID-19 recession were lower than those in the Great Recession (3% compared to 6%), which helped increased demand on the real estate market more.
- Second, although lowered mortgage rates in 2008 also contributed to an increase in demand, they also boosted the number of new houses that were built for sale. As new house listings became abundant (supply exceeded real demand), house prices started to drop quickly in 2008. In contrast, in the COVID-19 period, there was high demand for house and lower supply on the market, which led to prices climbing up.
- Finally, because of the experience from the housing bubble in 2008, the real estate market has been more strictly regulated by the US government to avoid another crash[6].

Hypothesis 3: Rejected.

2. Autocorrelation

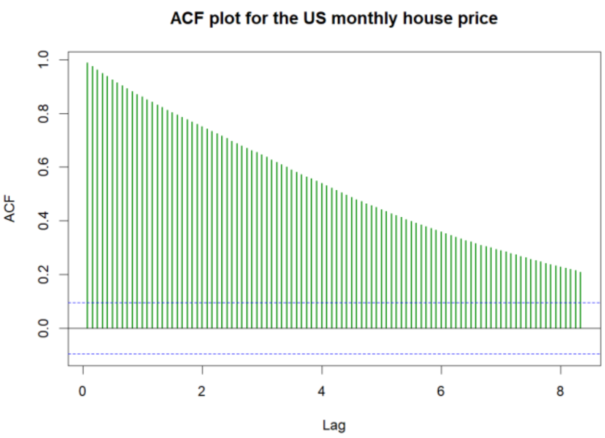

Figure 8: ACF plot for *house price*

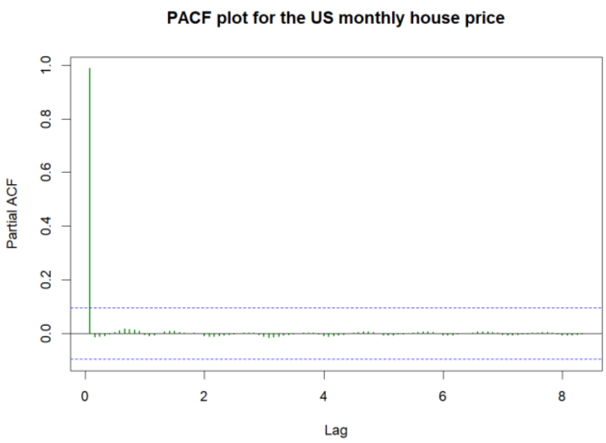

Figure 9: PACF plot for *house price*

The autocorrelation graph decreases gradually with all spikes above the zero bound of significance, which shows that house prices have high autocorrelation with the previous period and

[6] Zumwalt, W. (November 15, 2021). **"The Difference in the 2008 and 2021 Housing Bubble".** https://storymaps.arcgis.com/stories/ca5fe58da85240a19c4cfc144cd28552

therefore, it is not white noise. The partial autocorrelation graph indicates that we should take into consideration only one lag which has its spike higher than the significance line.

In general, there exists a trend as the ACF graph decreases gradually and there is no sign of seasonality as there is no visible fluctuation in any period of time.

3. Periodogram

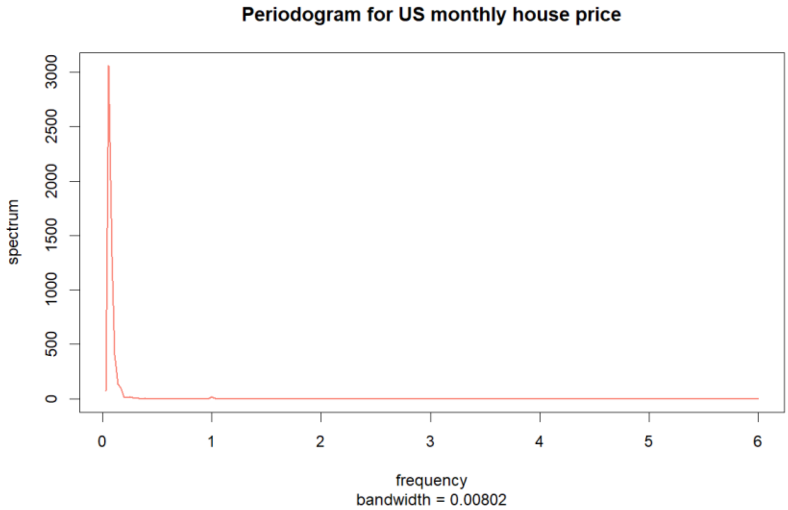

Figure 10: Periodogram for *house price*

From the periodogram, we can see that there is only one peak near the zero frequency. This indicates that there is a significant change at one wave, hence, there is a general trend. At the same time, it is possible to assume the non-stationary characteristic of our time series. As the line becomes almost completely flat after a very low peak when frequency equals 1, the graph shows that there are no small waves (almost flat).

4. Augmented Dickey-Fuller (ADF) test

H_0: Time series is non-stationary.

H_1: Time series is stationary.

Result:

Dickey-Fuller statistics: -0.94345

p-value: 0.9473

We fail to reject the null hypothesis, which means the time series of house prices is non-stationary.

Non-stationarity in the time series of house prices means that the average house price as well as the spread of house prices change over time. A possible explanation for this is that as the US economy grew over time, its citizens became more capable of buying houses. As a rule, increasing demands on the real estate market led to the increase in house prices.

5. Conclusion

House prices are highly autocorrelated by month, which means that the house price of a month is influenced by the house price(s) of the previous month(s).

In general, house prices in the US have an upward trend from the beginning till the end of the period. However, global financial crisis during 2007 – 2009 was the determinant factor that, for the first time, marked the rapid decline in contrast with the steady growth of the previous period. Moreover, after this global recession, the patterns of growth in house prices are not as smooth as before and in fact, the global recession completely changed the dynamic of the US house prices in the next decade.

Part 3: Decomposition of time series

Part 3.1: The mortgage rate time series

Trend and Seasonality

As shown in part 2, our mortgage rate time series does not have a seasonal characteristic. Therefore, we could skip the process of de-seasonalizing (using harmonic regression) and carried out polynomial regression immediately. The properties of the regression are shown in the table below:

```
Residuals:
     Min       1Q   Median       3Q      Max 
-1.77046 -0.37629 -0.03477  0.39445  1.51521 

Coefficients:
                                  Estimate Std. Error t value Pr(>|t|)    
(Intercept)                      9.955e+03  1.311e+03   7.592 2.09e-13 ***
poly(index(Y_m), 2, raw = TRUE)1 -9.719e+00  1.308e+00  -7.427 6.31e-13 ***
poly(index(Y_m), 2, raw = TRUE)2  2.372e-03  3.264e-04   7.268 1.81e-12 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.6107 on 417 degrees of freedom
Multiple R-squared:  0.9238,	Adjusted R-squared:  0.9234 
F-statistic:  2526 on 2 and 417 DF,  p-value: < 2.2e-16
```

Table 1: Polynomial regression

Regression equation:

$$y = 9.955e{+}03 + 2.372e{-}03 * index + - 9.9719e{+}00 * index^2$$

p-value of the regression is 2.2e-16 < 0.05, we can conclude that the regression is significant. p-value of the coefficients are 2.09e-13, 6.31e-13, 1.81e-12 < 0.05, we can conclude that all coefficients are significant.

Then we withdrew residuals from polynomial regression and received the following figure.

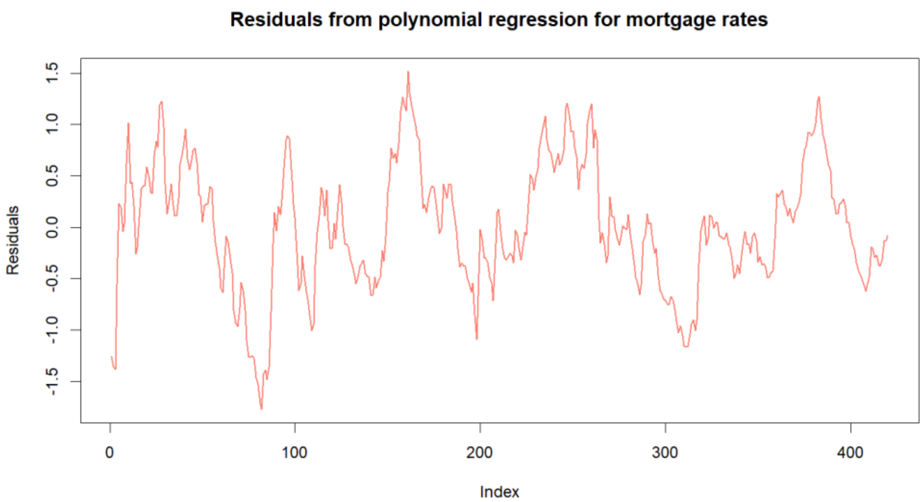

Figure 11: Residuals from polynomial regression for *mortgage rates*

In comparison with figure 1 and figure 2, this figure does not show any downtrend in the time series.

Augmented Dickey-Fuller (ADF) test

H_0: Time series is non-stationary.

H_1: Time series is stationary.

Result:

Dickey-Fuller statistic:	-3.9163
p-value:	0.01335

We performed the ADF test for the polynomial regression's residuals and the received result confirms stationarity.

Finally, we saved the data for polynomial regression residuals to proceed to the next stage.

Part 3.2: The house price time series

1. Trend

In order to remove the previously observed uptrend, we used a technique called differencing to stabilize the mean.

However, after **differencing for the first time**, our plot revealed the presence of seasonality that couldn't be observed in the previous analysis.

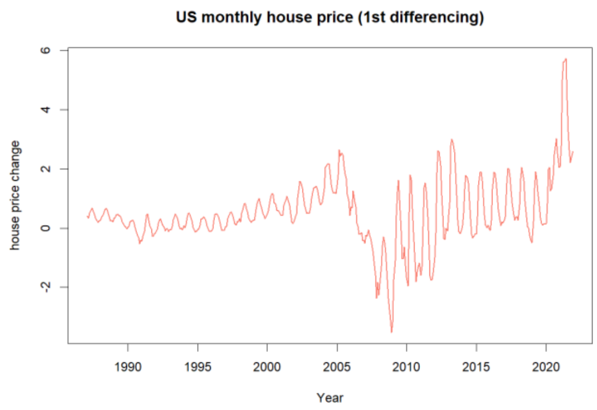

Figure 12: Graph of the US monthly house price (1987 – 2021) at the 1st difference

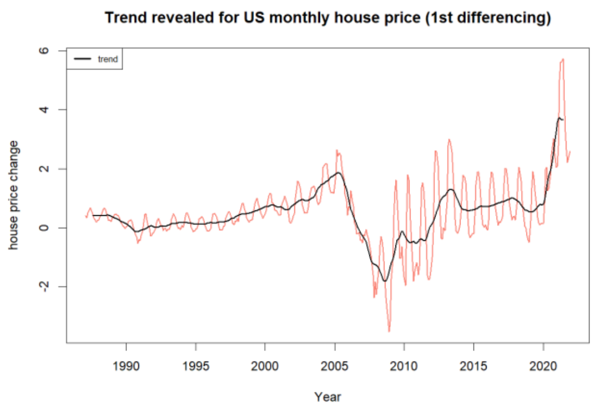

Figure 13: Graph of trend for the US monthly house price (1987 – 2021) at the 1st difference

As can be seen, the values for house price are now closer to 0. Similarly to our previous analysis, the global financial crisis of 2007 – 2009 marked a substantial change in the way house prices fluctuated. During this period, house prices fluctuated not in any repeated pattern, but after the crisis house prices started to oscillate more broadly than they did before 2006.

