Understanding Housing Market Trends and Risks : An Analytical Study

Kunal Bhardwaj[1]

[1] UIUC

Author Note

STAT 429: Time Series Analysis

The authors made the following contributions. Kunal Bhardwaj: Conceptualization, Data Curation, Visualization, Writing - Original Draft Preparation, Writing - Review & Editing.

Abstract

The objective of this study is to arrive at a model to predict the median sales price of houses sold in the US. The study tries to find the most significant predictors of price based on a regression analysis. A further ARCH-based analysis will try to estimate the volatility in house prices. The study finds that 'Median Sales Price of Houses Sold' and '30 Year Fixed Rate Mortgage Average' are significant predictors of housing prices. The forecasting performance for ARIMA and ARCH models were found to be fairly similar.

The study found that homeownership rate is not a significant predictor of median housing prices. It was also found that ARCH models failed to capture volatility patterns with a high degree of satisfaction. This may be due to the fact that the US housing market has been historically been very stable and immune to price shocks in the past. The results from all the selected models reveal that median house prices will remain stable or will go down slightly over the next five quarters.

Understanding Housing Market Trends and Risks : An Analytical Study

Introduction

Federal Reserve Economic Data (FRED) is a comprehensive database maintained by the Federal Reserve Bank of St. Louis. It provides access to a wide range of economic data, including economic indicators, financial and banking data, monetary data, and regional data for the United States. FRED aggregates data from various government agencies, international organizations, and other sources, making it a valuable resource for researchers, economists, policymakers, and the general public.

The dataset retrieved from *FRED* website comprises of six time series:

i) Median Sales Price of Houses Sold for the United States

ii) Monthly supply of New Houses in the United States

iii) New Privately owned Housing units started

iv) Home ownership rate in the United States

v) 30 Year Fixed Rate Mortgage Average in the United States

vi) Consumer Price Index (CPI) for All Urban Consumers: All Items Less Shelter in U.S. City Average

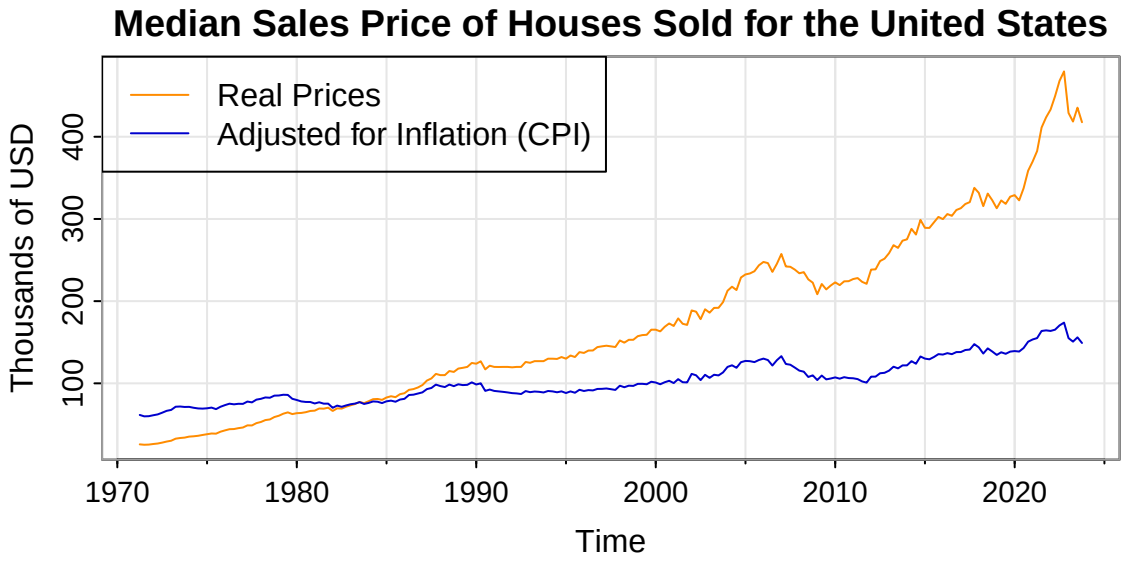
The objective of the project will be to predict Median Price *(i)* based on four other factors *(ii), (iii), (iv) & (v)*. The CPI data *(vi)* will not be used as a predictor but will be used to adjust Median Price based on inflation.

These four variables are fundamental drivers influencing the supply and demand dynamics of the housing market. For instance, the monthly supply of new houses and new housing units started *(ii)* offer insights into the supply side of the market, while the home

ownership rate *(iv)* reflects the demand for housing. Moreover, the 30 year fixed rate mortgage average *(v)* directly impacts affordability and purchasing power, crucial factors influencing housing demand. The structure of the data is outlined below.

Median Sales Price of Houses Sold for the United States (MSPUS) [Quarterly] [Q1'63 Q4'23] *OUTCOME VARIABLE*

The Median Sales Price of Houses Sold for the United States in US Dollars. The original data has not been seasonally adjusted.

