Time Series Analysis of Average House Sales Prices

Victoria Agboola

2024-04-29

Abstract

In this comprehensive study, I explored the modeling of average house sales prices in the United States from January 1963 to October 2023. The initial phase of my analysis involved deterministic trend fitting and seasonal adjustments, which unfortunately proved inadequate in capturing the complexity of the data, as diagnosed by persistent autocorrelation in the residuals. To address these shortcomings, I pivoted to the more sophisticated AutoRegressive Integrated Moving Average (ARIMA) modeling technique. This shift was motivated by ARIMA's ability to effectively manage non-stationarity and enhance forecast accuracy through its integration of differencing. The models I developed included a Moving Average (MA) of order 4, an Autoregressive (AR) of order 4, and an optimally selected ARIMA(2,2,1) model. Each model underwent a thorough diagnostic check to ensure adequacy and then faced evaluation against a reserved test set. The evaluation results revealed that the ARIMA(2,2,1) model exhibited the best performance, characterized by the most favorable Mean Absolute Error (MAE), Mean Squared Error (MSE), and Mean Absolute Percentage Error (MAPE). These findings underscore the robustness of ARIMA modeling in capturing and forecasting the dynamics of the housing market time series data, with the ARIMA(2,2,1) model standing out as a particularly effective forecasting tool.

Introduction

This report analyzes the average sales prices of houses in the United States, utilizing a dataset spanning from January 1963 to October 2023. The data, obtained from the Federal Reserve Economic Data (FRED) repository under the series "ASPUS," is reported quarterly and denominated in U.S. dollars. It represents a comprehensive measure of housing market trends over a significant period, offering insights into both economic stability and consumer confidence.

The objective of this analysis is to dissect and understand the underlying patterns in the housing market, encapsulated by the sales prices. Recognizing trends, seasonalities, and cycles within this time series data not only serves as an academic endeavor but also provides practical forecasting models that can inform stakeholders and policymakers in their decision-making processes.

Employing time series analytical techniques, this project endeavors to model the complexities inherent in economic data, such as non-stationarity and seasonality. The process entails rigorous exploratory data analysis, the application of various time series models, and the evaluation of these models based on their predictive performance.

The following sections will offer a detailed account of the exploratory data analysis conducted, describe the rationale behind the selection of specific time series models, and discuss the results derived from this methodical approach. This investigation culminates in a set of conclusions that synthesize findings, limitations, and prospective avenues for future research in the field.

Data Analysis

To begin the analysis, I loaded the quarterly average house sales price data from 1963 to 2023 into the analysis environment and formatted it as a time series. This conversion is crucial as it allows for the effective application of various time series analysis techniques later on in the study.

```
## $tsp
## [1] 1963.00 2023.75    4.00
##
## $class
## [1] "ts"

## [1] 244
```

Following the initial data structuring, a detailed visualization was created to better understand the trends and fluctuations over the decades.

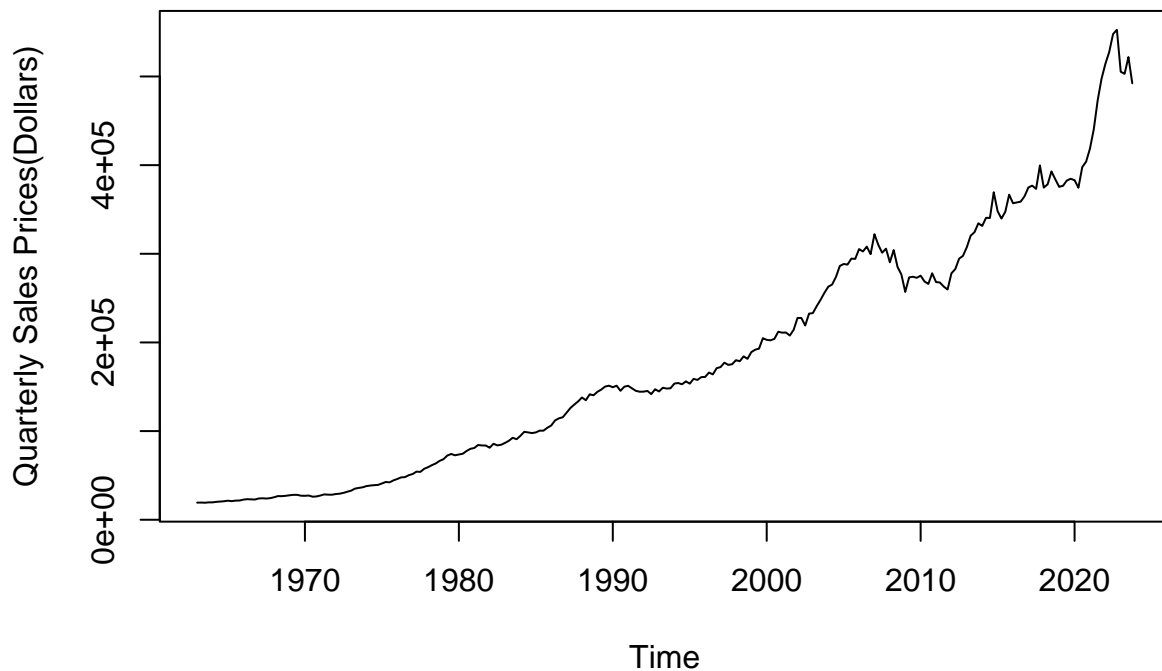## Average Sales Prices of Houses Sold 1963 – 2023

Figure 1: Average Sales Prices of Houses Sold 1963 - 2023

Figure 1 illustrates the quarterly sales prices of houses across the United States from 1963 to 2023. The graph effectively captures the significant trends and fluctuations in the housing market, emphasizing the long-term growth despite periodic volatilities. The upward trend and occasional sharp dips observed in the plot reflect the broader economic conditions influencing the housing market. This visual representation aids in providing a clear perspective on the overall market stability and investment potential, offering an intuitive understanding of the market's historical trends.

Following this initial visualization, Figure 2 presents the ACF plot showing the correlation of the time series with itself at different lags. This plot is essential for identifying the extent of serial dependence within the data, which is crucial for the choice of subsequent statistical modeling approaches. The persistence of autocorrelation across multiple lags indicates that past values have a considerable influence on future values, underscoring the need for models that can incorporate this dependency.

Polynomial Trend Analysis

In the initial phase of the analysis, I segmented the dataset into a training set, spanning from 1963 to 2011, and a testing set covering the years 2012 to 2023. This division aimed to evaluate the effectiveness of polynomial trend regression models in capturing the nuanced trends observed in the historical housing
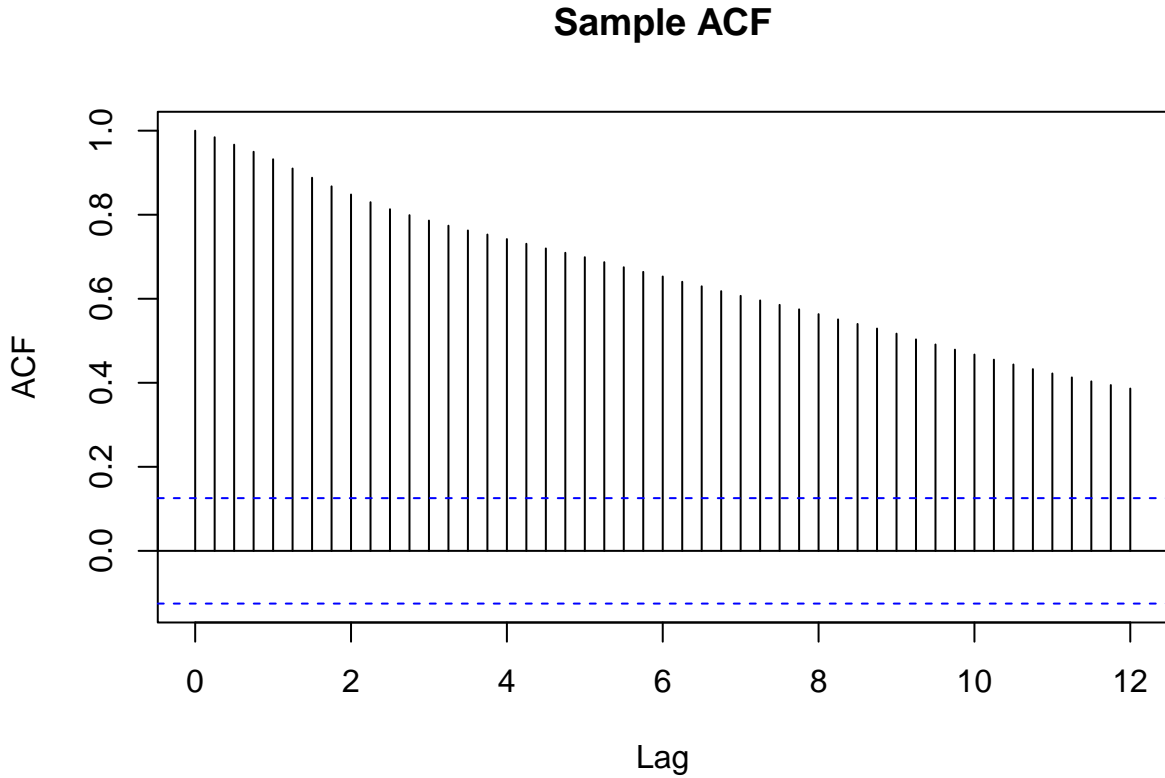
Sample ACF

Figure 2: ACF plot of Quarterly Sales Prices

market data. By exploring polynomial trends, ranging from linear to potentially higher-order trends, I sought to gain insights into the underlying dynamics driving housing prices over time. This approach served as a preliminary exploration to discern the presence of any discernible trends and paved the way for further analyses aimed at uncovering the complexities inherent in the dataset.