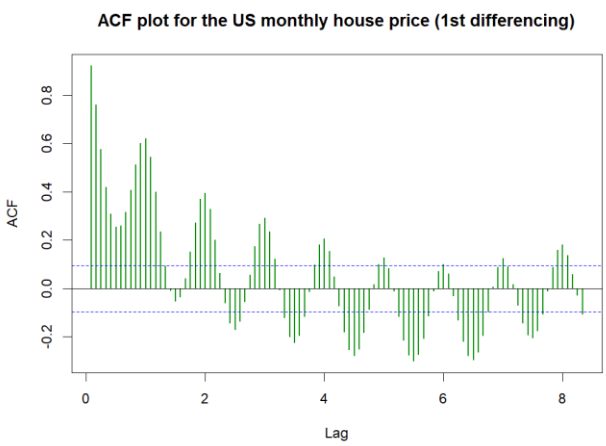

Figure 14: ACF plot for *house price* (1st differencing)

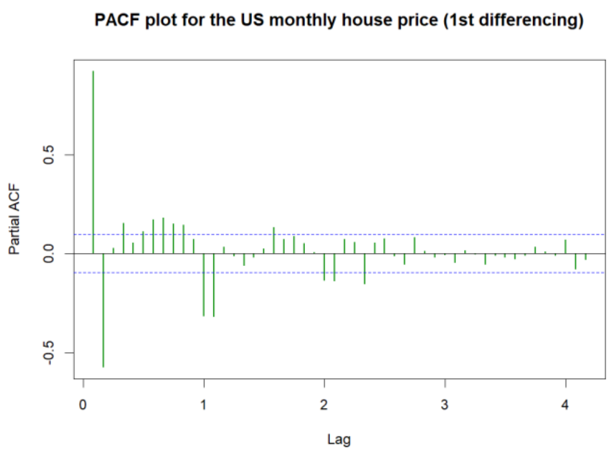

Figure 15: PACF plot for *house price* (1st differencing)

The ACF graph on the left indicates that the waves follow a repeated pattern after every 12 lags, which means that there exists a seasonal variation within a year (12 months). The PACF plot on the right also shows a few significant lags before a sharp cut-off after lag 12, which also proves annual seasonality.

In the preliminary analysis, we rejected the 1st hypothesis on the presence of seasonality in house prices. However, thanks to what the ACF graph has revealed, we have grounds to confirm this hypothesis after all.

We continued to apply **differencing for the second time**. This time we can be certain that we have successfully removed any remaining signs of a trend, and what's left is a clearer visualization of a seasonal time series.

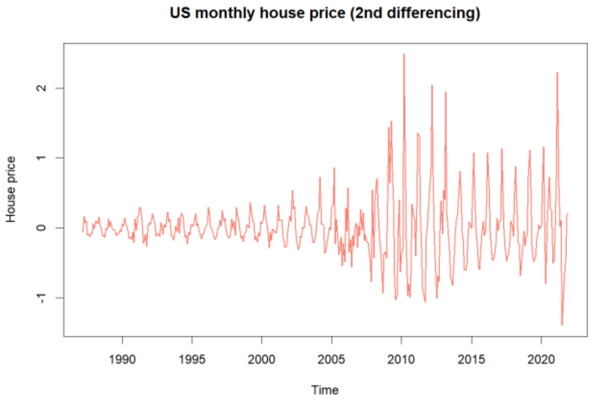

Figure 16: Graph of the US monthly house price (1987 – 2021) at the 2nd difference

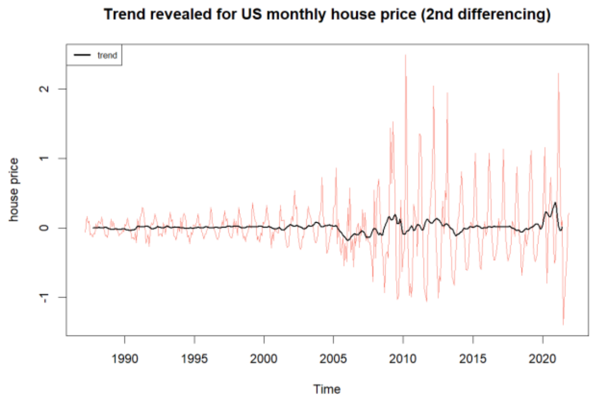

Figure 17: Graph of trend for the US monthly house price (1987 – 2021) at the 2nd difference

The largest waves are in the years from 2009 to 2013, as the market was affected by the 2007 – 2009 recession and by the government's measures to stabilize the economy as whole. The difference in the range of fluctuation between the post-recession period and pre-recession period is quite obvious. Apart from the fact that the trend has now been removed, visually, we also noted that our time series has an additive seasonality, because the magnitude of oscillation does not increase over time, but became larger in a sudden way due to the 2007 – 2009 GFC. From this period onwards, the range of fluctuations became narrower but still wider than that of the pre-GFC period. In the three periods (pre-GFC, during GFC, post-GFC), the magnitude of fluctuations are stable and did not amplify over time. Therefore, we reinforced the statement in part 2's analysis that the global recession in 2007 – 2009 had changed the real estate market forever.

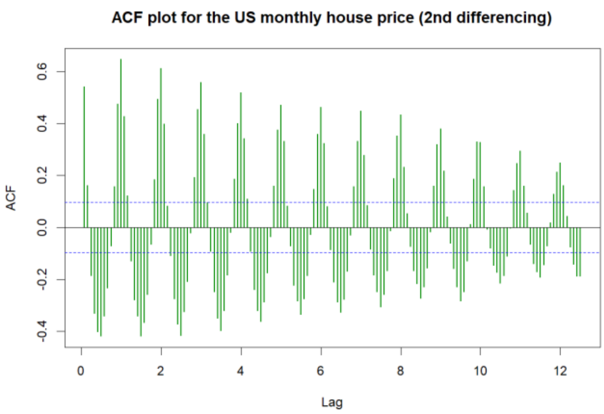

Figure 18: ACF plot for *house price* (2nd differencing)

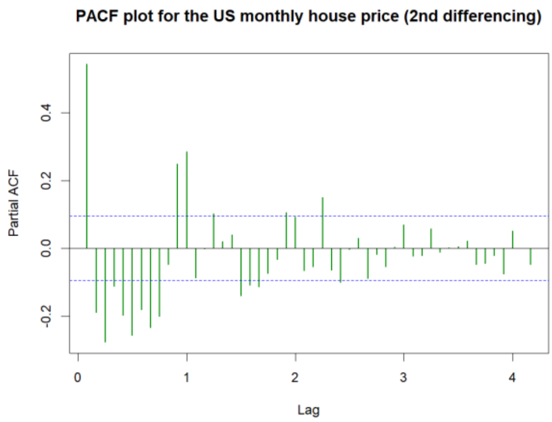

Figure 19: Plot for *house price* (2nd differencing)

As seasonality is revealed more clearly, it is observable that house prices changed its pattern every 6 month as the spikes in the ACF plot reach their peaks after every 12 lags and return to roughly the same level after every 12 lags. In addition, the ACF is slowly decaying, which means that future values of house price are heavily correlated with their past values. In the PACF plot, almost all of the first 12 lags are significant, and lag 12 is followed by a drop in PACF values and an approach to zero towards the end.

Hypothesis 1: There is a seasonal pattern with house prices within a year, as they tend to decrease during winter months and increase during summer months.

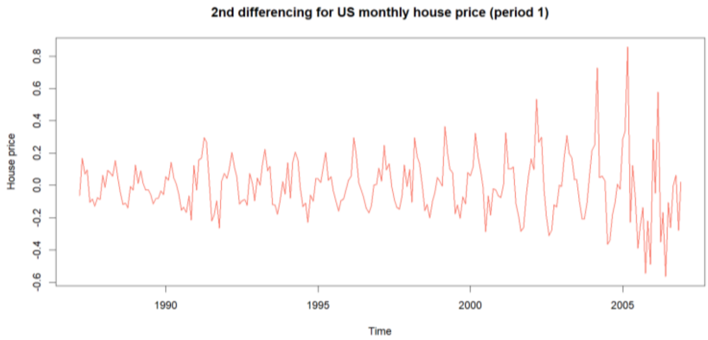

Figure 20: General plot for *house price* of the pre-GFC period (2nd differencing)

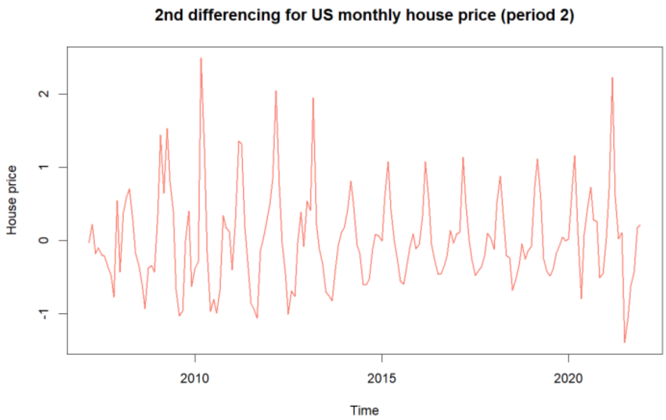

Figure 21: General plot for *house price* of the post-GFC period (2nd differencing)

Indeed, we split the data into two periods to observe more closely. Period 1 (before the global financial recession) is from 1987 to 2006 and period 2 (including the global financial recession and the COVID-19 recession) is from 2007 to 2021. For both periods, we observed that although seasonality does exist, there is no evidence to confirm that house prices go up during summer months and down during winter months. In fact, there are many periods where house prices peaked in winter, then gradually decreased until they reached bottom in summer, and started climbing up again till winter. Therefore, we rejected the 1st hypothesis.

Hypothesis 1: Rejected.

2. Seasonality

From this step onwards, we decided to apply two different approaches to our research on house price. For the first approach, we carried out spectral analysis with harmonic regression to test our assumptions of seasonality. For the second approach, we applied the 3rd round of differencing to completely eliminate any signs of seasonality.

Approach 1: Spectral analysis

Now that the upward trend has been removed from our series, we start the process of eliminating seasonality by carrying out spectral analysis, in which we apply harmonic regression.

> ***Note**: The reason why we did not perform regression on cycles for the original data, but for the data that had been differenced for two times, is because without differencing, our data did not reveal signs of seasonality in either the general graph or the ACF plots.*

Periodogram

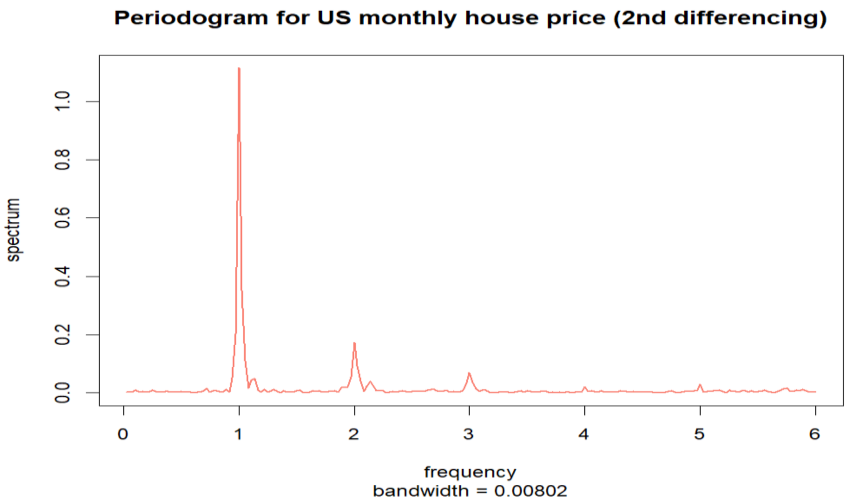

Figure 22: Periodogram for *house price* (2nd differencing)

From the graph above, we can see the evident peaks of the spectrograms which indicate potential seasonality within periods of 12 months, 6 months, or 4 months. After these peaks, the waves are extremely small with mini peaks at the 4 and 5 frequencies, so we did not take them into account.