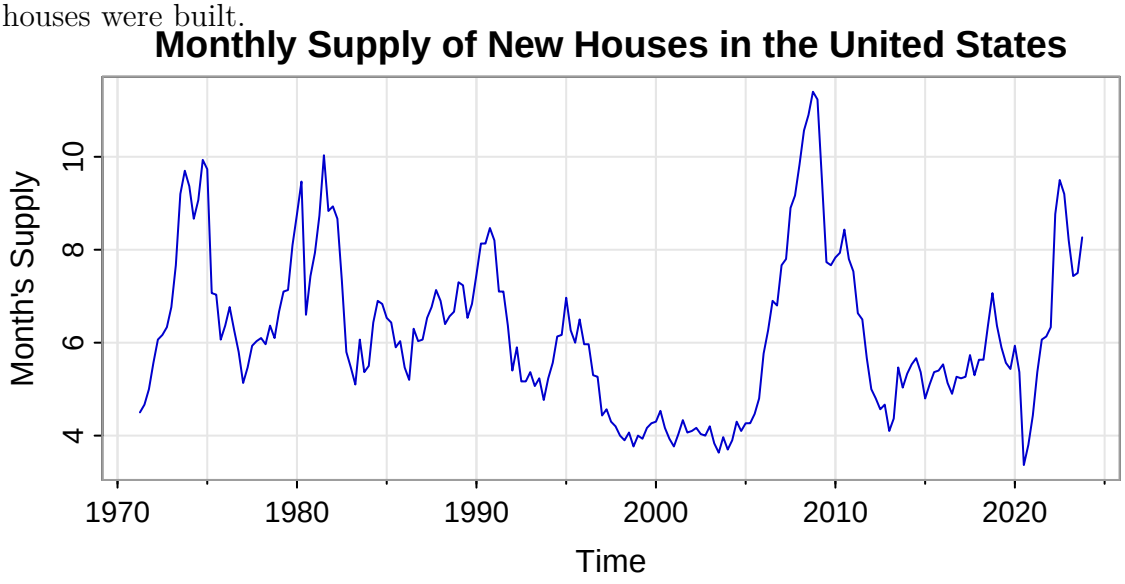


From the plot of Median Sales Price of Houses Sold, we can see that there exists an obvious trend in the data. The Median Prices will be adjusted for inflation based on 1982:1984 prices to make median housing prices from different years directly comparable. This is crucial for understanding long term trends in housing prices and assessing changes in affordability over time.

Moreover, adjusting median housing prices for inflation should lead to more accurate forecasts by avoid biases introduced by price shocks in commodity prices, the effect of which have not been included in the model.

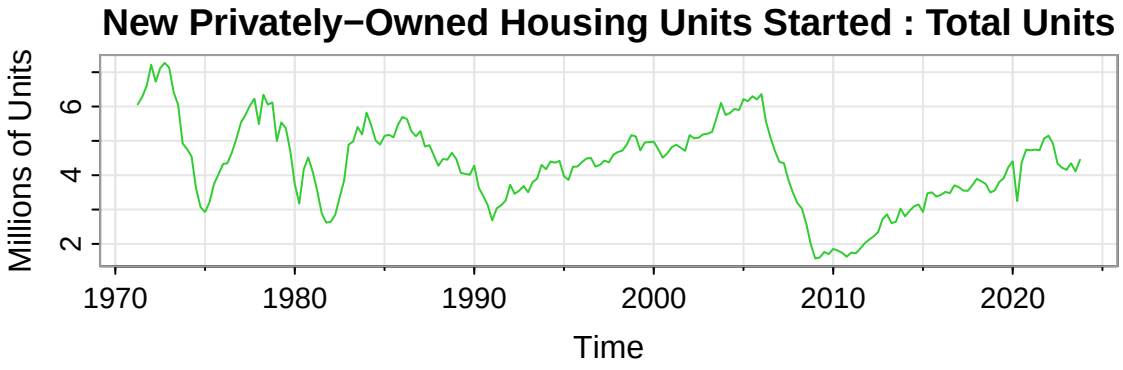
Monthly Supply of New Houses in the United States (MSACSR) [Monthly] [Jan'63 - Dec'23] ***Predictor variable 1***

The months' supply is the ratio of new houses for sale to new houses sold. This statistic provides an indication of the size of the new for-sale inventory in relation to the number of new houses currently being sold. The months' supply indicates how long the current new for-sale inventory would last given the current sales rate if no additional new houses were built.



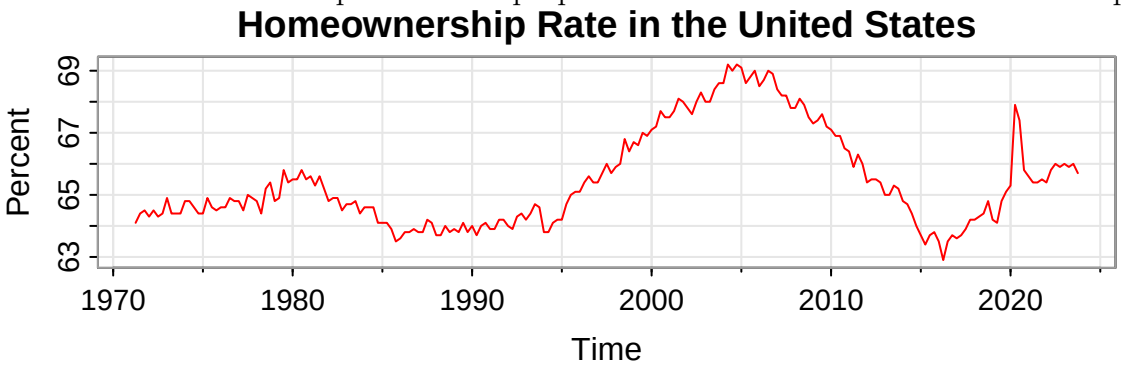
New Privately-Owned Housing Units Started: Total Units (HOUST) [Monthly] [Jan'59 - Jan'24] ***Predictor variable 2***

As provided by the Census, start occurs when excavation begins for the footings or foundation of a building. Increases in housing starts and permits indicate growing supply, which can help alleviate housing shortages and moderate price growth. Conversely, declines in construction activity may contribute to supply constraints and upward pressure on prices.

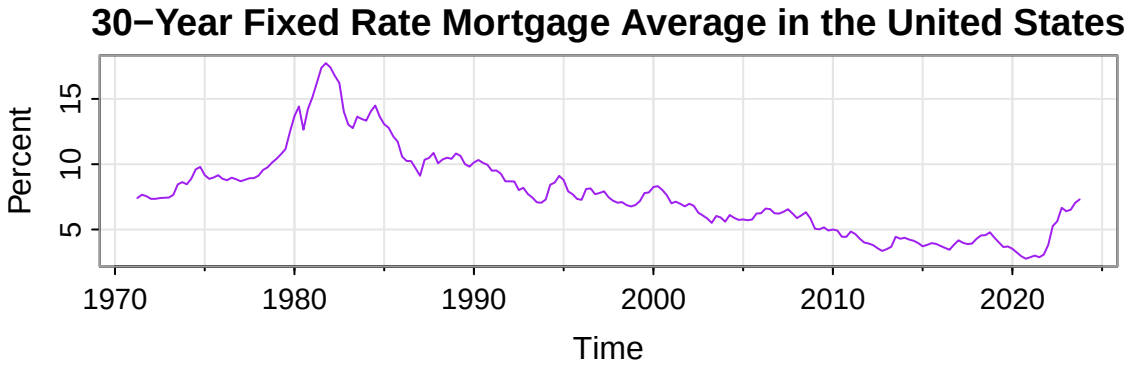


Homeownership Rate in the United States (RHORUSQ156N) [Quarterly] [Q1 '65 - Q4'23] ***Predictor variable 3***