To explore the potential of polynomial trend models in capturing the intricate dynamics of housing price trends, I proceeded to fit linear, quadratic, and cubic trend models to the training dataset. Beginning with a linear trend model, I used ordinary least squares regression to fit a straight line to the data, providing a baseline for comparison. Next, I extended the analysis to quadratic and cubic trends, incorporating polynomial terms of degree two and three, respectively, to capture potential non-linear patterns in the data. Each model's adequacy was visually assessed through residual diagnostic plots, which included plots of residuals against fitted values, quantile-quantile plots, scale-location plots, and residual histograms. These plots provided insights into the presence of any patterns or systematic deviations in the residuals, aiding in the evaluation of the models' suitability for capturing the underlying trends in the housing price data.
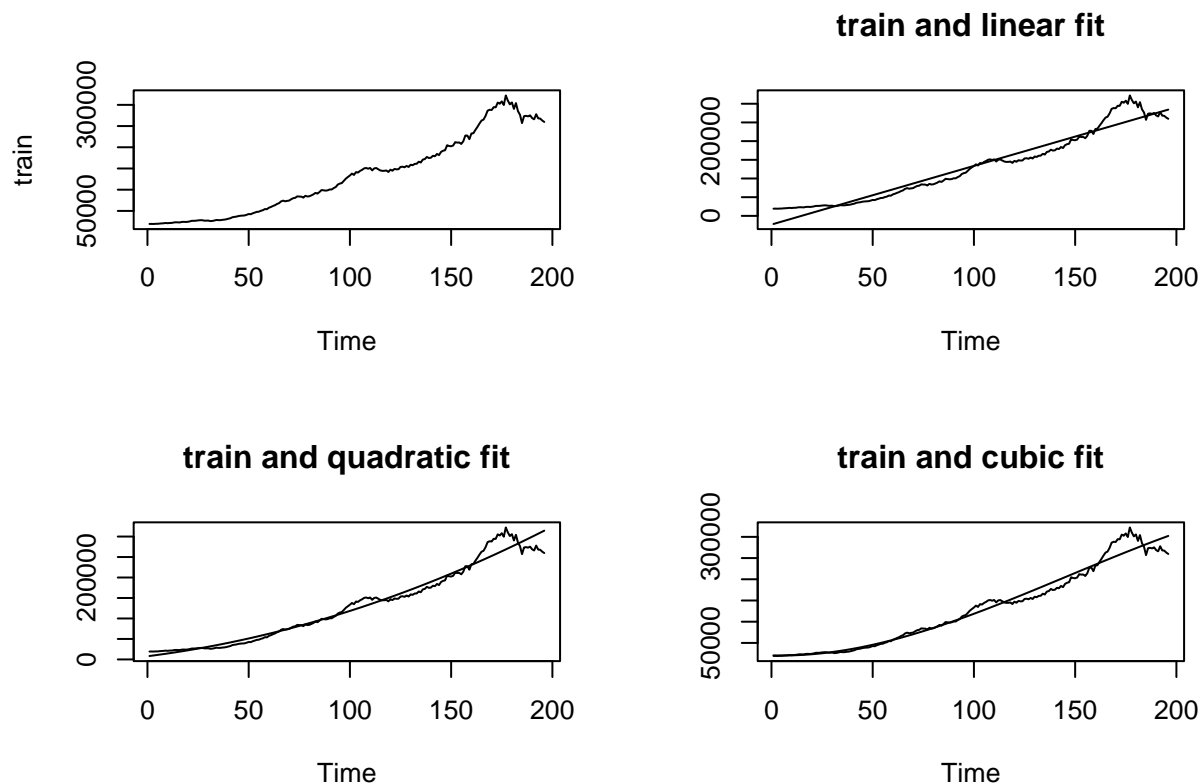
After fitting linear, quadratic, and cubic trend models to the dataset, the next step was to select the most appropriate model among them. This model selection process involved comparing the full model (cubic trend model) to the simpler linear model using ANOVA test. The ANOVA test results indicated a preference for the full model. Additionally, the Akaike Information Criterion (AIC) was calculated for each model, with the cubic trend model exhibiting the lowest AIC value, further supporting its preference over the linear and quadratic models.

```
## The AIC for linear fit is: 22.7589
```

```
## The AIC for quadratic fit is: 22.20852
```

```
## The AIC for cubic fit is: 22.12326
```

Seasonality Fitting Using Regression with Seasonal Indicators

In this section, seasonal indicators were incorporated into the regression model to account for any recurring patterns or fluctuations in the housing price data. Three models were fitted: one with linear trends and seasonal indicators, another with quadratic trends and seasonal indicators, and a third with cubic trends and seasonal indicators. Each model aimed to capture the seasonal variations present in the dataset and assess their impact on housing prices over time. The models were evaluated based on their coefficients and goodness-of-fit statistics to determine the effectiveness of incorporating seasonal indicators in modeling the housing market trends.

```
## # A tibble: 2 x 5
##   Model                                  AIC        MSE    MAE  MAPE
##   <chr>                                <dbl>      <dbl>  <dbl> <dbl>
## 1 Quadratic Model with Seasonal Indicators  22.2 1697657676. 26318.  6.00
## 2 Cubic Model with Seasonal Indicators      22.2 5577778373. 54669. 12.4
```

The assessment of model fit involved analyzing in-sample and out-of-sample prediction errors using AIC, mean squared error (MSE), mean absolute error (MAE), and mean absolute percentage error (MAPE) metrics. On the training set, the model with a cubic trend and seasonal indicators demonstrated better performance

based on the information criterion estimation. However, when evaluating the models on the test data, the quadratic model with seasonal indicators exhibited superior performance, showcasing smaller MSE, MAE, and MAPE values. Despite slight differences in goodness-of-fit statistics, the emphasis on out-of-sample predictive accuracy favored the quadratic model for forecasting purposes. This prioritization of out-of-sample performance underscores the model's ability to generalize effectively to unseen data, enhancing its reliability for future forecasts.

Detrending Fitted Models

To evaluate the adequacy of the models, a thorough examination of the residuals was conducted by detrending each model's output. This involved extracting the residuals from the linear, quadratic, and cubic trend models with seasonal indicators. Subsequently, the autocorrelation functions (ACFs) of the detrended time series were scrutinized in comparison to the observed series. Despite initial expectations of achieving model adequacy through deterministic trend fitting, the ACF plots revealed persistent autocorrelation with slow decay in the detrended series. This outcome suggests that the deterministic components of the models alone were insufficient to capture the nuanced dynamics of the time series. As a result, it becomes evident that additional modeling considerations are necessary to fully account for the stochastic elements inherent in the data.
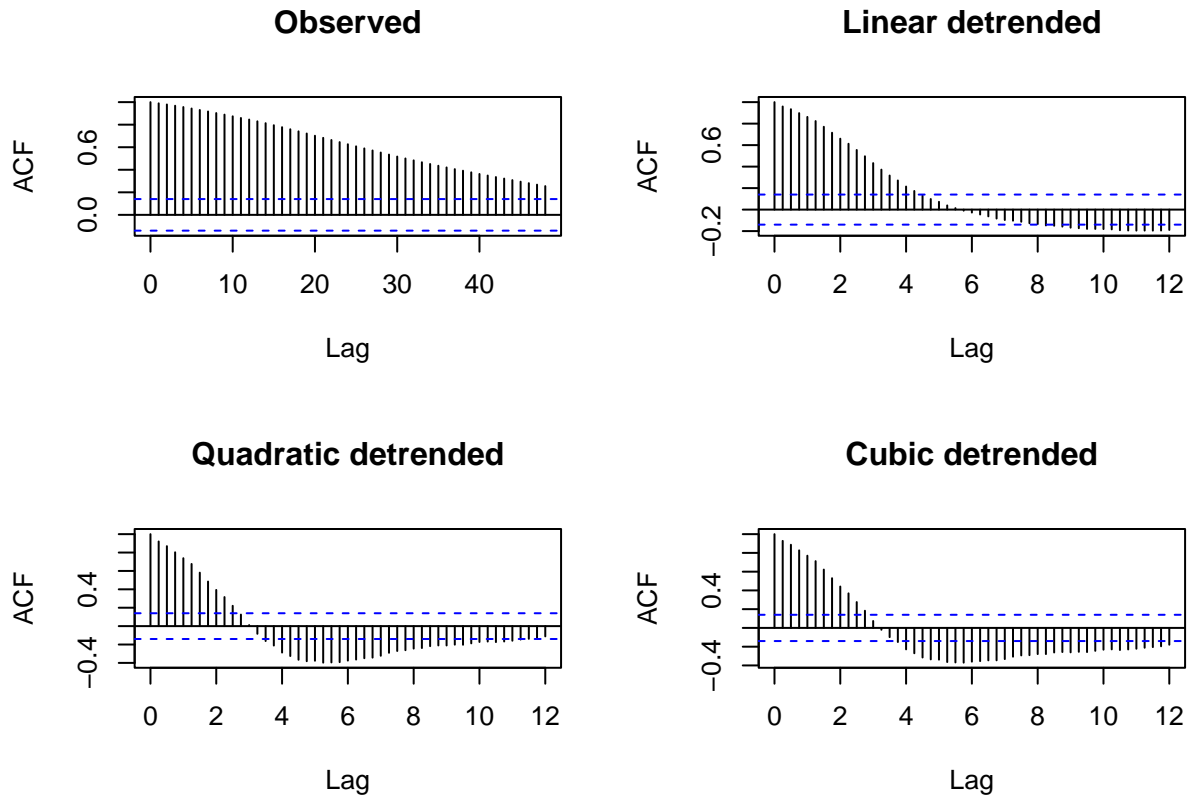

Figure 3: Detrended Time Series

Transition to ARIMA Modeling

After evaluating the detrending methods, it became apparent that they were insufficient in capturing the complexity of the time series data. Residual checks revealed persistent non-stationarity, indicating the shortcomings of deterministic approaches. Consequently, a transition to ARIMA modeling was pursued.

ARIMA offers an integrated solution, accounting for trends and fluctuations within the data itself. Its capability to handle non-stationarity and potential for enhanced forecast accuracy make it a logical next step.