Harmonic regression

After having identified the harmonics, we built a regression model for them using the least squares method.

```
Residuals:
     Min       1Q   Median       3Q      Max
-1.12625 -0.20550  0.02141  0.15762  1.91137

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.006100   0.016515   0.369    0.712
omega1c1     0.346100   0.023399  14.791   < 2e-16 ***
omega1s1     0.121977   0.023314   5.232 2.68e-07 ***
omega2c2    -0.010351   0.023370  -0.443    0.658
omega2s2     0.144289   0.023342   6.182 1.53e-09 ***
omega3c3     0.007573   0.023356   0.324    0.746
omega3s3     0.092030   0.023356   3.940 9.56e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3376 on 411 degrees of freedom
Multiple R-squared:  0.423,     Adjusted R-squared:  0.4146
F-statistic: 50.21 on 6 and 411 DF,  p-value: < 2.2e-16
```

Table 2: Result of harmonic regression for *house price* after the 2nd differencing

The table above shows that the one-year frequency is the most significant with the p-values of both sine and cosine are below the significant level. The other frequencies of 6 months and 4 months are also significant, and although not always, it also proves that these harmonics are present. Now we check the residuals to see if this model helped us improve the existing peaks.

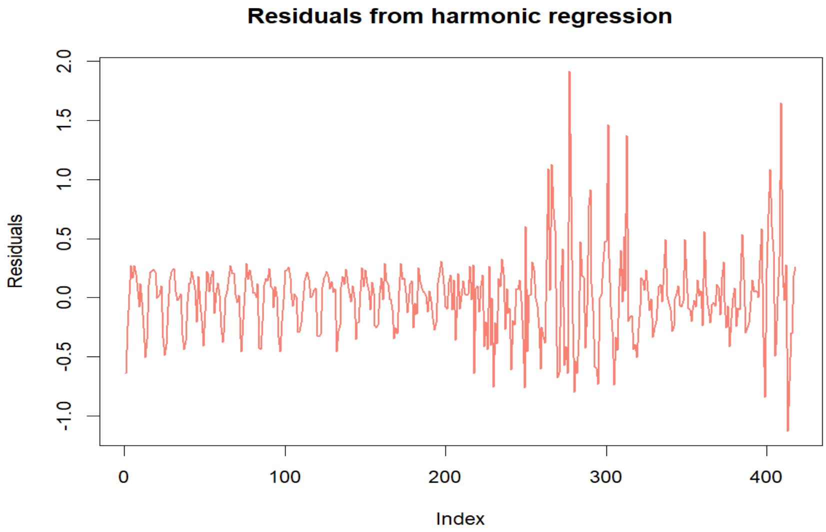

Figure 23: Residuals from harmonic regression after the 2nd differencing

Compared with the peaks in fig. 16, the peaks are now closer to 0, which is a good sign.

We also tried to build regression on polynomials for residuals. However, the result is revealed to be insignificant with p-value = 0.48 (> 0.05). At the same time, the coefficients are also insignificant.

```
Residuals:
     Min       1Q   Median       3Q      Max
-1.14572 -0.20796  0.02343  0.16121  1.90491

Coefficients:
                                 Estimate Std. Error t value Pr(>|t|)
(Intercept)                    -2.004e-02  3.287e-02  -0.610    0.542
poly(index(res_h2), 1, raw = TRUE)  9.564e-05  1.360e-04   0.703    0.482

Residual standard error: 0.3354 on 416 degrees of freedom
Multiple R-squared:  0.001188,	Adjusted R-squared:  -0.001213
F-statistic: 0.4949 on 1 and 416 DF,  p-value: 0.4822
```

Table 3: Polynomial regression for residuals from harmonic regression of *house price*

Therefore, for further analysis, we decided not to use the residuals of polynomial regression, but the previous residuals of harmonic regression.

Augmented Dickey-Fuller (ADF) test

We performed the ADF test for the harmonic regression's residuals and the received result confirms stationarity.

H_0: Time series is non-stationary.

H_1: Time series is stationary.

Result:

Dickey-Fuller statistic: -12.005

p-value: 0.01

Finally, we save the data for regression residuals (after having removed trend by differencing and detected seasonality) to proceed to the next stage.

Approach 2: Third-round differencing

We applied **differencing for the third time**. This time we have managed to clear the seasonal characteristics of our time series.

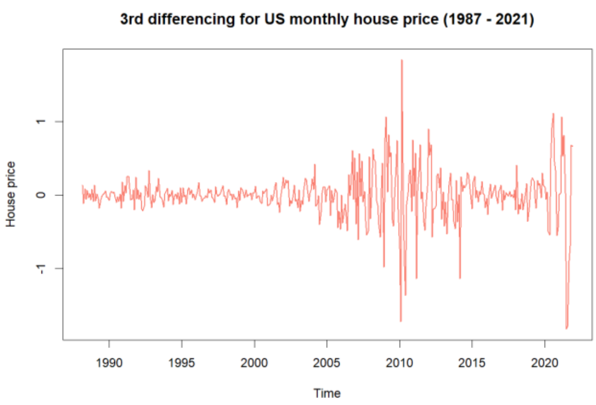

Figure 24: Graph of the US monthly *house price* changes after the 3rd differencing (1987 – 2021)

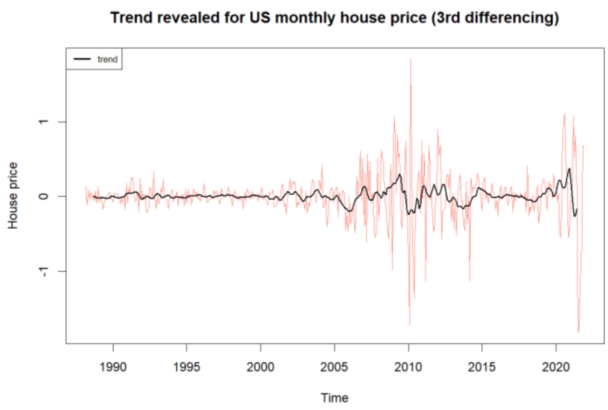

Figure 25: Graph of trend for the US monthly *house price* changes after the 3rd differencing (1987 – 2021)

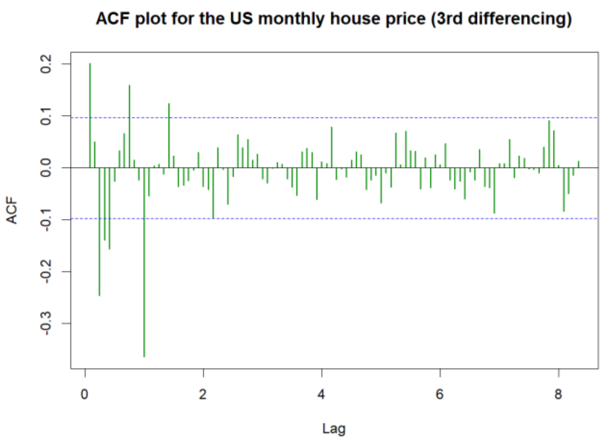

Figure 26: ACF plot for *house price* (3rd differencing)

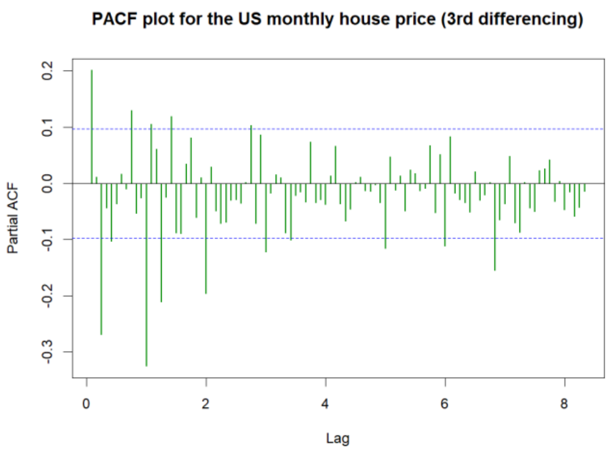

Figure 27: PACF plot for *house price* (3rd differencing)

As can be seen from the ACF plot on the left, there is no repeated pattern a specific number of lags and the graph on the right shows a de-trended and de-seasonalized graph of house prices. Not surprisingly, there are significant turbulences during the global recession and a few years later. Finally, the PACF graph informs us of the significant lags whose spikes are above threshold. The spikes alternate between positive and negative correlations, which indicates a higher order moving average term. We'll proceed to more details when we use this graph to choose the suitable predicting model.

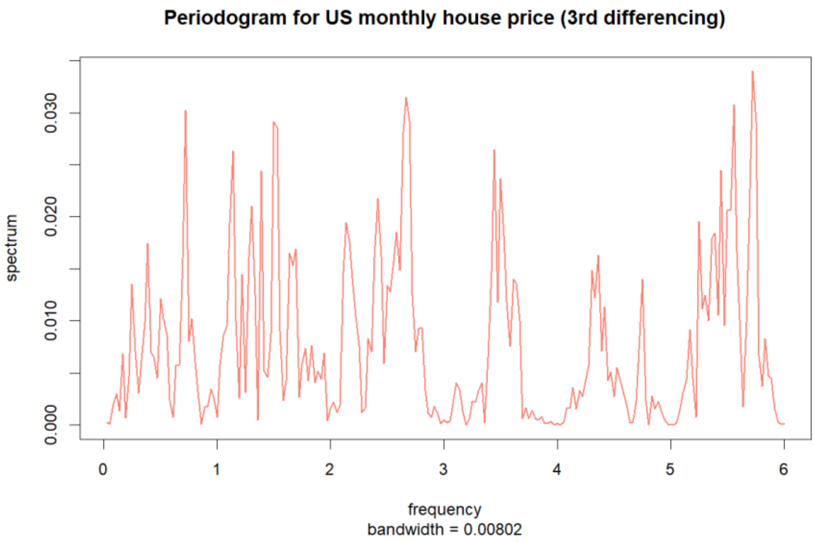

Figure 28: Periodogram for the US monthly *house price* at the 3rd difference

The periodogram for house price at the 3rd difference has no dominant spikes or distinct waves to signify seasonality or trend.

Augmented Dickey-Fuller (ADF) test

H_0: Time series is non-stationary.

H_1: Time series is stationary.

Result:

Dickey-Fuller statistic:	-7.6923
p-value:	0.01

We performed the ADF test for the harmonic regression's residuals and the received result confirms stationarity.

Finally, we saved the three-times differenced data of *house price* for further analysis.

3. Conclusion

Our data of house price appeared to have an upward trend, but showed no signs of seasonality until we applied differencing for the second time.

At the first approach, after carrying out the regression for harmonics, we derived the regression residuals. The ACF plot revealed annual seasonality (and surprising, stationarity), because of which we attempt to use **the seasonal ARIMA model** for these residuals in the next part of the research.

At the second approach, we applied the 3rd round of differencing, and received a series cleaned out of trend and seasonality, hence stationarity. We saved the three-times-differenced series and attempt to use **the ARIMA model** for it.