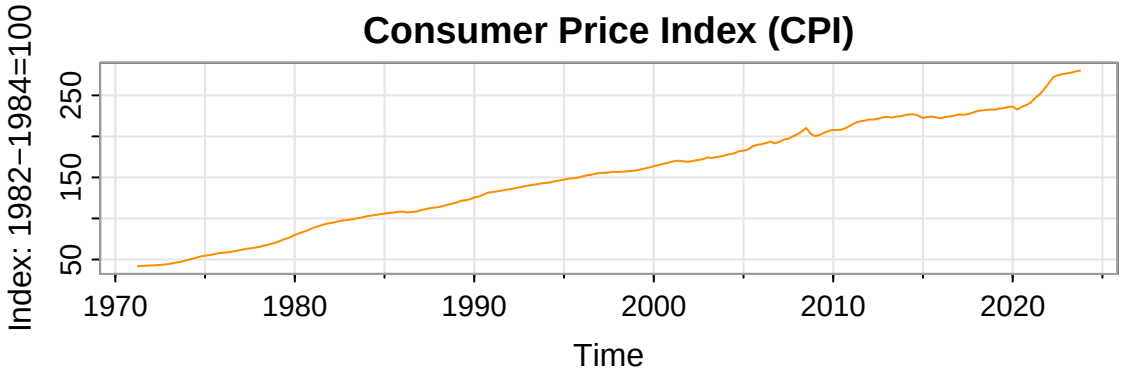
The homeownership rate is the proportion of households that is owner-occupied.



30-Year Fixed Rate Mortgage Average in the United States (MORTGAGE30US) [Weekly] [Apr'71 - Feb'24] ***Predictor variable 4***



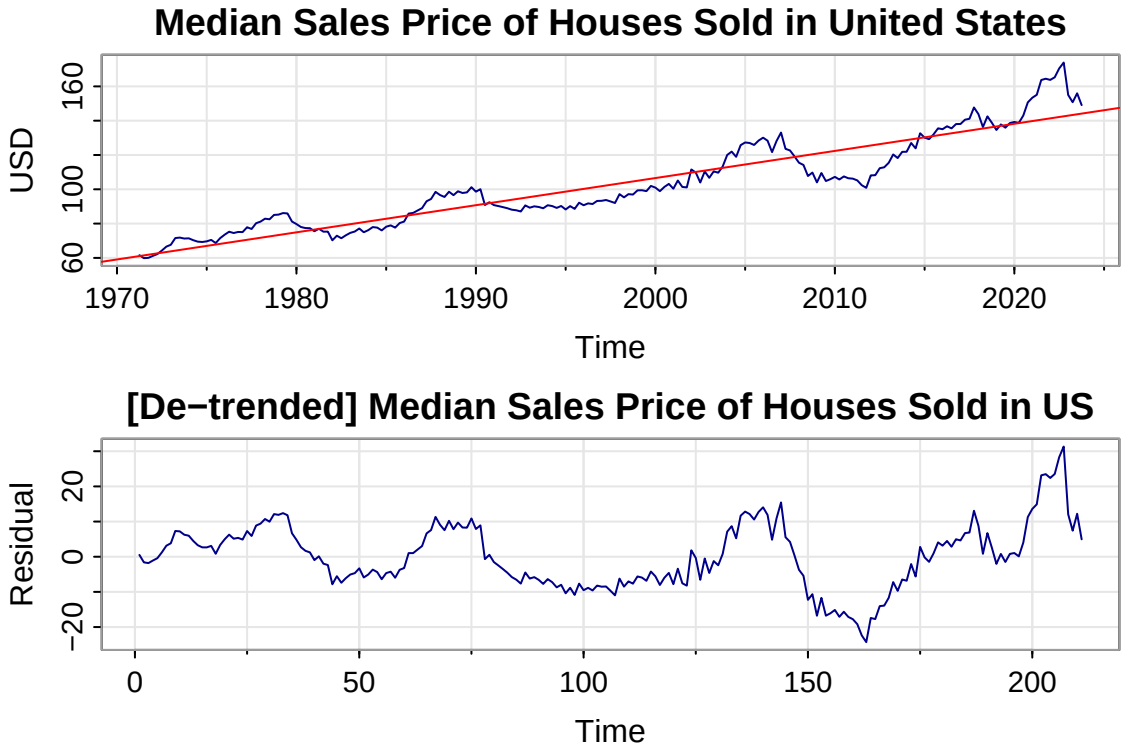
Consumer Price Index for All Urban Consumers: All Items Less Shelter in U.S. City Average (CUSR0000SA0L2) [Monthly] [Jan'47 - Jan'24] ***Additional Variable [NOT to be used as a predictor]***



Plans for Analysis A

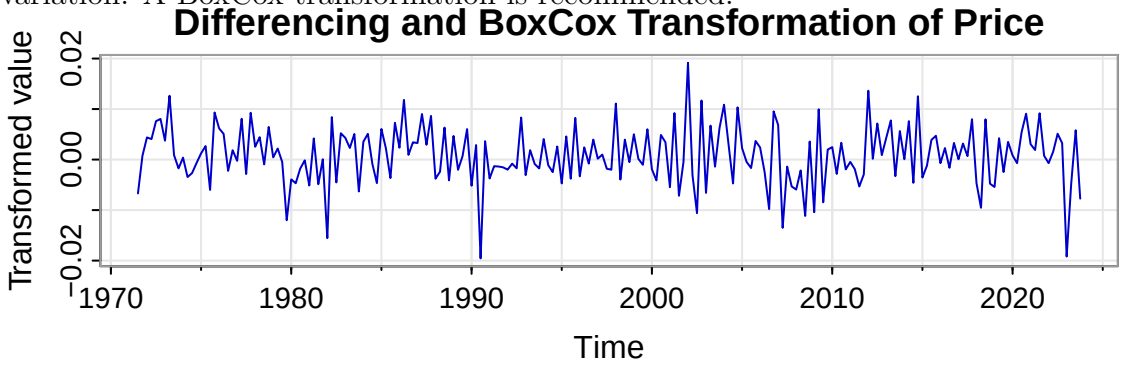
The time series data of median price [MSPUS] will be regressed on time *(t)* and the four other independent variables. The Median Prices (outcome variable) will be pre processed by adjusting for inflation, followed by de-trending and log transformation to make it stationary and homoscedastic. The four predictor variables will be converted into quarterly series if they are not already. The Median Price will be regressed on time and other predictors to arrive at the full model. Multiple model sizes will be analyzed to find the optimum model to be selected, based on AIC/BIC criteria. Based on the p/ACF of the residuals, we may conduct regression with autocorrelated errors.