Model diagnostics will be employed to confirm the suitability of this approach.

The decision to adopt ARIMA aligns with the goal of developing a robust model capable of accurately capturing and predicting the time series data.

The sample ACF in Figure 2 typically shows slow decay as the lag increases, we may consider the series to be nonstationary. We also see the upward trend in the time series plot in Figure 1.
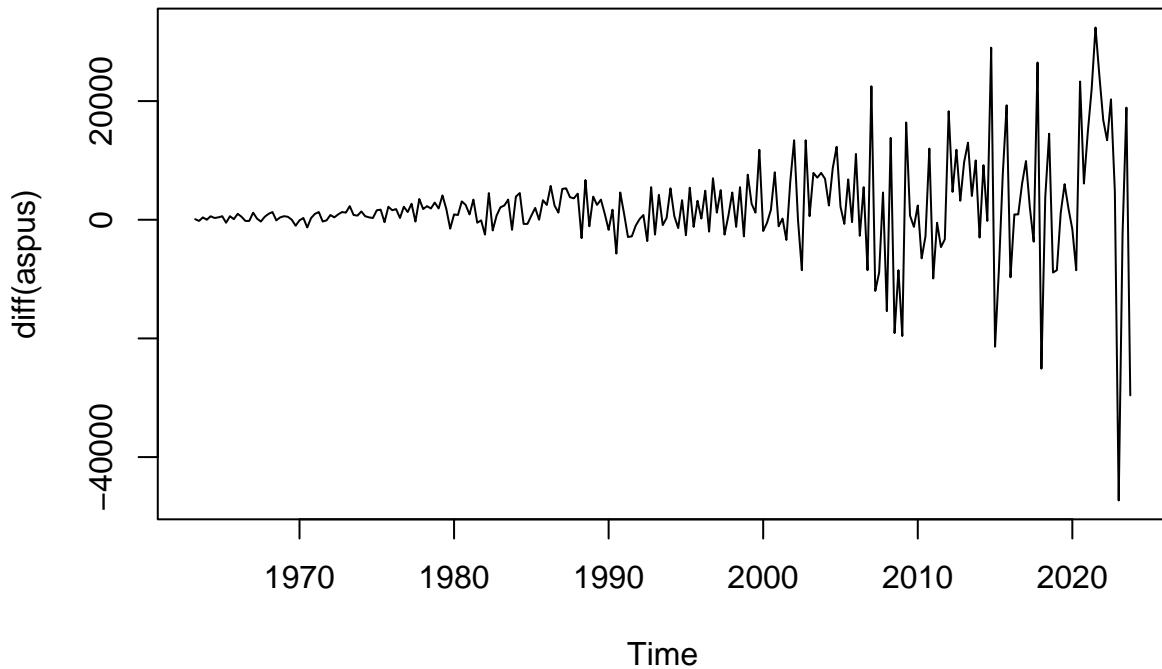
We may consider the differenced series:

Figure 4: Time series with single differencing

The time series displays increased volatility in the later years following a single differencing. To address this, the data were transformed into growth rates by taking the difference of the logarithms, which helps to stabilize the variance across the series.

We then used the Augmented Dickey-Fuller (ADF) unit root tests to check the stationarity of the growth rate.

```
##
## Title:
##   Augmented Dickey-Fuller Test
##
## Test Results:
##   PARAMETER:
##     Lag Order: 7
##   STATISTIC:
##     Dickey-Fuller: -4.585
##   P VALUE:
##     0.01
##
## Description:
##   Mon Apr 29 18:49:19 2024 by user:
```

Upon confirmation of stationarity, the next step involved identifying potential models by examining the
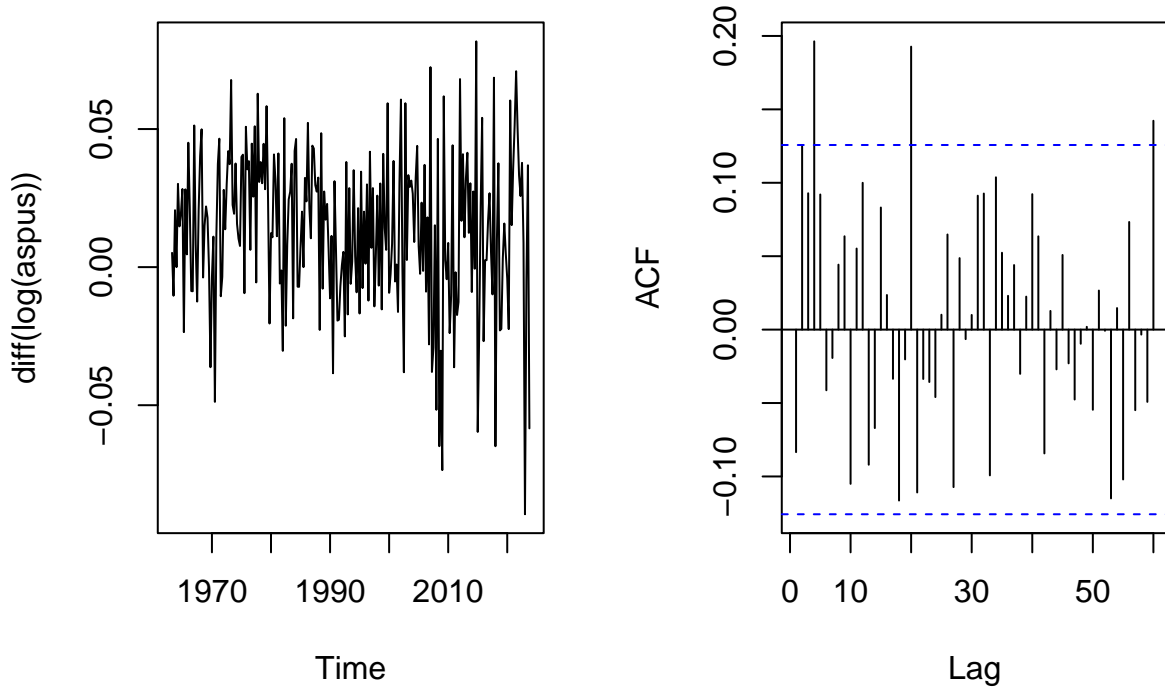
Growth Rate of Quarterly US ASI

Figure 5: Time Series and ACF plot of Growth Rate

autocorrelation function (ACF) of the differenced log series. This approach aimed to uncover any significant autocorrelation patterns that could inform the selection of an appropriate model.

Identifying p and q for ARIMA model

```
### Split the time series into calibration portion and holdout portion
aspusgr <- log(aspus)
aspusgr.all=aspusgr
aspusgr.calib=aspusgr.all[1:231]
aspusgr.hold=aspusgr.all[232:243]
```

To identify the appropriate model for the stationary series, I employed the autocorrelation function (ACF) and partial autocorrelation function (PACF) plots. These plots revealed significant spikes at specific lags, providing valuable insights into the underlying structure of the time series. In particular, the ACF plot displayed notable spikes at lag 4 and lag 20, while the PACF plot exhibited significant spikes at lag 4 and lag 5.

In line with the principle of parsimony, I opted to begin with a Moving Average (MA) model of order 4 (MA(4)) to ascertain its ability to capture the complexity of the data and serve as a suitable framework for forecasting purposes. This decision was guided by the observed spikes in the ACF and PACF plots, indicating potential dependencies at these lags. By starting with a MA(4) model, I aimed to strike a balance between model simplicity and predictive capability, laying the groundwork for further model refinement and evaluation.
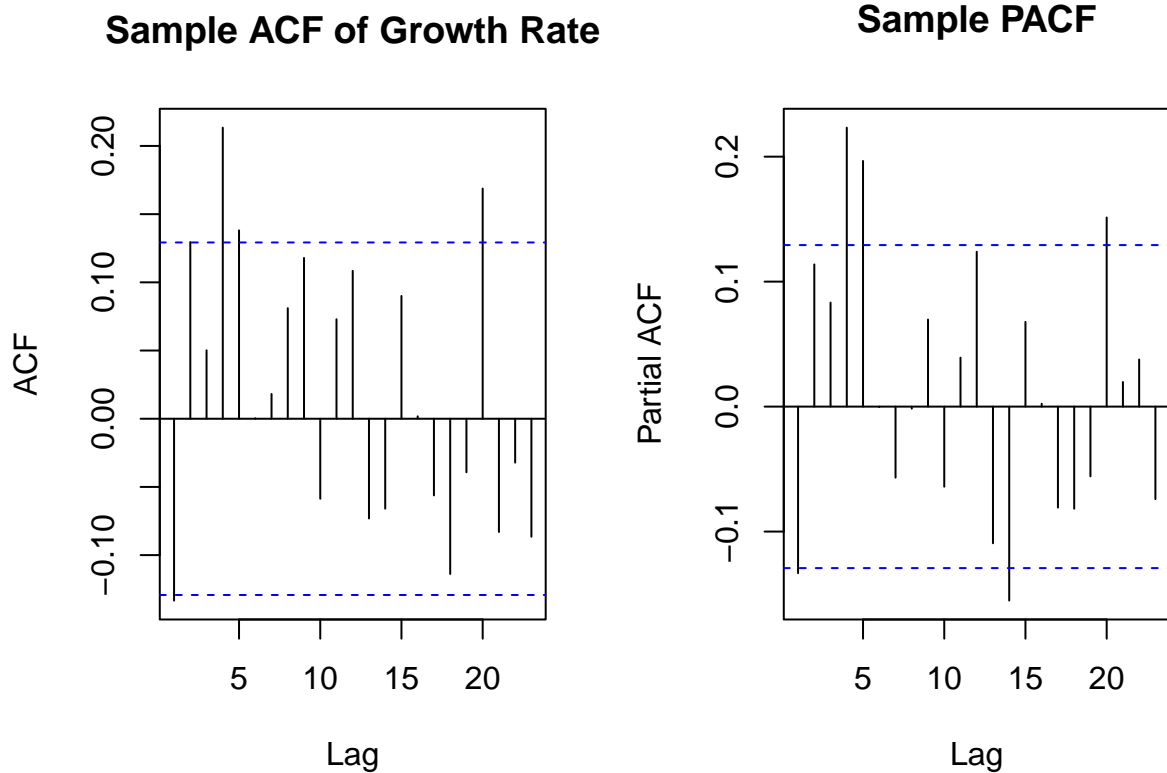
7