Part 4: Autonomous dynamic time series models

Part 4.1: Dynamic model for mortgage rates

We split our data into two parts: the part being trained includes values from January 1987 to December 2020, and the part being tested includes values from January 2021 to December 2021, which also means we will try to predict mortgage rates in the US for 12 successive months.

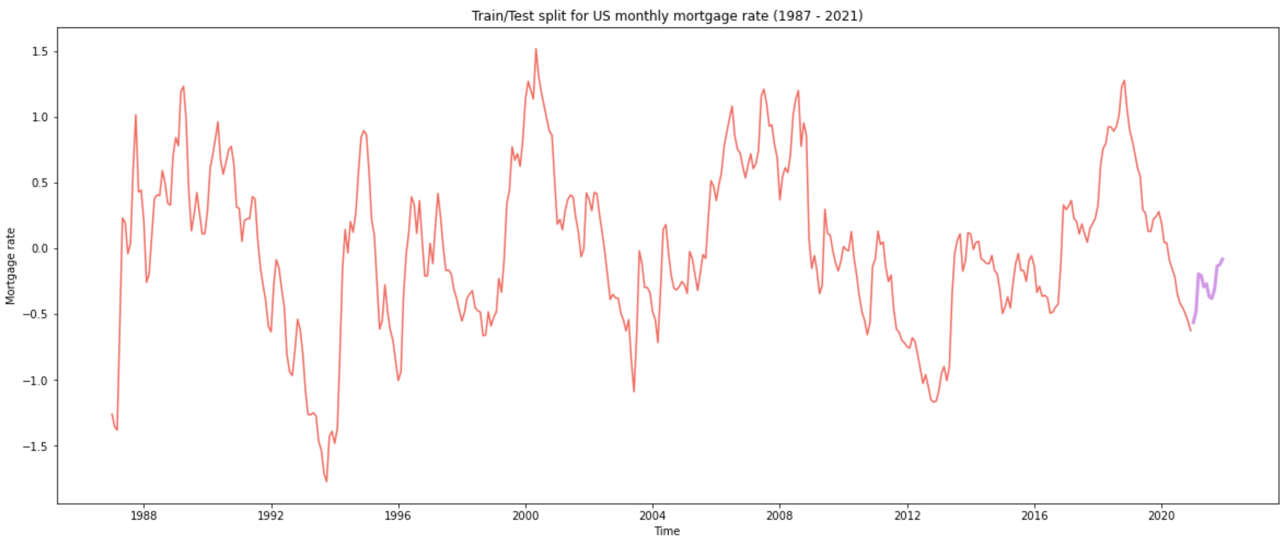

Figure 29: Division of train set and test set (of 12 months) for *mortgage rates*

The following steps are based on the Box-Jenkins methodology.

Step 1: Identification: Stationarity & Defining parameters of the model.

The 2 ADF tests from the previous part showed that our data, as well as harmonic regression's residuals, is stationary.

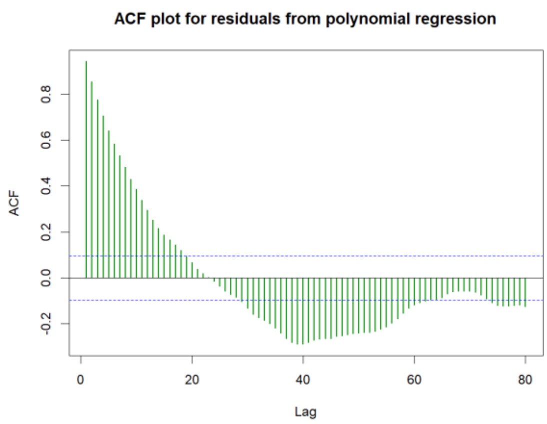

Figure 30: ACF plot for residuals from polynomial regression for *mortgage rates*

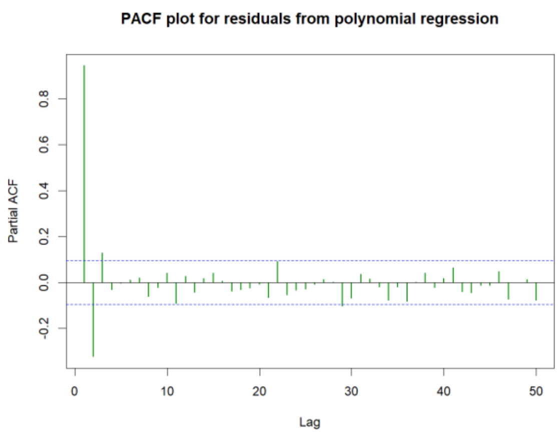

Figure 31: PACF plot for residuals from polynomial regression for *mortgage rates*

From the PACF plot, however we can see that partial autocorrelation dies down after lag 3. We also took into account lag 2, because it is far more significant than lag 3. The ACF plot trails off from positive to negative, which indicates that there still exists trend in our data.

Therefore, we decided to include a first differencing for residuals from polynomial regression for mortgage rates.

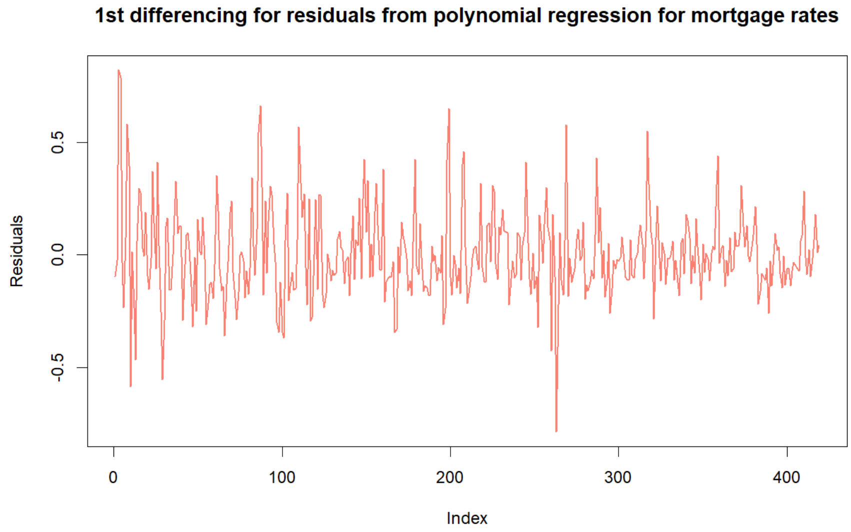

Figure 32: General plot for residuals from polynomial regression for *mortgage rates* (1^{st} differencing)

After applying the method of differencing, our data lost 1 value, which equal to the value of mortgage rate recorded in December 2021. We consider it a fair loss in exchange for a detrended dataset.

Hence, we split our data into two parts: the part being trained includes values from January 1987 to November 2021, and the part being tested includes values from December 2021 to November 2021, which also means we will try to predict mortgage rates in the US for 12 successive months.

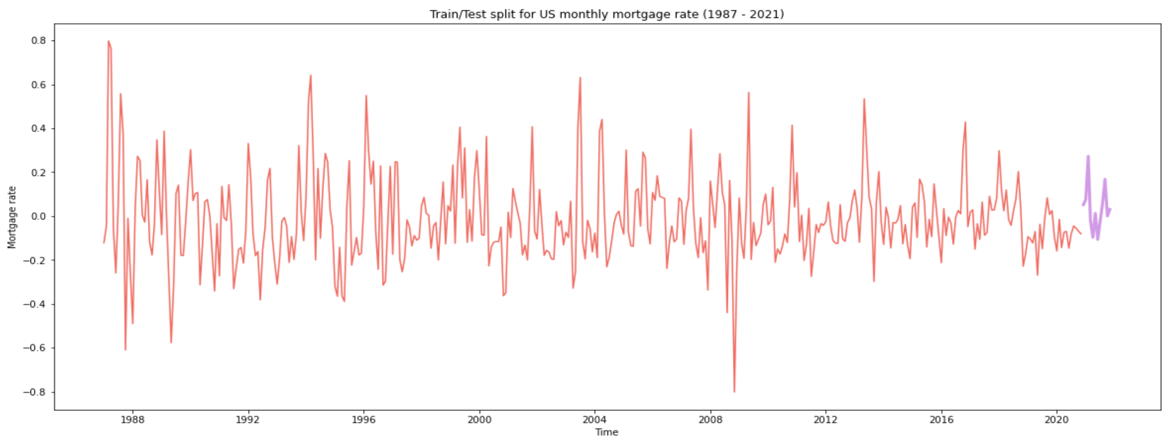

Figure 33: Division of train set and test set (of 12 months) for *mortgage rates* (after 1^{st} differencing)

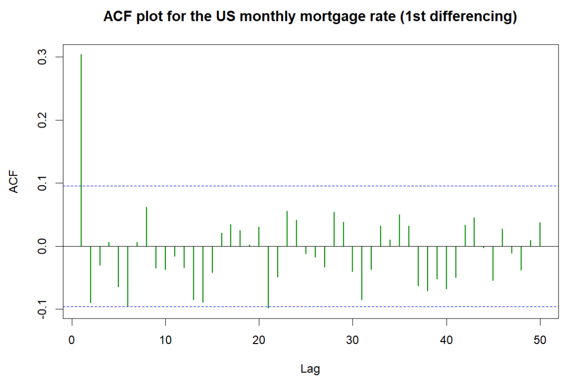

Figure 34: ACF plot for *mortgage rates* (1^{st} differencing)

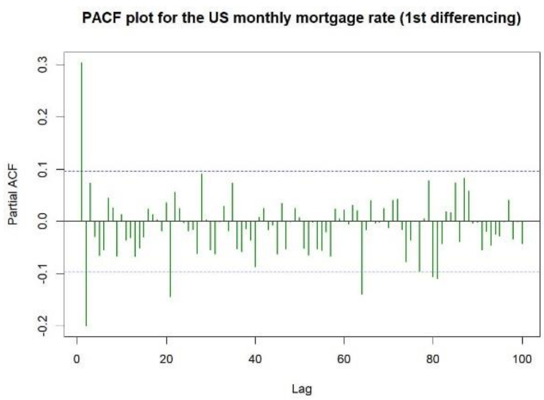

Figure 35: PACF plot for *mortgage rates* (1^{st} differencing)

Both the ACF and PACF plots show a decaying characteristic. From the ACF plot, we can see that lag 1 and 21 are significant. From the PACF plot, we should take into consideration lags 1, 2, 21, 64, 80 and 81, which have spikes beyond the significant level.

Therefore, from these plots we decided to try the ARMA (p,q) model with p = 1, p = 2, p = 21, p = 64, p = 80 or p = 81 and q = 1 or q = 21.

Step 02: Estimation

Apart from our own speculations derived from the ACF and PACF plots, we used the auto.arima function in the R program. It suggested an MA (1) model, which includes a first-order of MA component.

In order to optimize the suggested model, we withdrew the residuals and checked the ACF and PACF plots to adjust a few parameters and delivered a comparison between the potential models at the end.