Methods



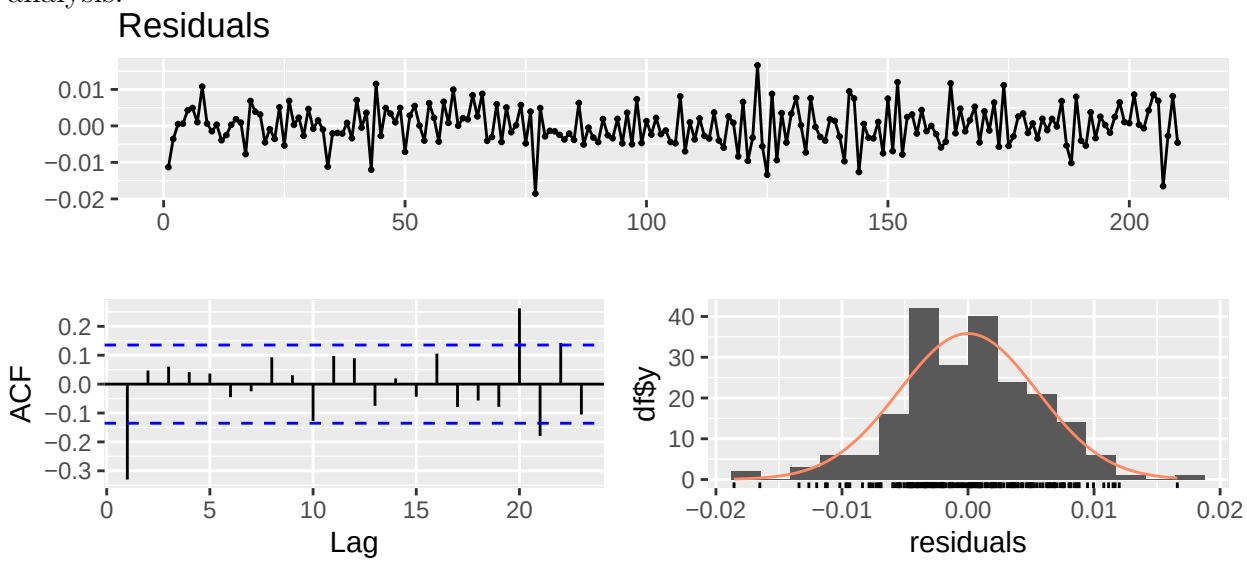
Preliminary adjustments indicate that median prices can be make stationary before regression with some adjustment for the heteroscastic behavior present in the series.

ADF & KPSS tests conclude stationarity. But the series exhibits an obvious heteroscedasticity (evidenced by BP-Test), where higher levels are associated with higher variation. A BoxCox transformation is recommended.

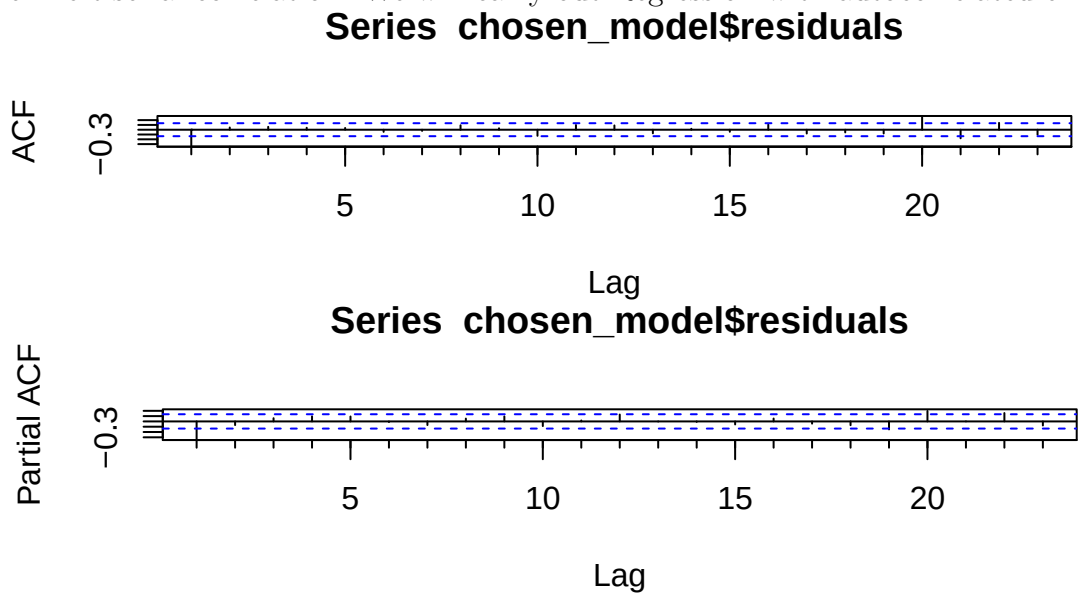


The differenced and BoxCox transformed series passes tests of stationarity and is homoscedastic. We can now proceed with regression. We will use the step-wise algorithm to find the optimum model.