Figure 6: Sample ACF of Growth Rate

The following output shows a summary of the MA(4) or ARIMA(0,1,4) model on the growth rate calibration data:

```
##
## Call:
## TSA::arima(x = aspusgr.calib, order = c(0, 1, 4), include.mean = TRUE)
##
## Coefficients:
##           ma1      ma2     ma3     ma4
##       -0.1259  0.1694  0.1310  0.3052
## s.e.   0.0659  0.0690  0.0628  0.0590
##
## sigma^2 estimated as 0.0007698:  log likelihood = 497.82,  aic = -987.65
```

Box-Ljung test

data: ma4.resid X-squared = 2.6452, df = 3, p-value = 0.4496

Box-Ljung test

data: ma4.resid X-squared = 8.0097, df = 7, p-value = 0.3317

Box-Ljung test

data: ma4.resid X-squared = 13.703, df = 11, p-value = 0.2499

After fitting the MA(4) model and estimating its coefficients, the next step involved extracting the residuals to perform model diagnostics and assess its adequacy. The autocorrelation function (ACF) and partial autocorrelation function (PACF) plots of the residuals revealed patterns indicative of white noise, with no discernible autocorrelation present. This observation was further confirmed through the Ljung-Box test
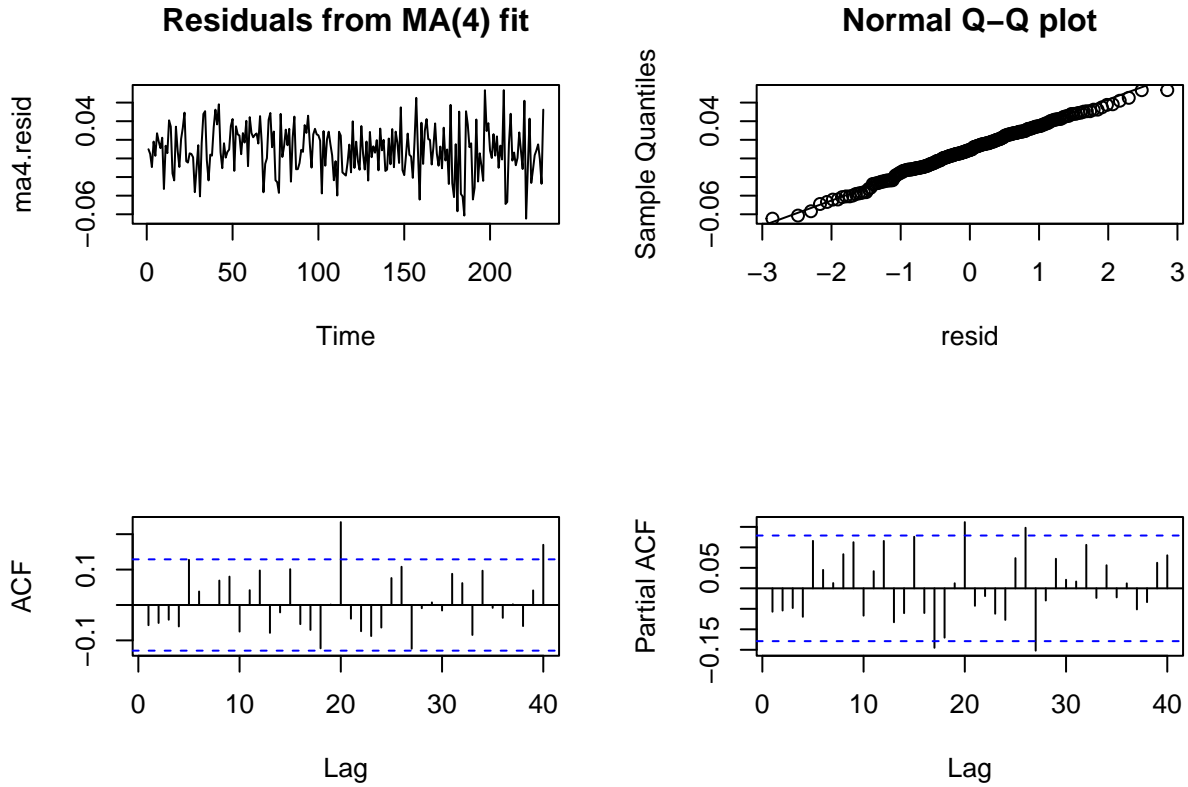
Figure 7: Residual Diagnostic Plots of MA(4)

applied at intervals of 4, 8, and 12, all yielding p-values greater than 0.05, indicating the absence of significant autocorrelation effects in the model.

Furthermore, a normality test using the Shapiro-Wilk test affirmed the normal distribution of the residuals. With the adequacy of the MA(4) model confirmed, predictions were made on the holdout set, with a forecast horizon of 12 time periods. Additionally, a 95% confidence interval was computed for the predictions, providing a measure of uncertainty. The time series plot of the actual values overlaid with the predicted values offered visual insights into the model's predictive performance and its ability to capture the underlying patterns in the data.

Lastly, I evaluated the prediction error on the holdout set using Mean Absolute Error (MAE), Mean Squared Error (MSE), and Mean Absolute Percentage Error (MAPE). This analysis provides insights into the accuracy and reliability of the model's forecasts on unseen data.
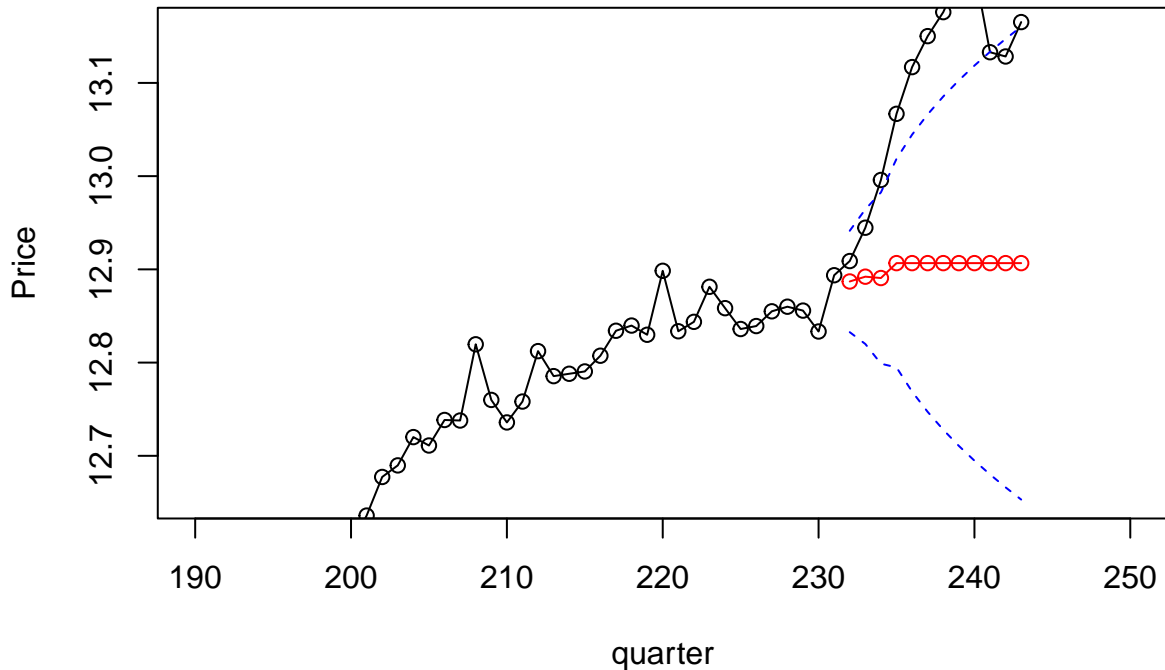
Figure 8: Actual sales prices of houses (solid line) overlaid with MA(4) model predictions (red points) and 95% confidence intervals (blue dashed lines)

AR(4)

After completing the fitting and evaluation of the MA(4) model, I proceeded to explore the adequacy of an AR(4) model. This decision stemmed from observing a significant spike at lag 4 in the partial autocorrelation function (PACF) plot, prompting an investigation into whether an autoregressive model might be suitable. Additionally, considering various model options is essential for selecting the best model for forecasting purposes.

```
##
## Call:
## TSA::arima(x = aspusgr.calib, order = c(4, 1, 0), include.mean = TRUE)
##
## Coefficients:
##           ar1      ar2      ar3      ar4
##       -0.0736   0.1916   0.2088   0.3039
## s.e.   0.0632   0.0616   0.0617   0.0629
##
## sigma^2 estimated as 0.0007363:  log likelihood = 502.9,  aic = -997.8
```