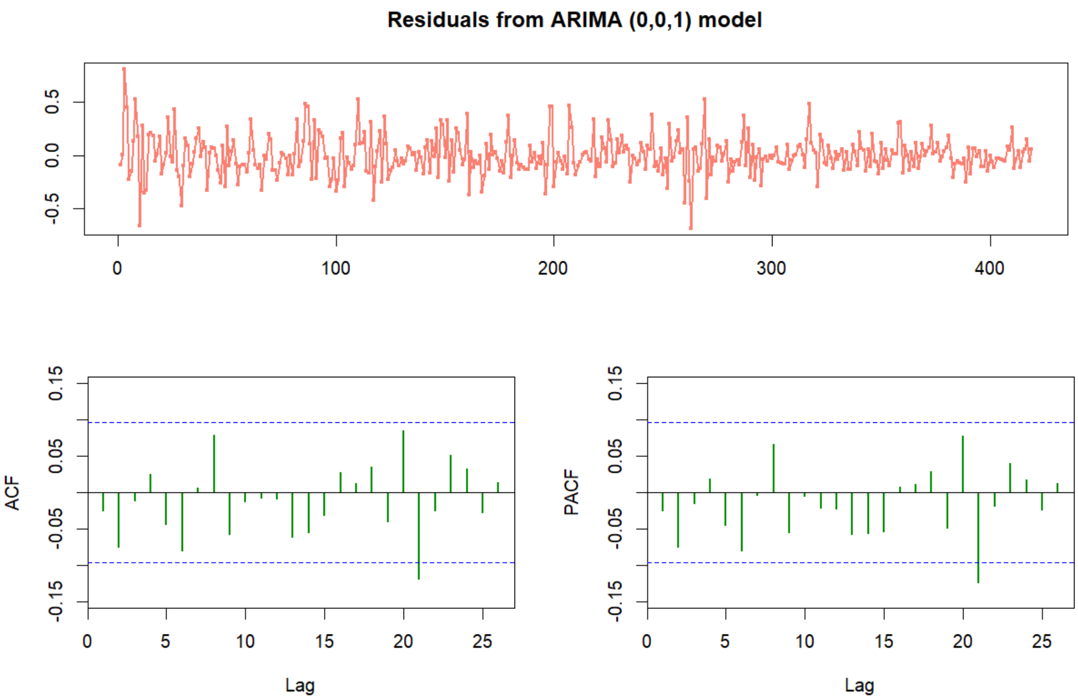

Figure 36: Residuals from the MA (1) model for *mortgage rates*

Both the ACF and PACF plots show a sharp cut-off after lag 21. In the ACF plot, there is a sharp cut-off after the significant spike of lag 21. In the PACF plot, only lag 21 is significant. However, we already took into account these lags in the previous observation. Hence, we adjusted the parameters, ran the models on Python and received the following results:

Model	AIC	BIC
ARIMA (0, 0, 1)	-114.57	302.35
ARIMA (1, 0, 1)	-115.07	292.67
ARIMA (1, 0, 21)	-121.01	281.89
ARIMA (2, 0, 1)	-141.26	191.47
ARIMA (2, 0, 21)	-133.20	215.57
ARIMA (21, 0, 1)	-154.65	113.94

ARIMA (21, 0, 21)	-181.60	-5.21
ARIMA (64, 0, 1)	-198.52	-102.31
ARIMA (64, 0, 21)	-201.28	-105.07
ARIMA (80, 0, 1)	**-215.29**	**-203.27**
ARIMA (80, 0, 21)	-214.67	-198.64
ARIMA (81, 0, 1)	-214.19	-194.15
ARIMA (81, 0, 21)	-207.00	-106.78

Table 4: Comparison of models for *mortgage rates* based on AIC and BIC criteria

We took into account both the AIC and BIC criteria, and therefore decided to choose the ARMA (80, 1) model as the optimal model for further analysis. This model is characterized by an eightieth AR order and a first MA order.

	coef	std err	z	P>|z|	[0.025	0.975]
const	-0.0168	0.004	-4.210	0.000	-0.025	-0.009
ar.L1	0.6306	0.419	1.506	0.132	-0.190	1.452
ar.L2	-0.3234	0.157	-2.066	0.039	-0.630	-0.017
ar.L3	0.1090	0.119	0.917	0.359	-0.124	0.342
ar.L4	-0.0300	0.078	-0.382	0.702	-0.184	0.124
ar.L5	-0.0332	0.081	-0.410	0.682	-0.192	0.126
ar.L6	-0.0928	0.079	-1.178	0.239	-0.247	0.062
ar.L7	0.0258	0.084	0.307	0.759	-0.139	0.191
ar.L8	0.0778	0.074	1.046	0.296	-0.068	0.223
ar.L9	-0.1345	0.079	-1.703	0.089	-0.289	0.020
ar.L10	0.0021	0.089	0.024	0.981	-0.172	0.176
ar.L11	-0.0171	0.082	-0.208	0.835	-0.178	0.144
ar.L12	-0.0323	0.088	-0.368	0.713	-0.204	0.140
ar.L13	-0.0736	0.099	-0.747	0.455	-0.267	0.120
ar.L14	0.0287	0.086	0.334	0.739	-0.140	0.198
ar.L15	-0.0667	0.091	-0.732	0.464	-0.246	0.112
ar.L16	-0.0144	0.080	-0.180	0.858	-0.172	0.143
ar.L17	0.0092	0.073	0.126	0.899	-0.133	0.152
ar.L18	0.0252	0.072	0.348	0.728	-0.117	0.167
ar.L19	-0.1036	0.074	-1.391	0.164	-0.250	0.042
ar.L20	0.1227	0.089	1.371	0.170	-0.053	0.298
ar.L21	-0.2142	0.100	-2.153	0.031	-0.409	-0.019
ar.L22	0.0652	0.110	0.590	0.555	-0.151	0.282
ar.L23	0.0229	0.084	0.273	0.785	-0.142	0.188
ar.L24	-0.0186	0.077	-0.241	0.810	-0.170	0.133
ar.L25	-0.0788	0.080	-0.986	0.324	-0.235	0.078
ar.L26	0.0261	0.086	0.302	0.763	-0.143	0.196
ar.L27	-0.1140	0.078	-1.471	0.141	-0.266	0.038
ar.L28	0.0940	0.092	1.024	0.306	-0.086	0.274
ar.L29	0.0036	0.084	0.043	0.966	-0.161	0.168
ar.L30	-0.0786	0.084	-0.939	0.348	-0.243	0.085
ar.L31	-0.0456	0.089	-0.514	0.607	-0.220	0.128
ar.L32	-0.0143	0.082	-0.174	0.862	-0.176	0.147
ar.L33	0.0223	0.089	0.251	0.801	-0.152	0.196
ar.L34	-0.0677	0.090	-0.756	0.450	-0.243	0.108
ar.L35	0.1056	0.083	1.278	0.201	-0.056	0.268
ar.L36	-0.0833	0.090	-0.923	0.356	-0.260	0.094
ar.L37	-0.0935	0.084	-1.114	0.265	-0.258	0.071
ar.L38	-0.0214	0.089	-0.241	0.809	-0.195	0.152
ar.L39	0.0143	0.070	0.203	0.839	-0.124	0.152
ar.L40	-0.1190	0.077	-1.544	0.123	-0.270	0.032
ar.L41	0.0031	0.092	0.034	0.973	-0.178	0.184
ar.L42	0.0383	0.086	0.445	0.656	-0.130	0.207
ar.L43	-0.1100	0.093	-1.183	0.237	-0.292	0.072
ar.L44	0.0297	0.097	0.306	0.760	-0.161	0.220
ar.L45	-0.0789	0.078	-1.006	0.314	-0.232	0.075
ar.L46	0.0344	0.089	0.385	0.700	-0.141	0.209
ar.L47	-0.0499	0.084	-0.596	0.551	-0.214	0.114
ar.L48	0.0023	0.082	0.028	0.977	-0.158	0.163
ar.L49	-0.0003	0.085	-0.003	0.998	-0.166	0.166
ar.L50	-0.0149	0.092	-0.163	0.871	-0.195	0.165
ar.L51	-0.0345	0.084	-0.410	0.682	-0.199	0.130
ar.L52	-0.0635	0.081	-0.779	0.436	-0.223	0.096
ar.L53	-0.0122	0.091	-0.133	0.894	-0.191	0.167
ar.L54	-0.0174	0.085	-0.205	0.838	-0.184	0.149

ar.L55	-0.0857	0.077	-1.108	0.268	-0.237	0.066
ar.L56	0.0072	0.082	0.088	0.930	-0.153	0.167
ar.L57	-0.0479	0.075	-0.641	0.521	-0.194	0.099
ar.L58	0.0292	0.077	0.379	0.705	-0.122	0.180
ar.L59	-0.0594	0.086	-0.691	0.489	-0.228	0.109
ar.L60	0.0332	0.081	0.409	0.683	-0.126	0.193
ar.L61	-0.0136	0.083	-0.165	0.869	-0.176	0.149
ar.L62	-0.0091	0.083	-0.111	0.912	-0.171	0.153
ar.L63	0.0710	0.074	0.956	0.339	-0.075	0.217
ar.L64	-0.1701	0.079	-2.144	0.032	-0.326	-0.015
ar.L65	-0.0100	0.110	-0.091	0.928	-0.226	0.206
ar.L66	0.0547	0.075	0.728	0.466	-0.093	0.202
ar.L67	-0.0148	0.082	-0.180	0.857	-0.176	0.146
ar.L68	-0.0354	0.088	-0.404	0.686	-0.207	0.136
ar.L69	0.0176	0.081	0.218	0.827	-0.140	0.176
ar.L70	-0.0168	0.079	-0.213	0.832	-0.172	0.138
ar.L71	0.0177	0.086	0.205	0.838	-0.152	0.187
ar.L72	0.0359	0.080	0.451	0.652	-0.120	0.192
ar.L73	0.0093	0.083	0.112	0.911	-0.153	0.172
ar.L74	-0.0800	0.085	-0.940	0.347	-0.247	0.087
ar.L75	-0.0758	0.096	-0.786	0.432	-0.265	0.113
ar.L76	0.0518	0.086	0.602	0.547	-0.117	0.220
ar.L77	-0.0934	0.085	-1.099	0.272	-0.260	0.073
ar.L78	-0.0285	0.100	-0.285	0.775	-0.224	0.167
ar.L79	0.1683	0.079	2.122	0.034	0.013	0.324
ar.L80	-0.1696	0.073	-2.323	0.020	-0.313	-0.026
ma.L1	-0.2635	0.431	-0.611	0.541	-1.109	0.581
sigma2	0.0266	0.002	12.917	0.000	0.023	0.031

Ljung-Box (L1) (Q):	0.07	Jarque-Bera (JB):	20.78
Prob(Q):	0.80	Prob(JB):	0.00
Heteroskedasticity (H):	0.46	Skew:	0.27
Prob(H) (two-sided):	0.00	Kurtosis:	3.97

Table 5: Result of the ARMA (80, 1) model for *mortgage rates*

The table above shows the result of the ARMA (80, 1) model, where we focused on the coefficients and found that only the following components are significant: autoregressive terms with lags of 2, 21, 64, 79, 80.

Step 3: Diagnostic Checking: Residual Errors

We checked the residuals from the chosen model and performed the Ljung-Box test.