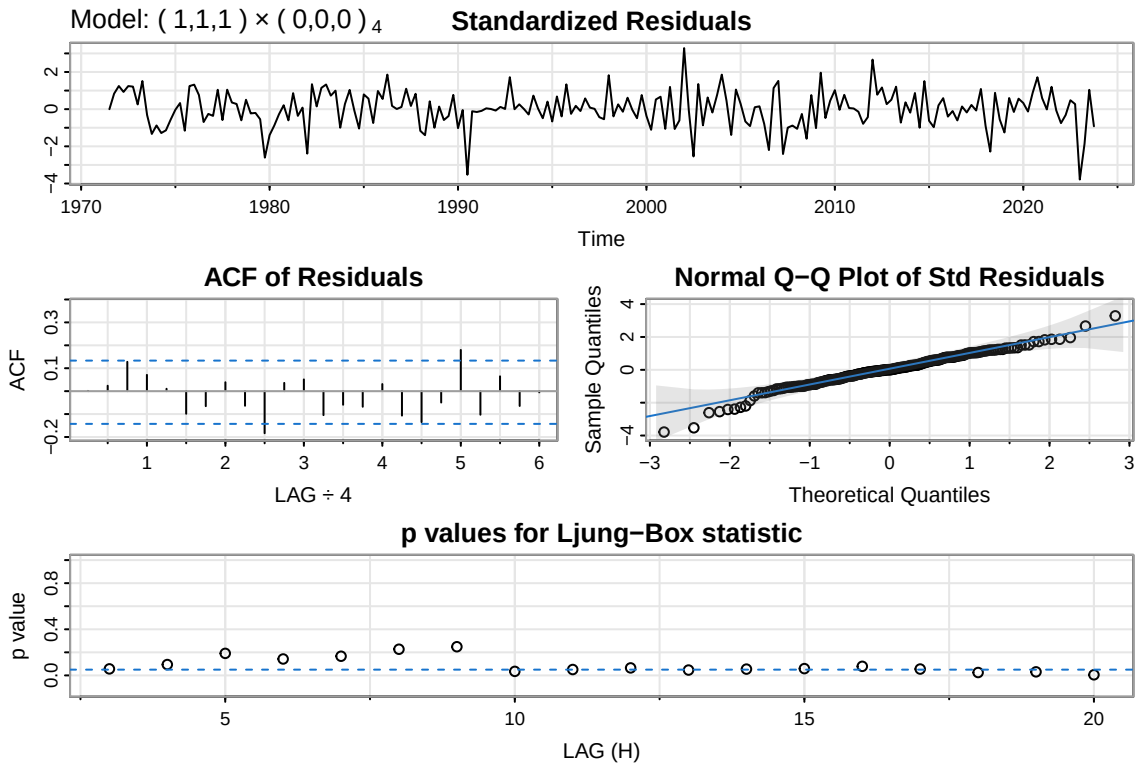
The stepwise algorithm concludes that Predictor 1 (MSACSR) and Predictor 4 (MORTGAGE30US) are significant predictors of housing prices based on AIC criteria. **Refer appendix for full results of the algorithm.** We will now carry out residual analysis:



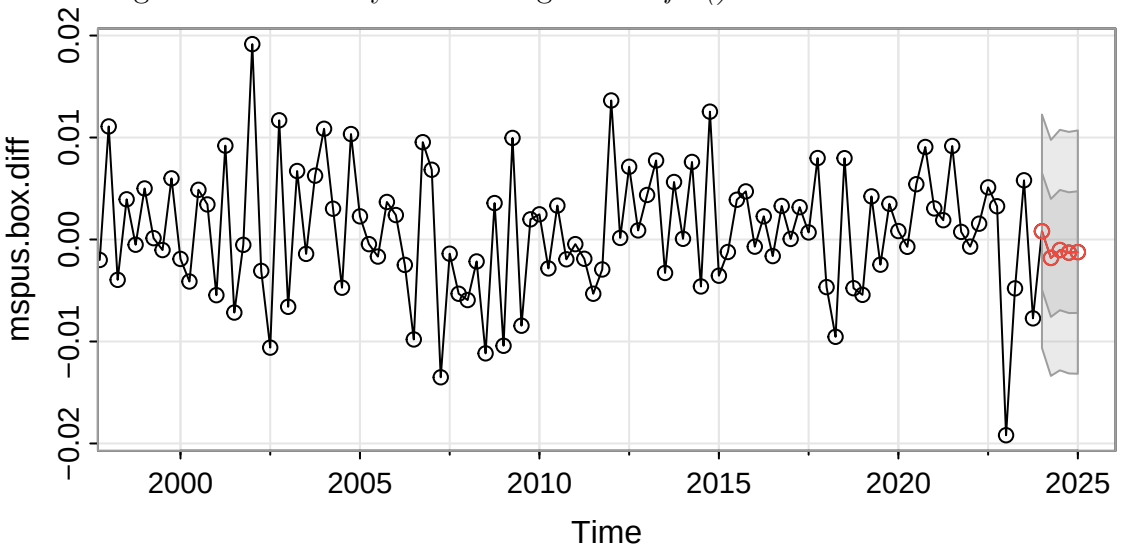
The Ljung-Box test concludes that the residuals are not independently distributed; they exhibit serial correlation. We will carry out Regression with autocorrelated errors.



ACF cuts off after 1. PACF cuts off after 1. ARMA(1,1) looks like a good fit for the residuals. We will fit a ARMA(1,1) model and carry out forecasting.



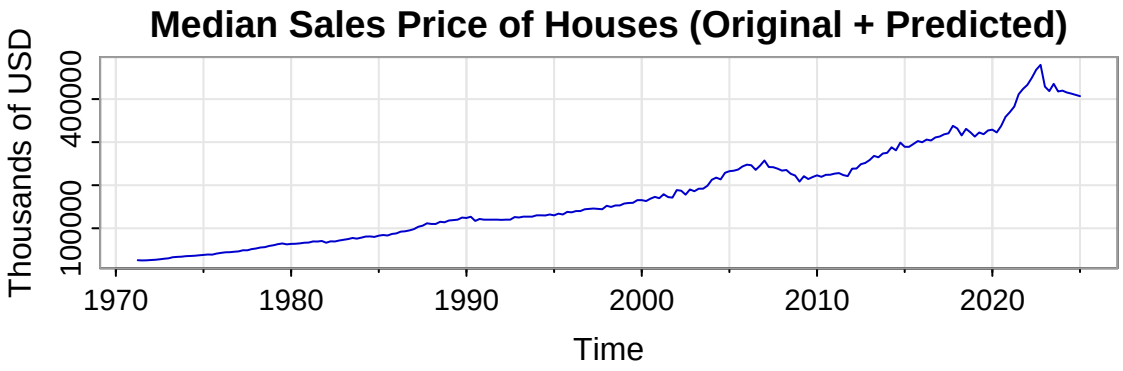
The ACF of residuals show that they now resemble white noise. We will carry out the forecasting of the stationary series using *sarima.for()* function.




```
## [1] "Forecasted Values: "
```

```
##          Qtr1     Qtr2     Qtr3     Qtr4
## 2024 419656.1 415300.8 412832.8 409781.8
```

```
## 2025 406890.7
```