Box-Ljung test

data: ar4.resid X-squared = 6.4272, df = 3, p-value = 0.09258

Box-Ljung test

data: ar4.resid X-squared = 15.044, df = 7, p-value = 0.03544

Box-Ljung test

data: ar4.resid X-squared = 19.477, df = 11, p-value = 0.05305

Following a similar process to that of the MA(4) model, I conducted a thorough analysis of the AR(4) model, which included fitting the model to the data and assessing model diagnostics. During the evaluation phase,
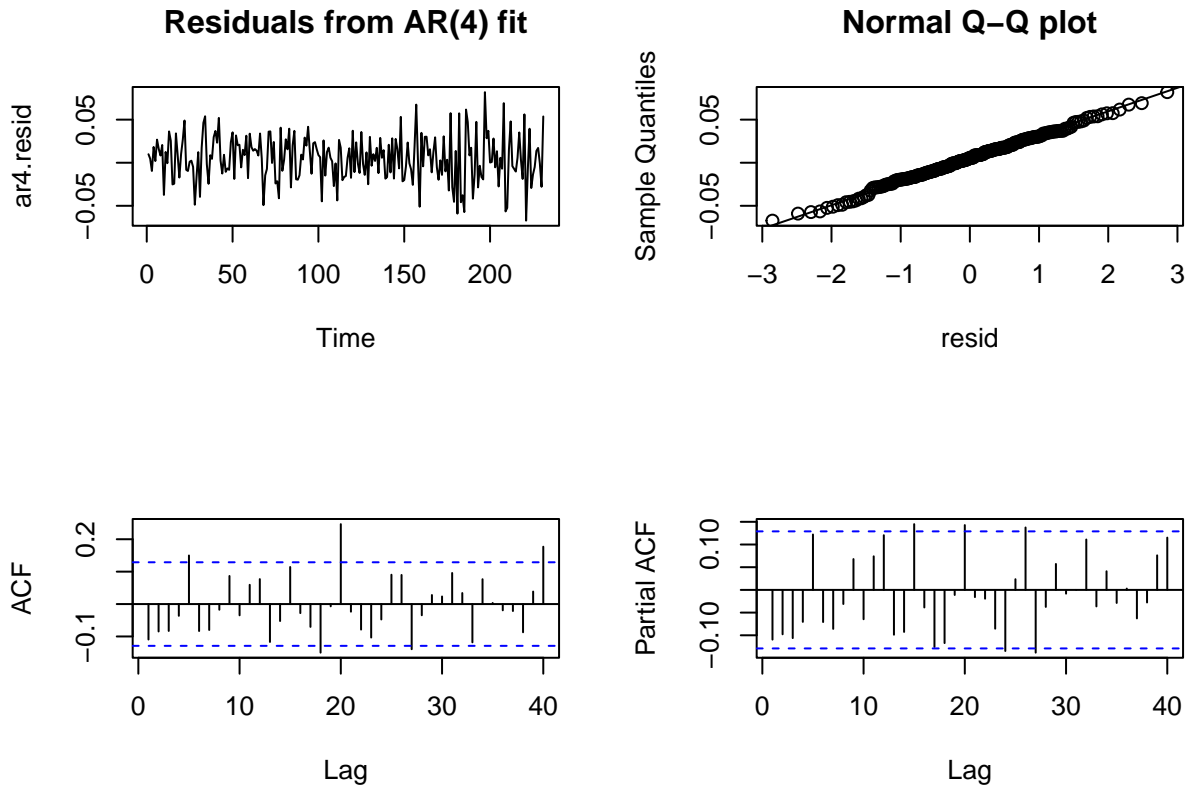
Figure 9: Residual Diagnostic Plots of AR(4)

the model's performance was scrutinized using the same criteria as applied to the MA(4) model. Crucially, the Box-Ljung test performed on the residuals resulted in p-values below 0.05, indicating inadequacy in the AR(4) model. Given this result, I decided not to continue with performance assessments on the test data, as the inadequacy of the model was already confirmed.

Automated Model Selection with auto.arima

Following the assessment of MA(4) and AR(4) models, I transitioned to the auto.arima function from the forecast package. This tool automatically selects the best ARIMA model based on the AIC, a metric that balances model fit and complexity. By leveraging auto.arima, I aimed to streamline model selection and identify the most suitable ARIMA model for the dataset. This approach not only saves time but also ensures that the chosen model is statistically robust and optimally balanced in terms of fit and complexity.

```
## Series: aspusgr.calib
## ARIMA(2,2,1)
##
## Coefficients:
##           ar1      ar2      ma1
##       -0.3080  -0.0626  -0.8334
## s.e.   0.0938   0.0894   0.0699
##
## sigma^2 = 0.0007353:  log likelihood = 501.94
## AIC=-995.89   AICc=-995.71   BIC=-982.15
```
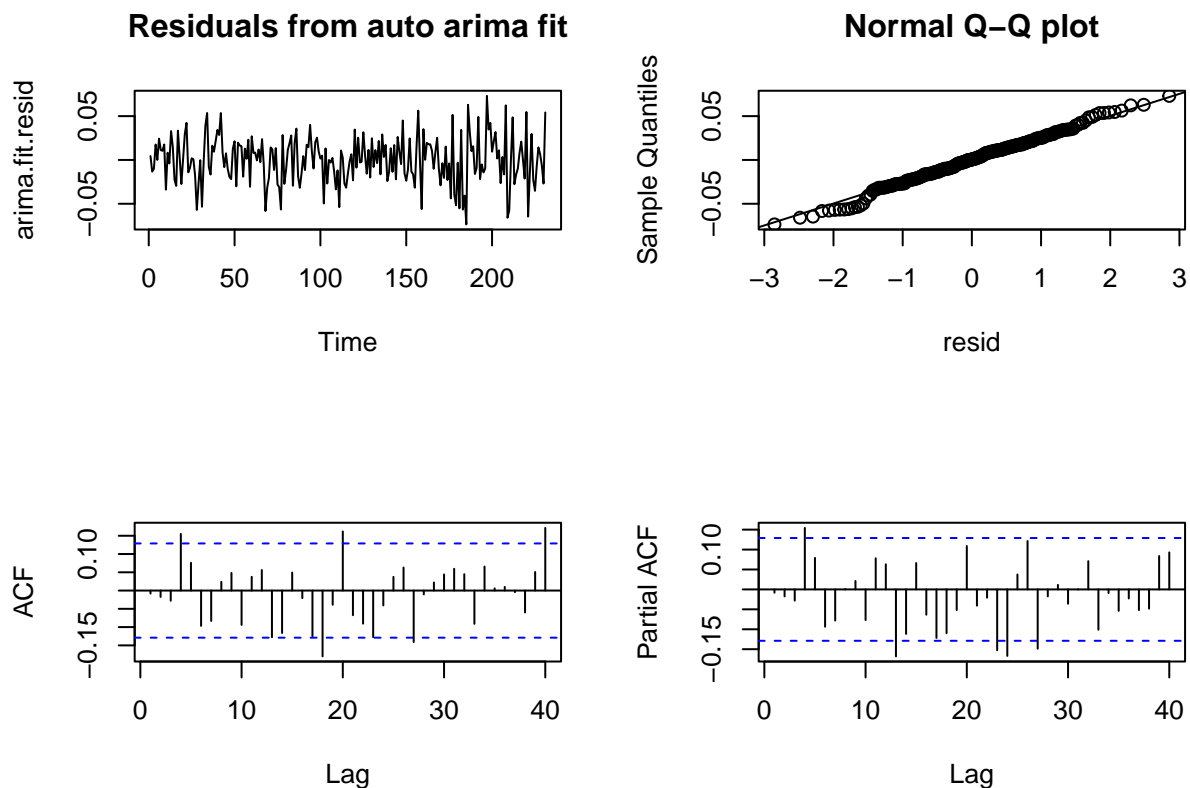

Figure 10: Residual Diagnostic Plots of auto-arima model(2,2,1)

Box-Ljung test

data: arima.fit.resid X-squared = 5.9652, df = 3, p-value = 0.1133

Box-Ljung test

data: arima.fit.resid X-squared = 11.38, df = 7, p-value = 0.1229

Box-Ljung test

data: arima.fit.resid X-squared = 15.219, df = 11, p-value = 0.1727

The ARIMA (2, 2, 1) model underwent diagnostic checks akin to those performed on the preceding models. These diagnostics involved assessing the residuals for autocorrelation using ACF and PACF plots. The absence of significant autocorrelation in the residuals indicated a satisfactory fit of the ARIMA model to the data.
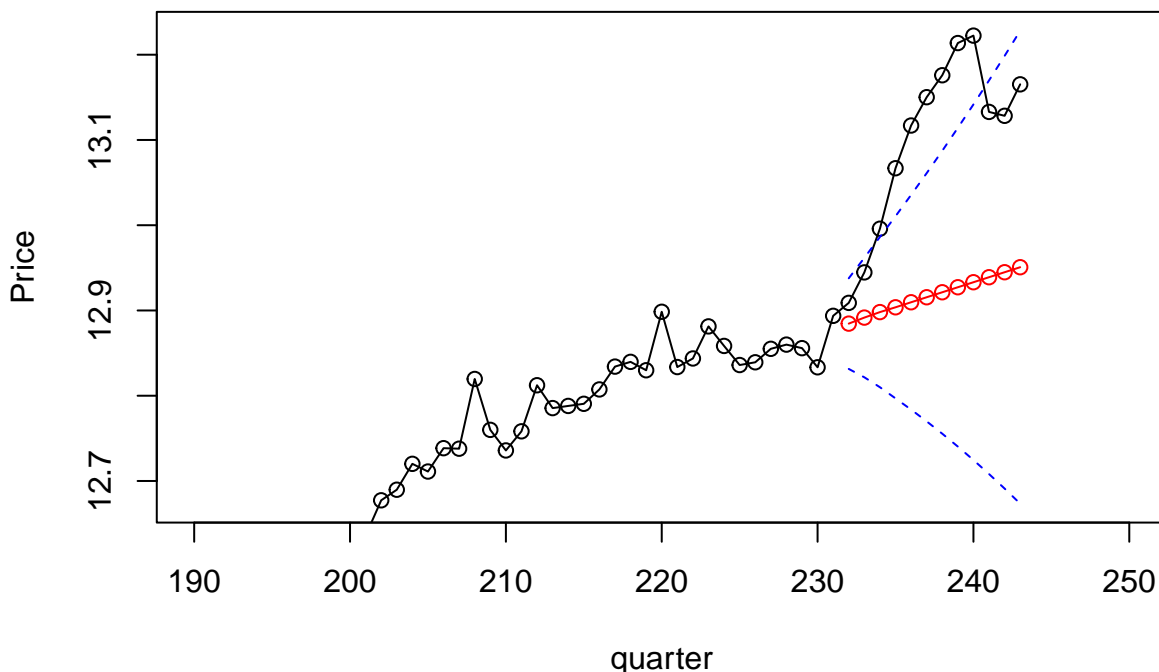

Figure 11: Comparison of actual house prices (black) against ARIMA (2, 2, 1) model predictions (red) with 95% confidence intervals (blue dashed lines) for the holdout test period.