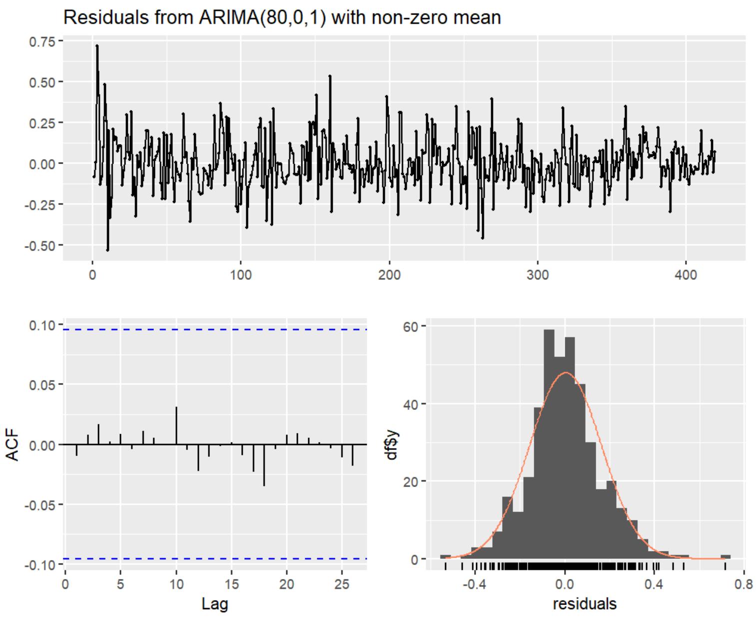

Figure 37: Residuals from the ARMA (80, 1) model

All spikes of the ACF plot are now within the significance limits, showing that there are no autocorrelations.

Ljung-Box Test

H_0: The residuals are independently distributed.

H_1: The residuals are not independently distributed and exhibit a serial correlation.

Results:

Ljung-Box statistics: 27.718847

p-value: 1.0

We failed to reject H_0. The Ljung-Box test shows that the residuals have no remaining autocorrelations. Therefore, we can confirm that the residuals of the ARMA (80, 1) model is white noise.

Thus, we now have a seasonal ARIMA model that passes the requirements and is ready for forecasting.

Forecasting process

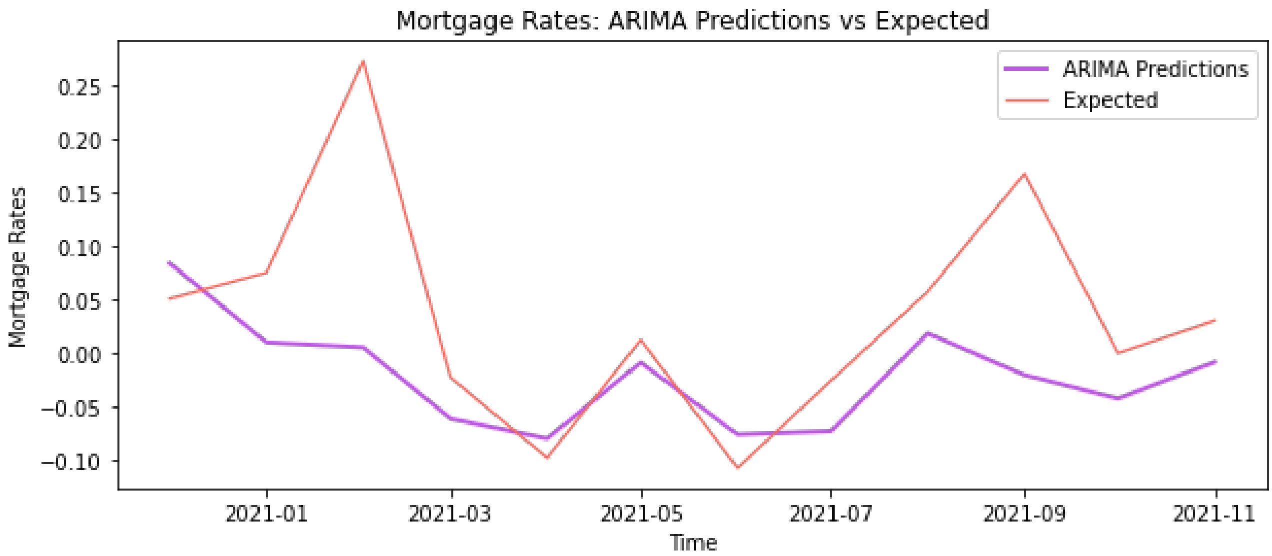

Figure 38: Comparison graph of expected values and ARIMA-predicted values

In the comparison graph, we observe that our model follows the patterns of the test set quite well. The predictions are quite accurate for the periods of March 2021 – July 2021 and October 2021 – November 2021. The largest gap is observed in the period from January 2021 to February 2021, and from August 2021 to October 2021. In 2020 and the first months of 2021, mortgage rates plummeted in response to the COVID-19 pandemic. These record-low rates were largely dependent on accommodating, Covid-era policies from the Federal Reserve[7]. However, right after the economic recovery in mid-2021, mortgage rates started to rise. Our model is not aware of the measures by the government, nor how fast the economy picked up, which explains the gap between the predictions and expected values.

[7] Peter M., 2023. Mortgage rates chart: Historical and current rate trends. *The mortgage reports*. Retrieved from: https://themortgagereports.com/61853/30-year-mortgage-rates-chart

Part 4.2: Dynamic model for house prices

In this part, we will build potential models based on the results of the previous part. For each approach, we will choose the most appropriate model, and try to compare which approach is better at predicting by looking at the comparison graphs.

Approach 01: Optimal model for the harmonic regression's residuals

After applying the method of differencing for two times to remove trend, our data lost 2 values, which equal to the 2 values of house prices recorded in November 2021 and December 2021. Hence, our house price time series is down to 418 values.

We split our data into two parts: the part being trained includes values from January 1987 to October 2019, and the part being tested includes values from November 2019 to October 2021, which also means we will try to predict house prices in the US for 24 successive months.

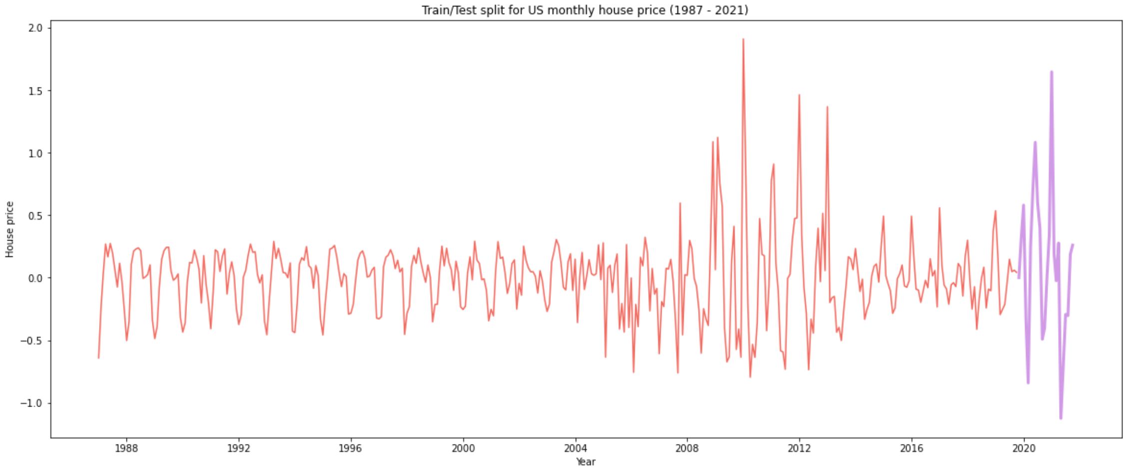

Figure 39: Division of train set and test set (of 24 months) for *house price*

The following steps are based on the Box-Jenkins methodology.

Step 1: Identification: Stationarity & Defining parameters of the model.

In part 3, we proved that our final data is free from stationarity using the ADF test.

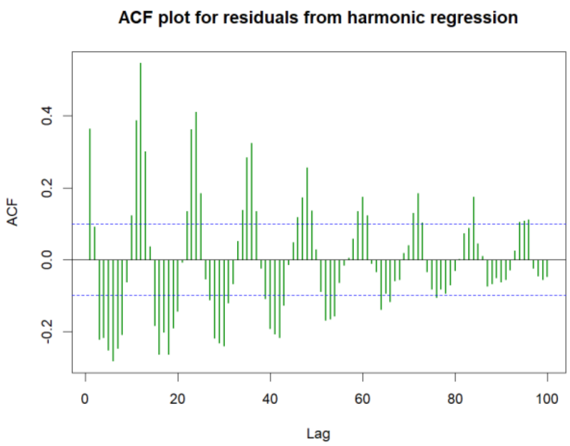

Figure 40: ACF plot for residuals from harmonics regression

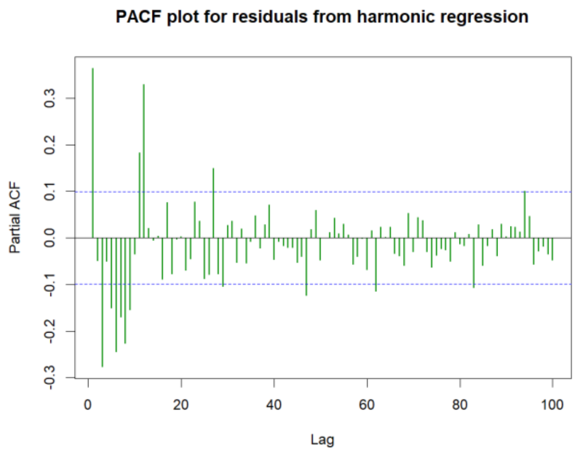

Figure 41: Plot for *house price* (2nd differencing) from harmonics regression

Both the ACF and PACF plots show a decaying characteristic, meaning the autocorrelation and partial autocorrelation become less significant gradually. In the ACF plot, there is clear evidence of annual seasonality as the spikes reach a peak after every 12 lags (sine-curved pattern). In the PACF plot, the majority of the first 12 lags are significant, until there comes a sudden cut-off after lag 12, after which the spikes begin tapering between the positive and negative zones.

With a gradual geometrically declining ACF and a PACF that is significant for only a few lags, the time series can be a seasonal-AR (12) process.

Step 02: Estimation

Apart from our own speculations derived from the ACF and PACF plots, we used the auto.arima function in the R program. It suggested a SARIMA model with parameters (3, 0, 2)(0, 1, 2)[12], which indicates a non-seasonal second-order MA component with a non-seasonal third-order AR component, and seasonal differencing with a seasonal second MA order, and a time span of annual seasonal pattern.

In order to optimize the suggested model, we withdrew the residuals and checked the ACF and PACF plots to adjust a few parameters and delivered a comparison between the potential models at the end.

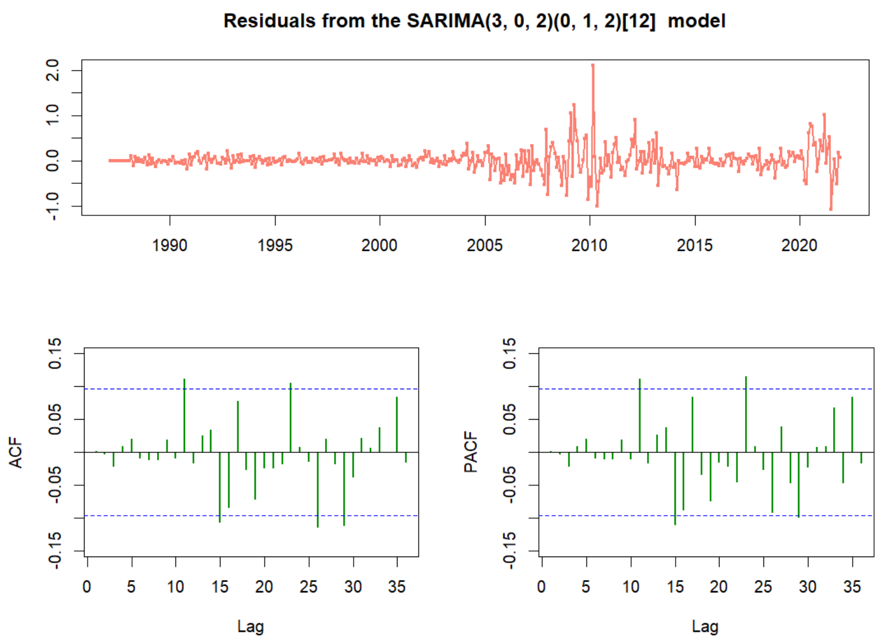

Figure 42: General plot, ACF and PACF plots for residuals from model SARIMA (3, 0, 2)(0, 1, 2)[12] for *house price*

The plots show that there are significant spikes corresponding to the moving-average and the autoregressive components. There is no tapering pattern, and the lags are significant in a random order, so for the SARIMA model we tried additional AR terms, corresponding to p = 11, p = 15, p = 23, p = 29, and p = 12 (from step 1).