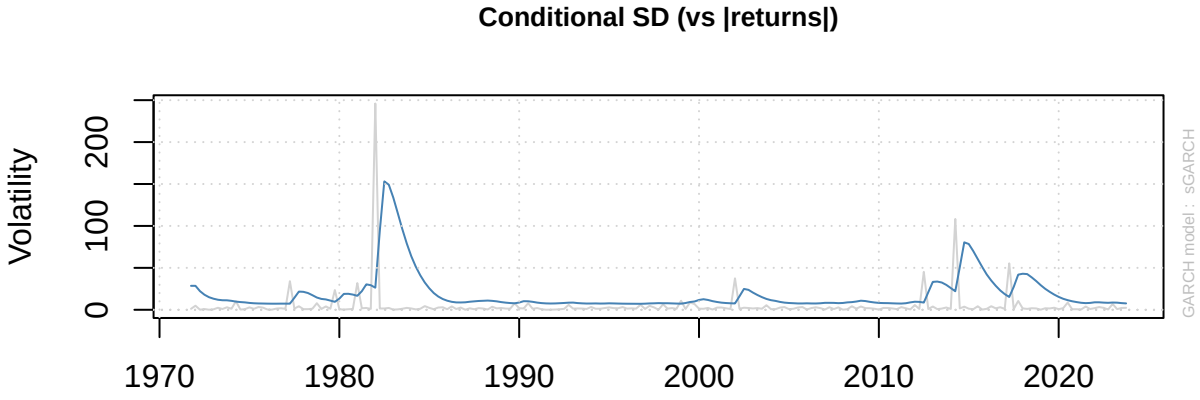


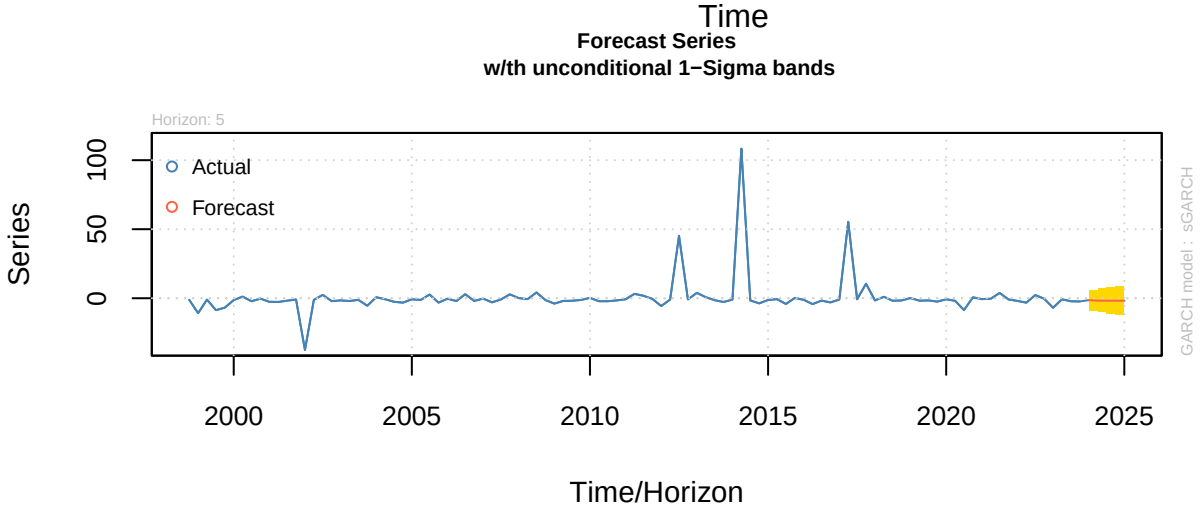
Analysis C

The plot of squared residuals from the ARIMA fit shows the residuals have conditional heteroscedasticity. An ARCH/GARCH model should work well to model this. We will fit three ARCH models: GARCH, IGARCH & APARCH.

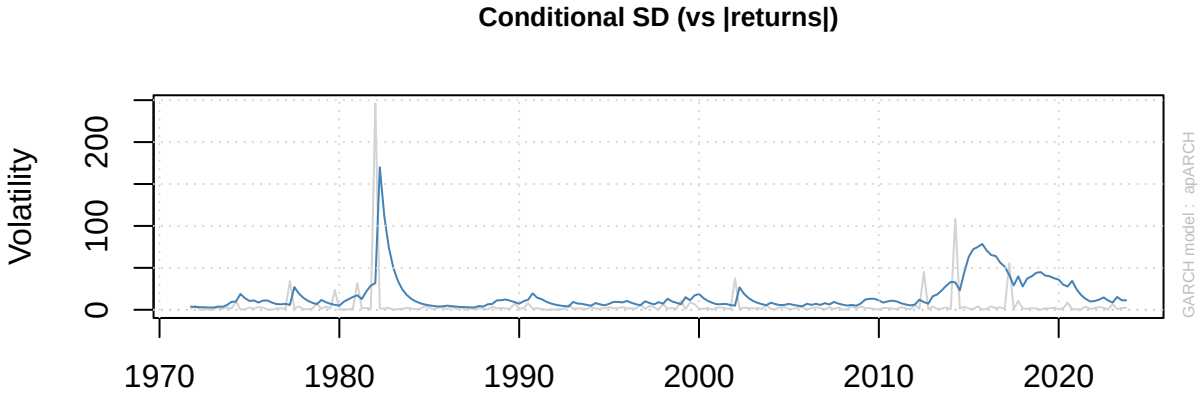
GARCH models assume stationarity (Shumway & Stoffer, 2000). Therefore we will use the log-differenced series for GARCH model which is stationary. We need to evaluate the mean model ARMA Order and GARCH order simultaneously to fit a ARMA-GARCH model.The mean model uses ARMA that generates the forecast for the mean of the time series, while the GARCH model generates the forecast for the variance. The mean model will use an ARMA(1,1) and we will use different GARCH orders to find the best ICs to evaluate goodness of fit.