Table of Results for the ARIMA Models

```
## # A tibble: 2 x 4
##   Model              MAE    MSE  MAPE
##   <chr>            <dbl>  <dbl> <dbl>
## 1 MA(4)            0.199 0.0481  1.52
## 2 Auto Arima(2,2,1) 0.184 0.0405  1.40
```

The models are evaluated based on three common metrics: Mean Absolute Error (MAE), Mean Squared Error (MSE), and Mean Absolute Percentage Error (MAPE). The Auto ARIMA(2,1,1) model demonstrates superior performance across all metrics, with a MAE of 0.1835378, MSE of 0.04045227, and MAPE of 1.396267, compared to the MA(4) model which exhibits a MAE of 0.1993138, MSE of 0.04807906, and MAPE of 1.516112. Notably, while an AR(4) model was also considered during the modeling process, it failed to pass the Box-Ljung test, a check for randomness in the residuals, indicating that the model's residuals were not independently distributed. Consequently, the AR(4) model was not utilized for predictions on the test set, leaving the Auto ARIMA(2,2,1) as the preferred model given its robustness and lower error metrics, indicating a closer fit to the observed data.

Including Exogenous Predictor

In an effort to augment the analytical framework for examining the Average Sales Price of Houses Sold in the United States (ASPUS), I expanded our exploration into the realm of ARMAX modeling. By integrating Gross Domestic Product (GDP) as an exogenous variable, the goal is not necessarily to forecast, but rather to investigate the potential of this expanded model to better fit the complex nature of the data. The inclusion

of GDP is particularly aimed at understanding the ramifications of significant economic events, such as the notable increase in ASPUS witnessed during the COVID-19 pandemic in 2020. This addition is expected to shed light on the intricate effects that such macroeconomic factors may exert on the housing market, providing a more holistic view of the variables at play. The use of the ARMAX model thus serves as an investigative tool to determine the viability of incorporating external economic indicators in explaining the variations observed in ASPUS.

```
## $tsp
## [1] 1963.00 2023.75    4.00
##
## $class
## [1] "ts"

## [1] 244
```

Having achieved stationarity in the doubly differenced GDP data, I then examined the sample Autocorrelation Functions (ACFs) of both the doubly differenced ASPUS and GDP. The ACF plots were assessed to visually confirm the behavior indicative of white noise. After the first significant lag, both sample ACFs exhibited the randomness expected of a stationary white noise process, providing a solid foundation for further ARMAX modeling with the inclusion of GDP as an exogenous predictor. This process set the stage for a nuanced analysis that considers the interplay of broader economic forces with the housing market.
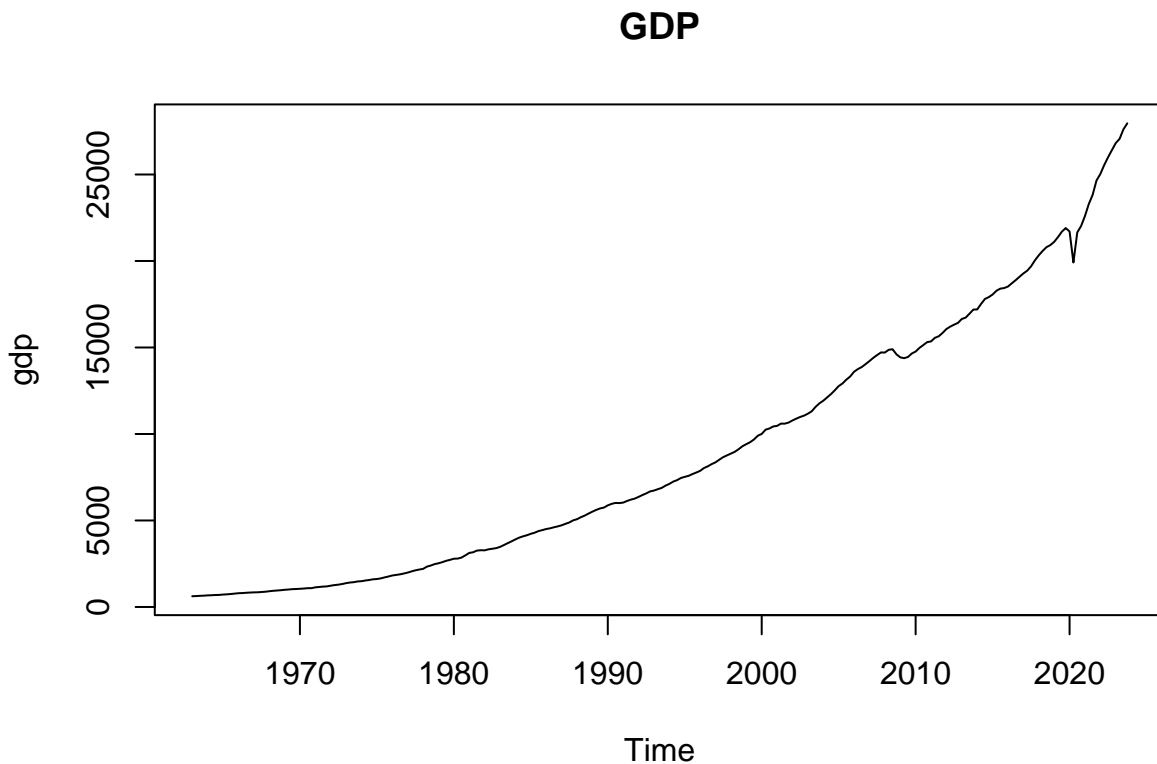
GDP

Figure 12: Time Series Plot of GDP

Quarterly Gross Domestic Product (GDP) data was introduced, measured in billions of dollars, aligning with the timeframe of the Average Sales Price of Houses Sold in the United States (ASPUS). Upon visual examination of the plotted GDP time series, the upward trend and fluctuations in amplitude were as anticipated, characteristic of economic growth over time, particularly marked by the sharp contraction and subsequent recovery around 2020 — likely reflecting the economic impact of the COVID-19 pandemic.

To address nonstationarity, an initial round of differencing was applied to the GDP data. However, an Augmented Dickey-Fuller test suggested that the differenced series was still nonstationary, returning a p-value

of 0.99, far above the typical threshold for stationarity. Consequently, a second round of differencing was performed, which successfully yielded a stationary series.

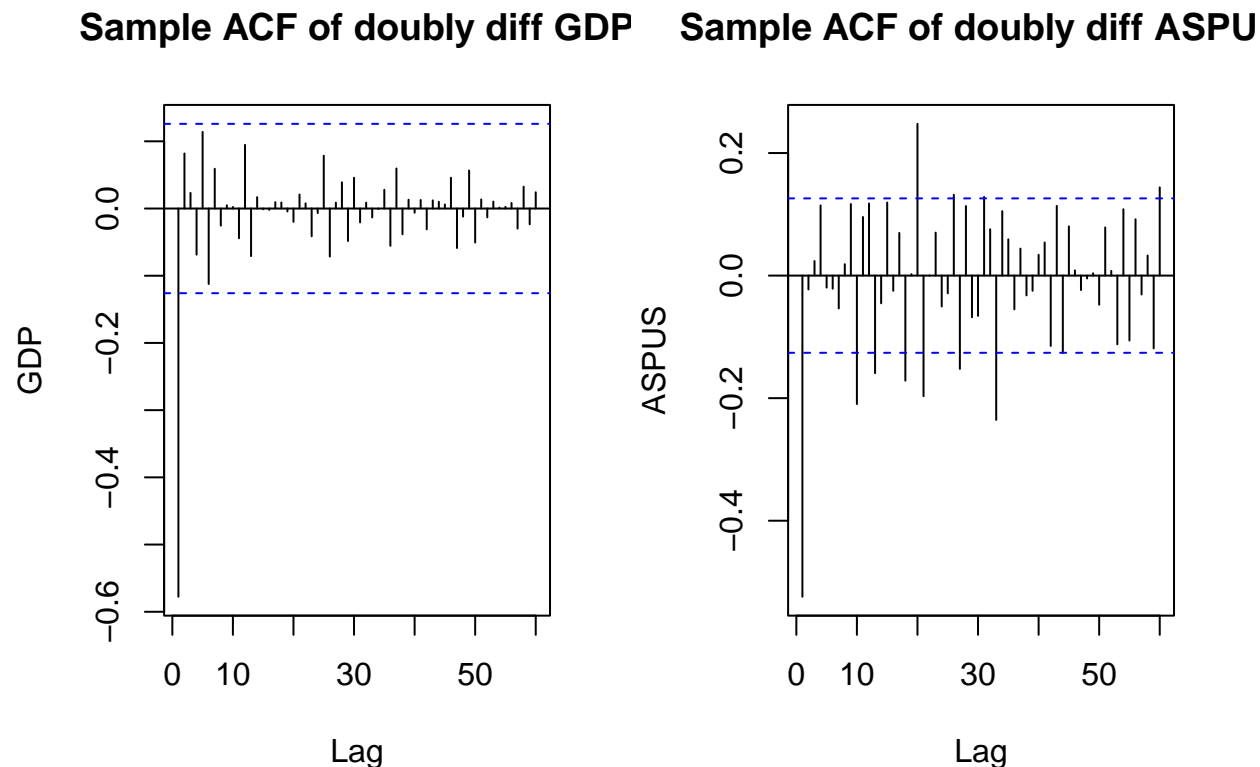**Sample ACF of doubly diff GDP**   **Sample ACF of doubly diff ASPU**