Hence, we adjusted the parameters, ran the models on Python and received the following results:

Model	AIC	BIC
SARIMA (3, 0, 2) (0, 1, 2) [12]	33.29	64.85
SARIMA (11, 0, 2) (0, 1, 2) [12]	12.12	79.25
SARIMA (12, 0, 2) (0, 1, 2) [12]	**12.02**	**79.09**
SARIMA (15, 0, 2) (0, 1, 2) [12]	9.55	88.46
SARIMA (23, 0, 2) (0, 1, 2) [12]	12.71	123.18
SARIMA (29, 0, 2) (0,1,2) [12]	17.59	151.74

Table 6: Comparison of models for *house price* based on AIC and BIC criteria (approach 01)

We took into account both the AIC and BIC criteria, and therefore decided to choose the SARIMA (12, 0, 2) (0, 1, 2) [12] model as the optimal model for further analysis. This model is characterized by a non-seasonal second MA order with a twelfth AR order, seasonal differencing with a seasonal second MA order, and a time span of annual seasonal pattern.

	coef	std err	z	P>|z|	[0.025	0.975]
ar.L1	0.1585	0.107	1.488	0.137	-0.050	0.367
ar.L2	0.0488	0.133	0.366	0.714	-0.213	0.310
ar.L3	-0.3049	0.036	-8.370	0.000	-0.376	-0.234
ar.L4	0.0202	0.056	0.360	0.719	-0.090	0.131
ar.L5	-0.0258	0.067	-0.383	0.702	-0.158	0.106
ar.L6	-0.1261	0.046	-2.762	0.006	-0.216	-0.037
ar.L7	-0.0641	0.040	-1.613	0.107	-0.142	0.014
ar.L8	-0.0495	0.039	-1.262	0.207	-0.126	0.027
ar.L9	-0.0315	0.041	-0.765	0.444	-0.112	0.049
ar.L10	-0.0411	0.038	-1.084	0.278	-0.115	0.033
ar.L11	0.1609	0.037	4.380	0.000	0.089	0.233
ar.L12	0.3493	0.082	4.238	0.000	0.188	0.511
ma.L1	-0.0096	0.115	-0.084	0.933	-0.234	0.215
ma.L2	-0.0278	0.127	-0.219	0.827	-0.276	0.221
ma.S.L12	-1.0598	0.089	-11.950	0.000	-1.234	-0.886
ma.S.L24	0.1603	0.067	2.386	0.017	0.029	0.292
sigma2	0.0533	0.002	21.822	0.000	0.049	0.058

Ljung-Box (L1) (Q):	0.00	Jarque-Bera (JB):	1070.58
Prob(Q):	1.00	Prob(JB):	0.00
Heteroskedasticity (H):	14.56	Skew:	0.95
Prob(H) (two-sided):	0.00	Kurtosis:	10.98

Table 7: Result of the SARIMA (12, 0, 2) (0, 1, 2) [12] model for *house price* (approach 1)

The table above shows the result of the SARIMA (12, 0, 2) (0, 1, 2) [12] model, where we focused on the coefficients and found that only the following components are significant: non-seasonal autoregressive terms with lags of 3, 6, 11, 12 and seasonal moving-average terms with lags of 12, 24.

Step 3: Diagnostic Checking: Residual Errors

We checked the residuals from the chosen model and performed the Ljung-Box test.

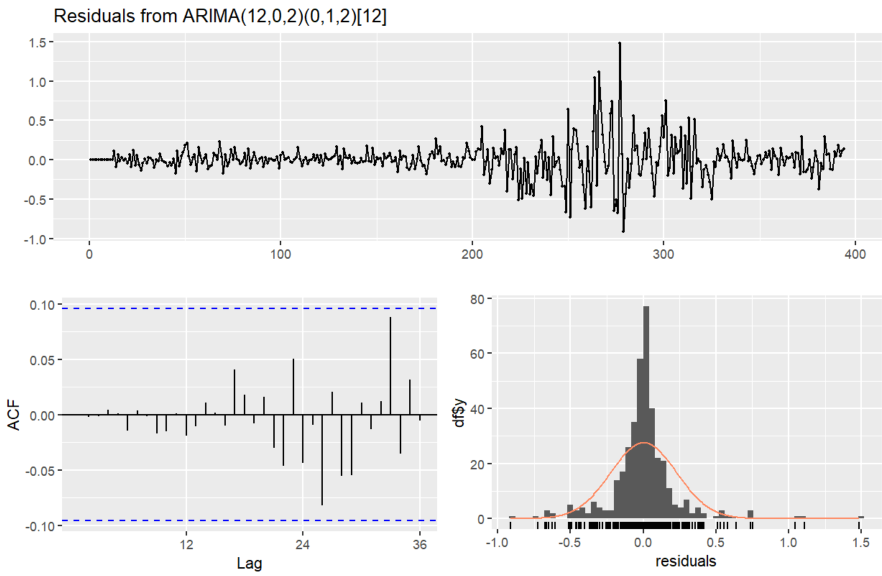

Figure 43: Residuals from the SARIMA (12, 0, 2)(0, 1, 2)[12] model

All spikes of the ACF plot are now within the significance limits, showing that there are no autocorrelations.

Ljung-Box Test

H_0: The residuals are independently distributed.

H_1: The residuals are not independently distributed and exhibit a serial correlation.

Results:

Ljung-Box statistics:	0.360938
p-value:	0.999999

We failed to reject H_0. The Ljung-Box test shows that the residuals have no remaining autocorrelations. Therefore, we can confirm that the residuals of the SARIMA (12, 0, 2)(0, 1, 2)[12] model is white noise.

Thus, we now have a seasonal ARIMA model that passes the requirements and is ready for forecasting.

Forecasting process

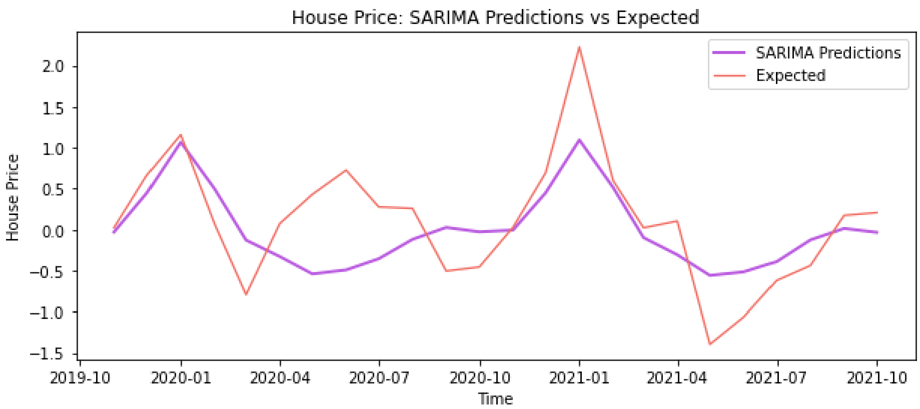

Figure 44: Comparison graph of expected values and SARIMA-predicted values

In the comparison graph, we observe that our model follows the patterns of the test set quite well. The predictions are quite accurate for the periods of November 2019 – March 2020, November 2020 – March 2021, and July 2021 – October 2021. The largest gap is observed in the period from April 2020 to August 2020, which can be explained by the fact that these are the first months of COVID-19, and for such exceptional times, it is very difficult for models based on a train set in normal times to predict such a heavy plunge.

Approach 02: Optimal model for a de-seasonalized time series (after three times of differencing)

After applying the method of differing for three times, our data lost 14 values in total, which equal to 14 values of house prices recorded monthly from November 2020 to December 2021. We consider it a fair lost in exchange for a stationary, maximumly detrend and de-seasonalized dataset, and also because the COVID-19 pandemic is an external factor that would probably undermine the accuracy of our predictions.

Hence, we split our data into two parts: the part being trained includes values from January 1987 to October 2019, and the part being tested includes values from November 2019 to October 2020, which also means we will try to predict house prices in the US for 12 successive months.

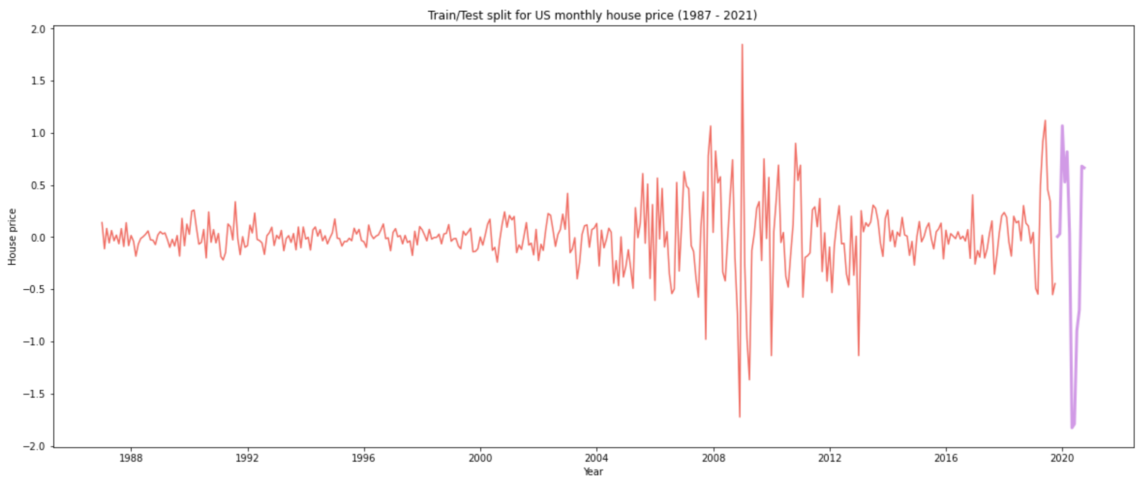

Figure 45: Division of train set and test set (of 12 months) for *house price*

The following steps are based on the Box-Jenkins methodology.

Step 1: Identification: Stationarity & Defining parameters of the model.

In the third part, we proved that our final data is free from stationarity with the ADF test.