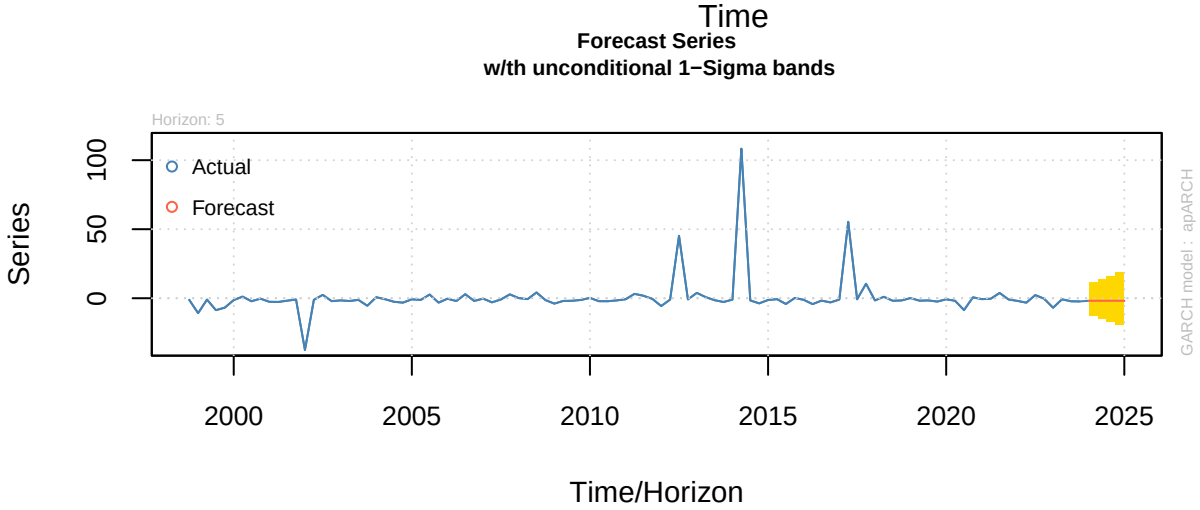
GARCH


Conditional SD (vs |returns|)
Volatility
0
100
200
1970
1980
1990
2000
2010
2020
Time
GARCH model : sGARCH

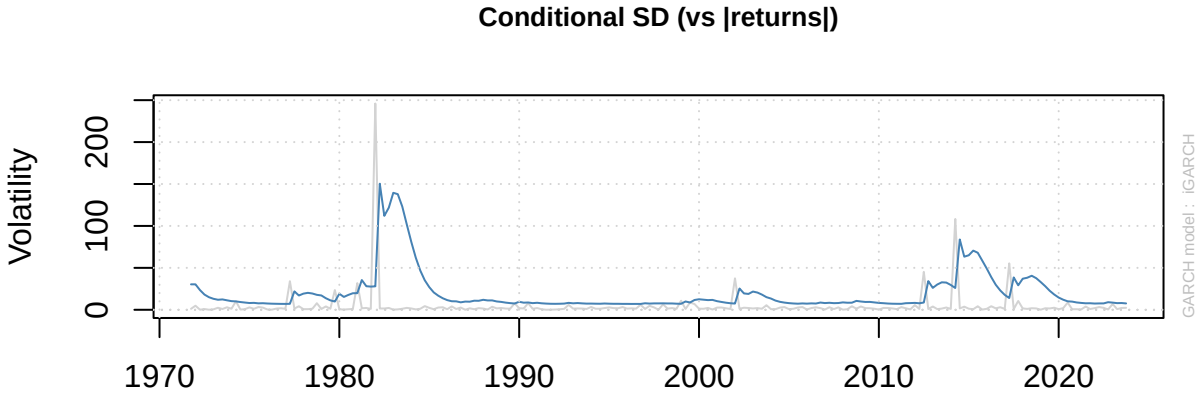

Forecast Series
w/th unconditional 1−Sigma bands
Horizon: 5
Actual
Forecast
Series
0
50
100
2000
2005
2010
2015
2020
2025
Time/Horizon
GARCH model : sGARCH

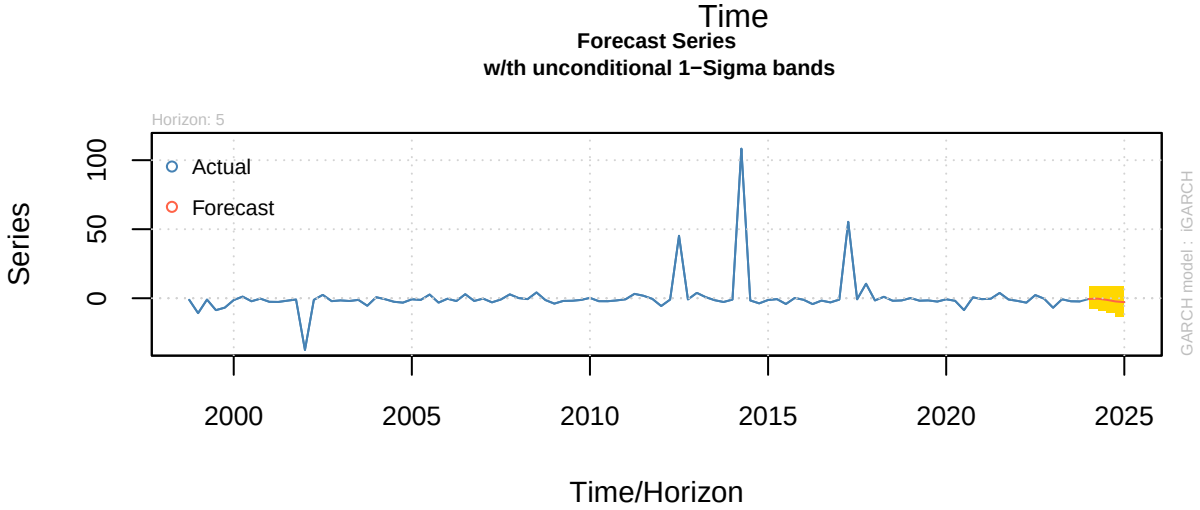
APARCH


Conditional SD (vs |returns|)
Volatility
0
100
200
1970
1980
1990
2000
2010
2020
Time
GARCH model : apARCH