Figure 13: ACF and PACF of doubly differenced GDP and ASPUS

In the analysis, Cross-Correlation Function (CCF) was employed to assess the relationship between GDP and ASPUS over time. The CCF was particularly informative, revealing significant lags that indicate how GDP and ASPUS interact across different quarters. The most significant finding is at lag $h = +3$, where a clear correlation suggests that GDP changes precede changes in ASPUS by three quarters. Correlations at $h = 0$ and $h = -3$ also emerged, indicating a possible immediate relationship and a potential influence of ASPUS on GDP three quarters earlier. These insights from the CCF are crucial for our next steps, particularly in determining how to incorporate GDP as an exogenous variable in an ARMAX model, to better understand the dynamics between these two economic measures.
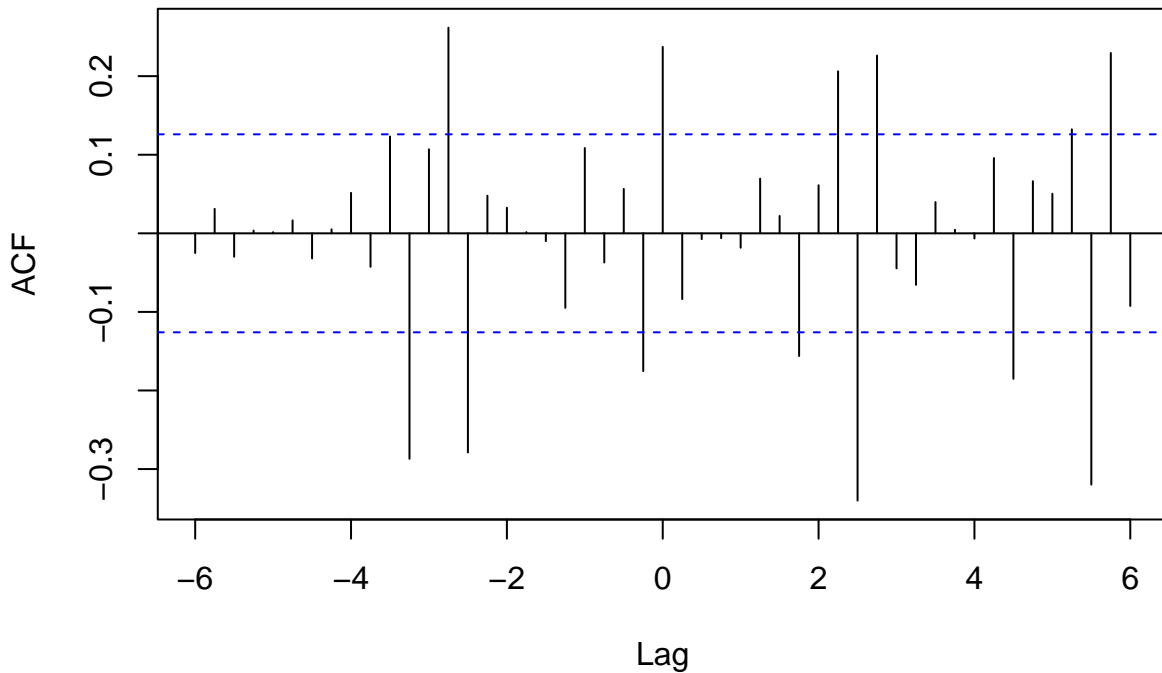
Sample CCF of GDP and ASPUS

Figure 14: Sample CCF plot of GDP and ASPUS

```
##
## Call:
## lm(formula = stable.aspus ~ stable.gdp, data = sales.price)
##
## Residuals:
##    Min    1Q Median    3Q    Max
## -51878  -2945    287   4046  46692
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -136.707    730.502  -0.187 0.851708
## stable.gdp     9.986      2.639   3.783 0.000195 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11360 on 240 degrees of freedom
## Multiple R-squared:  0.05628,    Adjusted R-squared:  0.05235
## F-statistic: 14.31 on 1 and 240 DF,  p-value: 0.0001955
```

When GDP was introduced as a predictor in a linear model at a lag of 3 quarters, the resulting estimate was statistically insignificant. This insignificance was consistent for other lags identified as significant in the CCF, including at -3 quarters.

Contrastingly, when GDP was included as a contemporaneous predictor, the estimate was significant. This indicates that while the CCF suggested potential predictive relationships at various lags, only the contemporaneous relationship between GDP and ASPUS proved to be significant in the context of a linear model. Consequently, the contemporaneous GDP value was retained as a predictor in the linear model, as it

provided a statistically significant explanation for variations in ASPUS.

After fitting a linear model with contemporaneous GDP as a predictor for ASPUS, the residuals were carefully examined for any autocorrelation that might suggest further modeling opportunities. The time series plot of the residuals, complemented by their ACF and PACF, was scrutinized for patterns. The stationarity of these residuals was confirmed via an ADF test, affirming the appropriateness of pursuing an ARMAX model.

Based on the behavior observed in the ACF and PACF, an ARMA(3, 0, 1) was selected to model the differenced ASPUS and GDP data, utilizing the contemporaneous GDP as an exogenous variable. The choice of this particular ARMA model was informed by the lags at which significant autocorrelations and partial autocorrelations appeared, suggesting that this model would adequately capture the dynamics present in the time series data.
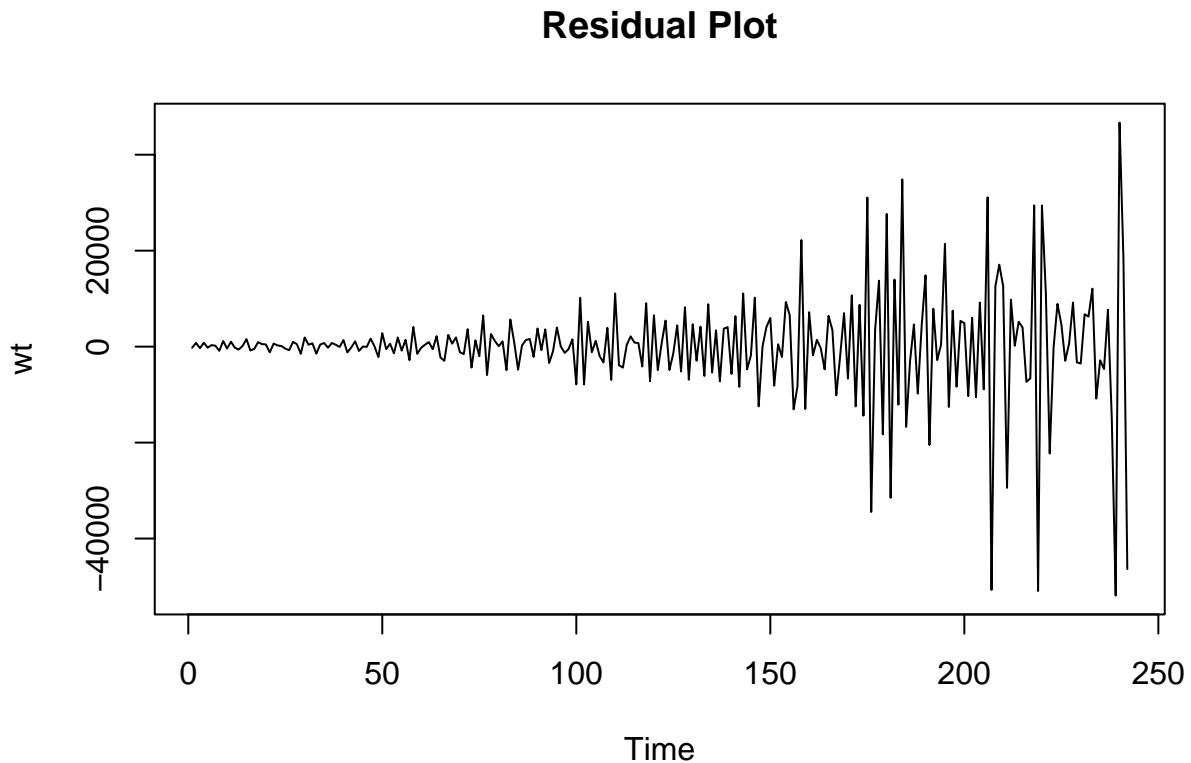
Residual Plot

Figure 15: Residual time series plot of wt

ACF of Regression Residuals
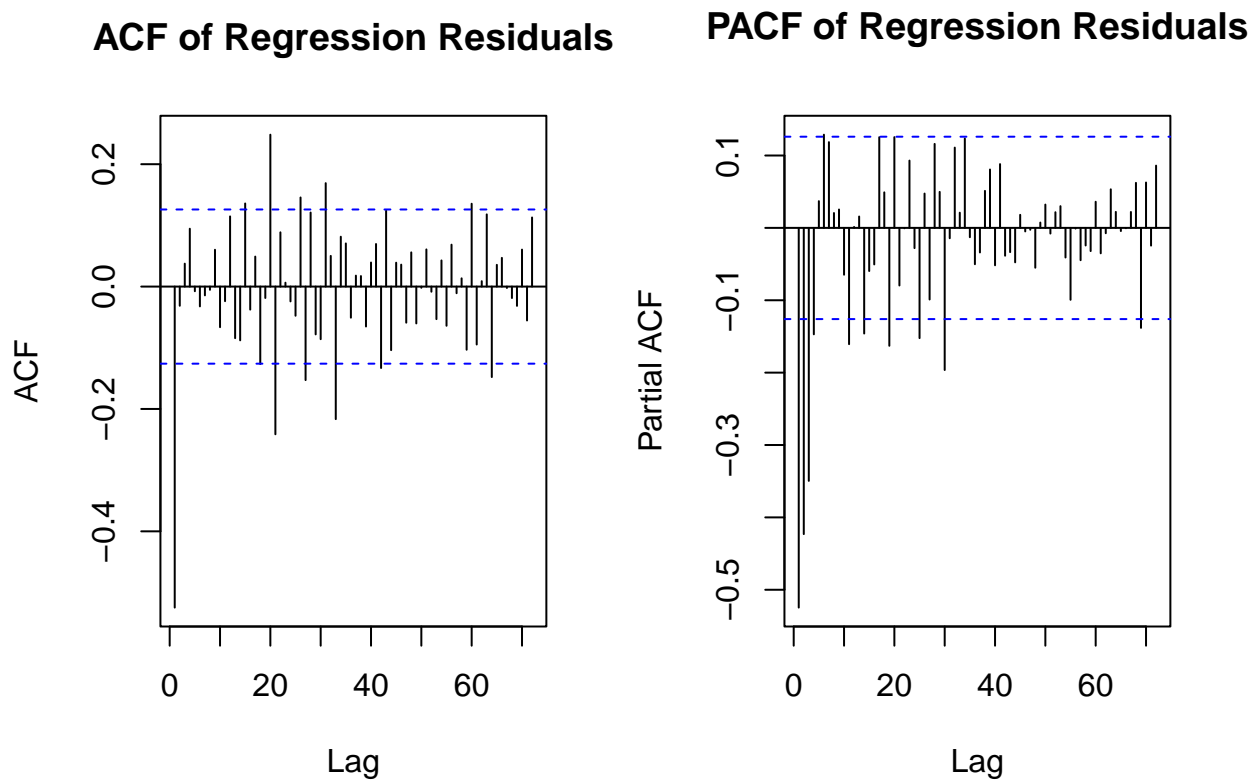
PACF of Regression Residuals