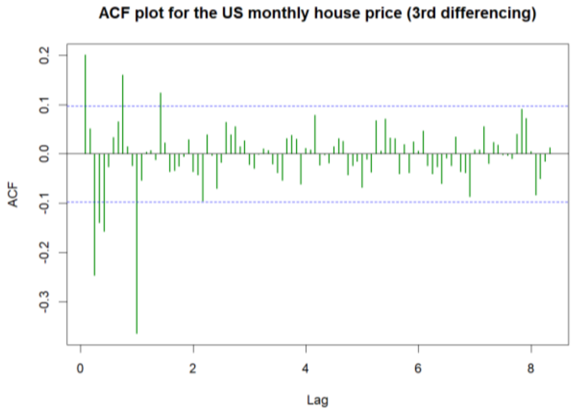

Figure 46: ACF plot for *house price* (3rd differencing)

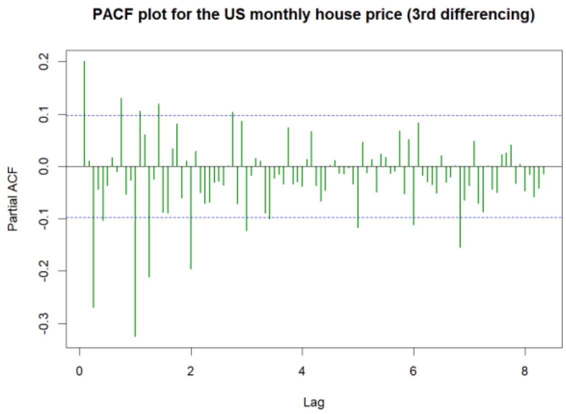

Figure 47: PACF plot for *house price* (3rd differencing)

From the ACF plot, it can be seen that there is a cut-off after lag 12, after which the lags are insignificant with the exception of lag 17. From the PACF plot, there seems to be a decaying characteristic but there are no significant lags, after which the PACF die down. Hence, from these plots, we can only derive an MA (12) or an MA (17) component. Further speculations will be made at the next step.

Step 02: Estimation

Apart from our own speculations derived from the ACF and PACF plots, we used the auto.arima function in the R program, and it suggested an ARIMA model with parameters (0,0,4), which indicates a fourth-order MA component.

In order to optimize the suggested model, we withdrew the residuals and checked the ACF and PACF plots to adjust a few parameters, then we delivered a comparison between the potential models at the end.

Figure 48: General plot, ACF and PACF plots for residuals from model ARIMA (3, 0, 4)

After deriving the residuals of the suggested model, we can see that both plots tail off to 0. We chose the following parameters for an ungraded ARMA (p, q) model from the suggested one: p = 5, p = 12, p = 14, p = 24, q = 5, q = 12, q = 17.

We adjusted the non-seasonal AR order and ran the models on Python and received the following results:

Model	AIC	BIC
ARIMA (3, 0, 4)	167.55	191.41
ARIMA (5, 0, 4)	145.16	188.90
ARIMA (12, 0, 4)	90.94	162.51
ARIMA (14, 0, 4)	90.17	169.69
ARIMA (24, 0, 4)	83.46	202.75
ARIMA (5, 0, 5)	121.15	168.86
ARIMA (5, 0, 12)	**50.07**	**125.62**
ARIMA (5, 0, 17)	60.96	156.23
ARIMA (12, 0, 5)	86.94	162.49

ARIMA (12, 0, 12)	65.26	168.65
ARIMA (12, 0, 14)	65.53	189.87
ARIMA (14, 0, 5)	89.92	185.35
ARIMA (14, 0, 12)	51.05	174.32
ARIMA (14, 0, 17)	60.77	231.75
ARIMA (24, 0, 5)	86.28	209.55
ARIMA (24, 0, 12)	59.55	210.65
ARIMA (24, 0, 17)	69.63	240.61

Table 8: Comparison of models for *house price* based on AIC and BIC criteria (approach 02)

We took into account both the AIC and BIC criteria, and therefore decided to choose the ARIMA (5, 0, 12) model as the optimal model for further analysis. This model is characterized by a twelfth MA order and a fifth AR order.

	coef	std err	z	P>|z|	[0.025	0.975]
const	0.0010	0.002	0.595	0.552	-0.002	0.004
ar.L1	0.1603	0.057	2.791	0.005	0.048	0.273
ar.L2	0.0672	0.054	1.241	0.214	-0.039	0.173
ar.L3	-0.1861	0.061	-3.033	0.002	-0.306	-0.066
ar.L4	0.0954	0.056	1.714	0.086	-0.014	0.205
ar.L5	-0.0190	0.071	-0.270	0.787	-0.157	0.119
ma.L1	0.0870	0.064	1.370	0.171	-0.037	0.211
ma.L2	-0.0564	0.074	-0.767	0.443	-0.200	0.088
ma.L3	-0.1464	0.062	-2.343	0.019	-0.269	-0.024
ma.L4	-0.1431	0.075	-1.910	0.056	-0.290	0.004
ma.L5	-0.0576	0.076	-0.757	0.449	-0.207	0.092
ma.L6	-0.0623	0.058	-1.067	0.286	-0.177	0.052
ma.L7	-0.0476	0.050	-0.956	0.339	-0.145	0.050
ma.L8	-0.0212	0.057	-0.373	0.710	-0.133	0.090
ma.L9	-0.0175	0.051	-0.344	0.731	-0.117	0.082
ma.L10	0.0227	0.052	0.434	0.664	-0.080	0.125
ma.L11	0.1942	0.046	4.253	0.000	0.105	0.284
ma.L12	-0.7079	0.054	-13.054	0.000	-0.814	-0.602
sigma2	0.0583	0.004	15.053	0.000	0.051	0.066

Ljung-Box (L1) (Q):	0.08	Jarque-Bera (JB):	935.42
Prob(Q):	0.78	Prob(JB):	0.00
Heteroskedasticity (H):	15.68	Skew:	0.85
Prob(H) (two-sided):	0.00	Kurtosis:	10.36

Table 9: Result of the ARIMA (5, 0, 12) model for *house price* (approach 2)

The table above shows the result of ARIMA (5, 0, 12) model, where we focused on the coefficients and found that only the following components are significant: autoregressive terms with lags of 1, 3, 4 and moving-average terms with lags of 3, 4, 11, 12.

Step 3: Diagnostic Checking: Residual Errors

We checked the residuals from the chosen model and performed the Ljung-Box test.

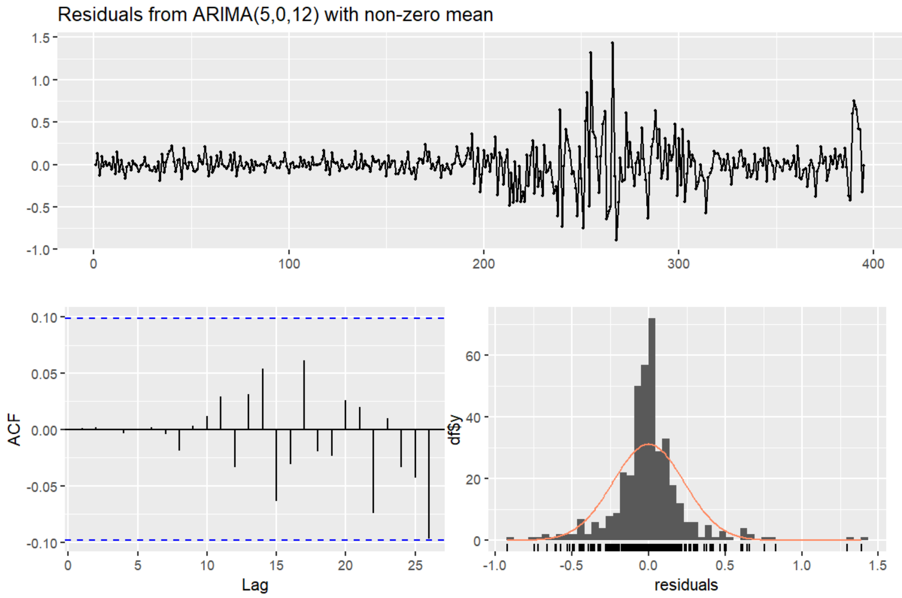

Figure 49: Residuals from the ARIMA (5, 0, 12) model

All spikes of the ACF plot are now within the significance limits, showing that there are no autocorrelations.

Ljung-Box Test

H_0: The residuals are independently distributed.

H_1: The residuals are not independently distributed and exhibit a serial correlation.

Results:

Ljung-Box statistics:	2.159917
p-value:	0.994961

We failed to reject H0. The Ljung-Box test shows that the residuals have no remaining autocorrelations. Therefore, we can confirm that the residuals of the SARIMA (12, 0, 2)(0, 1, 2)[12] model is white noise.

Thus, we now have a seasonal ARIMA model that passes the requirements and is ready for forecasting.

Forecasting process

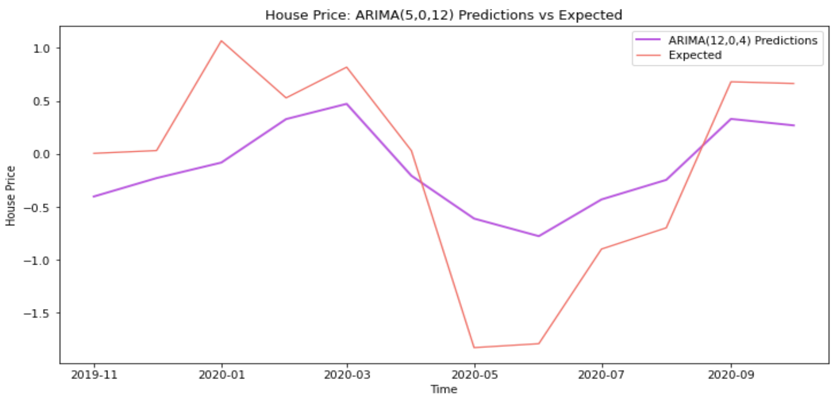

Figure 50: Comparison graph of expected values and ARIMA-predicted values

In the comparison graph, we observe that our model follows the trend of the test set quite well. The predictions are quite accurate for the periods of March 2020 – April 2020 and August 2020 – September 2020, where there are intersections. The largest gap is observed in the period from April 2020 to August 2020, which can be explained by the fact that these are the first months of COVID-19, and for such exceptional times, it is very difficult for models based on a train set in normal times to predict such a heavy plunge. (A similarly large gap was also observed in the model of our 1st experiment, which suggests that this is an unstable period that is difficult to predict).

Conclusion

To sum up, in our study, we have managed to study the behavior of house prices and mortgage rates in the US from 1987 to 2021, in which there are two periods of recession.

As regards to mortgage rate, we applied polynomial regression, derived from it residuals and applied a first-order differencing for them to clear trend. It took us quite some time to receive the result of the appropriate model for mortgage rates because of the high AR order, the model performed relatively well.

As regards to house price, after having removed the detected trend by two rounds of differencing, we took two approaches to handle seasonality and building models. For approach 01, we built a harmonic regression, withdrew residuals to forecast the next 2 years' values using a seasonal-ARIMA model. For approach 02, we differenced the data for the 3rd time, and used the de-seasonalized data to forecast the next 1 year's values using an ARIMA model. After comparing the performance of the two models with the real values, we conclude that the seasonal-ARIMA model for the residuals from harmonic regression is not only better at predicting, but also can predict well for at least 2 years, where as the ARIMA model for the third-order differenced data can only predict well enough for one year.

We also managed to test three hypotheses. The 1st hypothesis was rejected, because although there was underlying seasonality, house prices actually went up more during winter months and down during summer-autumn months. The 2nd hypothesis was confirmed, because during both the recession periods, mortgage rates always went down. The 3rd hypothesis was rejected, because whereas house prices dropped during the GFC in 2007 – 2009, they increased during the COVID-19 pandemic and the period afterwards.

Our final derivations from our research are:

- summer-autumn months are the most suitable time of the year for potential house buyers and house owners, because of the increase in both demand and supply;
- mortgage rates tend to decrease during recessions, but this should not be taken too seriously because in general, the decline over time;
- it is not always the case that house prices would drop during recessions because the FED can implement difference policies, so buyers should pay attention to updated policies from the central bank.