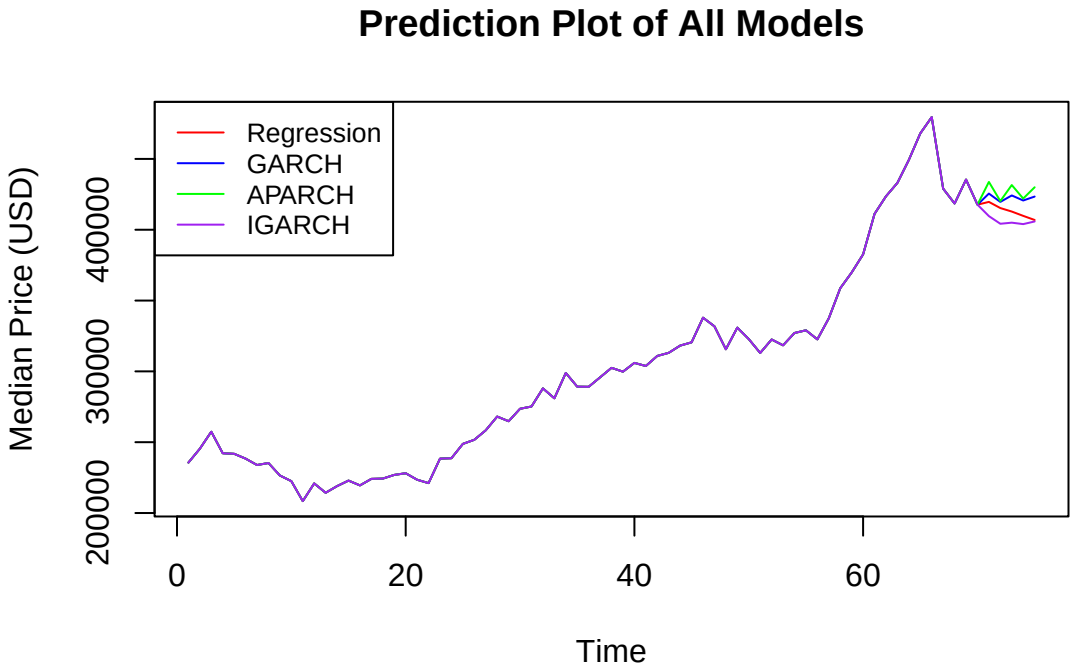

Forecast Series
w/th unconditional 1−Sigma bands
Horizon: 5
Actual
Forecast
Series
0
50
100
2000
2005
2010
2015
2020
2025
Time/Horizon
GARCH model : apARCH

IGARCH


Conditional SD (vs |returns|)
Volatility
0
100
200
1970
1980
1990
2000
2010
2020
Time
GARCH model : iGARCH


Forecast Series
w/th unconditional 1−Sigma bands
Horizon: 5
Actual
Forecast
Series
0
50
100
2000
2005
2010
2015
2020
2025
Time/Horizon
GARCH model : iGARCH

Results



We conclude that MSACSR (Median Sales Price of Houses Sold) and MORTGAGE30US (30 Year Fixed Rate Mortgage Average) are significant predictors of housing prices. ARIMA (1,1,1) shows the best performance among all models tested with autocorrelated errors.

We find that all three ARCH models fail to capture volatility patterns with a high degree of satisfaction. APARCH model does not seem to model negative volatility much differently from positive volatility. This may be due to very limited negative shocks experienced by the housing market, historically.We find that GARCH, APARCH & IGARCH models predict returns differently. This results in different values of the predicted values.

Discussion

GARCH reacts sharply to new information (higher alpha) but suffers from high persistence (too high beta). IGARCH also reacts sharply to new information but

consistently overestimates volatility (high persistence). APARCH fails to respond to new information quickly and consistently underestimates peaks.

References

Shumway, R. H., & Stoffer, D. S. (2000). *Time series analysis and its applications*. Springer.

Appendix

Step-wise Algorithm results to find the optimum regression model:

```
## Start:  AIC=-2178.18
## mspus.box.diff ~ time_mspus + houst.ts1 + mortgage.ts1 + msacsr.ts1 +
##     rhorusq156n.ts1
##
##                   Df   Sum of Sq       RSS     AIC
## - rhorusq156n.ts1  1 0.000048741 0.0062543 -2178.5
## - time_mspus       1 0.000049118 0.0062547 -2178.5
## - houst.ts1        1 0.000055033 0.0062606 -2178.3
## <none>                           0.0062056 -2178.2
## - mortgage.ts1     1 0.000101557 0.0063071 -2176.8
## - msacsr.ts1       1 0.000199669 0.0064052 -2173.5
##
## Step:  AIC=-2178.53
## mspus.box.diff ~ time_mspus + houst.ts1 + mortgage.ts1 + msacsr.ts1
##
##                   Df   Sum of Sq       RSS     AIC
## - houst.ts1        1 0.000033771 0.0062881 -2179.4
## <none>                           0.0062543 -2178.5
## + rhorusq156n.ts1  1 0.000048741 0.0062056 -2178.2
## - time_mspus       1 0.000087123 0.0063414 -2177.6
```

```
## - mortgage.ts1     1 0.000100425 0.0063547 -2177.2
## - msacsr.ts1       1 0.000217177 0.0064715 -2173.4
##
## Step:  AIC=-2179.4
## mspus.box.diff ~ time_mspus + mortgage.ts1 + msacsr.ts1
##
##                   Df  Sum of Sq       RSS     AIC
## <none>                           0.0062881 -2179.4
## + houst.ts1        1 0.00003377 0.0062543 -2178.5
## + rhorusq156n.ts1  1 0.00002748 0.0062606 -2178.3
## - mortgage.ts1     1 0.00010843 0.0063965 -2177.8
## - time_mspus       1 0.00016324 0.0064513 -2176.0
## - msacsr.ts1       1 0.00039291 0.0066810 -2168.7
```