Figure 16: ACF and PACF of wt

ARMAX fit result

```
## initial  value 9.340263
## iter   2 value 9.032501
## iter   3 value 9.009284
## iter   4 value 8.998532
## iter   5 value 8.989335
## iter   6 value 8.973655
## iter   7 value 8.972657
## iter   8 value 8.971891
## iter   9 value 8.970879
## iter  10 value 8.969853
## iter  11 value 8.968218
## iter  12 value 8.966824
## iter  13 value 8.966127
## iter  14 value 8.966103
## iter  15 value 8.966102
## iter  16 value 8.966100
## iter  17 value 8.966100
## iter  17 value 8.966100
## iter  17 value 8.966100
## final  value 8.966100
## converged
## initial  value 8.962749
## iter   2 value 8.962738
## iter   3 value 8.962736
## iter   4 value 8.962732
```

```
## iter   5 value 8.962731
## iter   6 value 8.962731
## iter   7 value 8.962731
## iter   7 value 8.962731
## iter   7 value 8.962731
## final  value 8.962731
## converged
## <><><><><><><><><><><><><><>
##
## Coefficients:
##            Estimate       SE t.value p.value
## ar1         -0.6824   0.1414 -4.8278  0.0000
## ar2         -0.5421   0.1363 -3.9769  0.0001
## ar3         -0.2503   0.1046 -2.3940  0.0174
## ma1         -0.3377   0.1449 -2.3302  0.0206
## intercept -60.4597 134.8686 -0.4483  0.6544
## xreg        12.6298   2.6075  4.8436  0.0000
##
## sigma^2 estimated as 60598212 on 236 degrees of freedom
##
## AIC = 20.82119  AICc = 20.82267  BIC = 20.92211
##
```
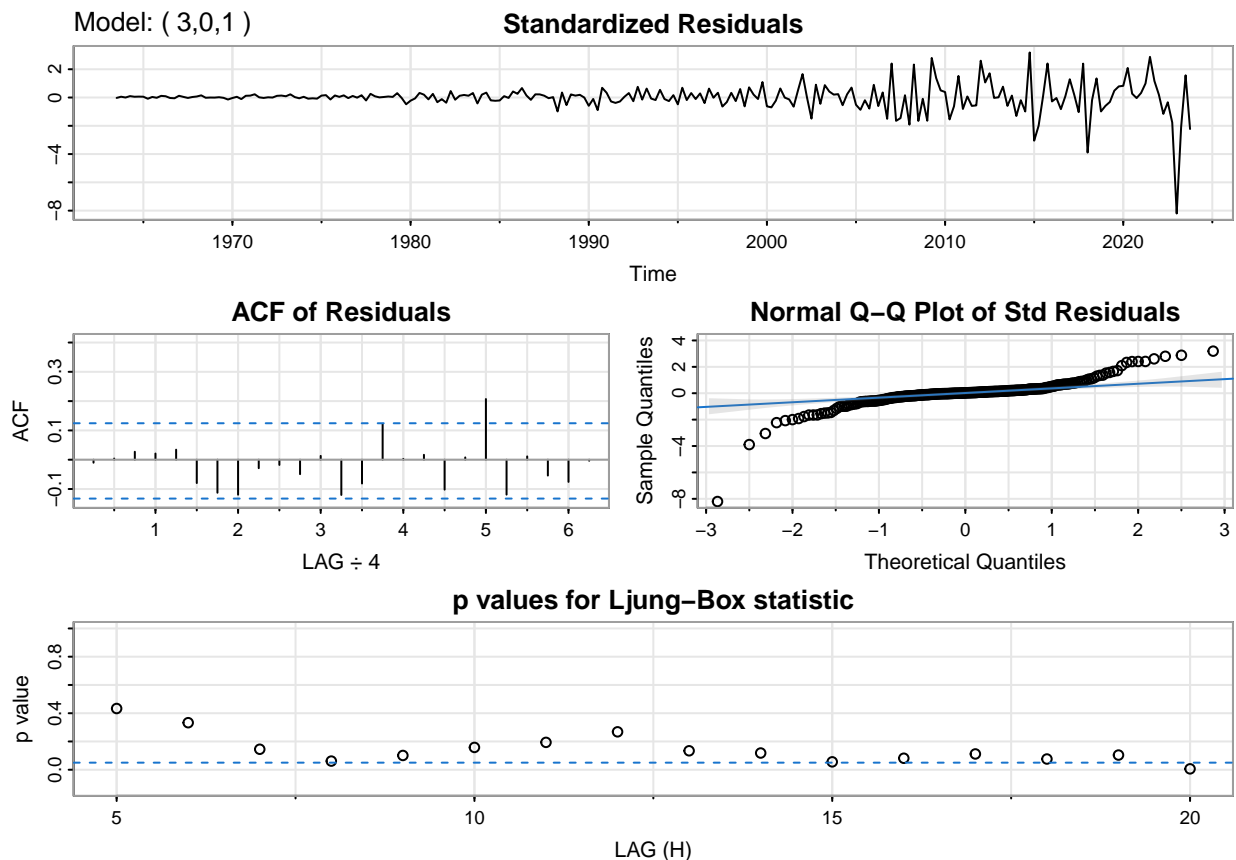

An ARMAX model was successfully fitted to the data using the SARIMA function, which integrates the significant contemporaneous relationship between GDP and ASPUS. The model's estimates indicate that all AR and MA coefficients are significant at conventional levels, with p-values well below the threshold of 0.05. Notably, the inclusion of contemporaneous GDP as an exogenous variable (xreg) in the ARMAX model

yielded a significant estimate, emphasizing its influence on ASPUS within the modeled period.

Model diagnostics were conducted to evaluate the goodness of fit. The standardized residuals over time do not display any obvious patterns or trends, suggesting that the model captures the data's underlying structure effectively. The ACF of residuals, which ideally should show no significant correlations for successive lags, appears to confirm this, with most lags within the confidence bounds. Moreover, the Normal Q-Q plot of standardized residuals adheres closely to the line, suggesting that the residuals are normally distributed

Lastly, the p-values for the Ljung-Box statistic, are above the typical significance level. This implies that the residuals can be considered random 'white noise', which is a desirable property indicating that the model has successfully captured the information in the data.

Overall, the ARMAX model demonstrates satisfactory statistical properties, indicating that it is well-specified and that contemporaneous GDP is a significant predictor of ASPUS within this modeling framework.

Limitations of Analysis

The analysis of the ARMAX model's capacity to predict the Average Sales Price of Houses Sold in the United States, while incorporating GDP as a contemporaneous variable, encounters several methodological constraints. Chief among these is the limitation arising from model validation—or the lack thereof. The model's predictive accuracy was not tested against an unseen test set, a step that is critical for assessing forecast reliability and for benchmarking against other models, such as previously fitted ARIMA models. Furthermore, the deviations in the tails of the Q-Q plot raise concerns about the normality of residuals, suggesting that extreme values could be more prevalent than the normal distribution would predict, thus potentially affecting the robustness of statistical conclusions. The analysis also potentially overlooks the multifaceted economic factors and structural breaks, like the economic impact of COVID-19, which may not be entirely explained by the variables and the linear approach adopted. The assumption of normality and the decision to not validate the model with out-of-sample testing reflect gaps that must be carefully weighed when interpreting the results and before applying the model for real-world prediction or analysis. Future iterations of this work would benefit from addressing these issues to strengthen the analytical framework and the subsequent insights drawn from it.

Conclusion

This study has undertaken a detailed examination of housing price trends in the United States from 1963 to 2023, utilizing a range of statistical models to decipher the underlying patterns. The Auto ARIMA(2,2,1) emerged as particularly effective, demonstrating superior performance over the MA(4) model in terms of prediction accuracy. Expanding upon these insights, the introduction of an ARMAX model, which incorporated contemporaneous GDP, marked a significant advancement in the analytical approach, revealing how immediate economic conditions impact housing prices.

The inclusion of GDP in the ARMAX model highlights its value as a predictor, emphasizing the dynamic relationship between economic activity and the housing market. This model enriches our understanding by illustrating the direct effects of economic changes on housing prices, thus providing a more holistic view of the economic factors at play.

The findings from this study enhance the existing body of knowledge on housing market dynamics and offer valuable insights for stakeholders in making informed decisions in real estate investment and policy formulation. Future efforts in this research area may build on this foundation, exploring further the complex interactions between the economy and housing markets.

Reference

U.S. Census Bureau and U.S. Department of Housing and Urban Development, Average Sales Price of Houses Sold for the United States [ASPUS], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/ASPUS, March 27, 2024.

U.S. Bureau of Economic Analysis, Gross Domestic Product [GDP], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/GDP, April 26, 2024.