Using the S&P 500 Index to Forecast U.S. Home Prices

Max Johnson and Rory Fan

2024-12-16

Abstract

The stock market and housing prices are widely recognized as key indicators of a nation's economic well-being. While the stock market tends to react quickly to market news and macroeconomic events, the real estate market is generally slower to adjust, reflecting longer-term economic trends. Given the common economic drivers that influence both markets, this study seeks to investigate the lead-lag relationship between the stock market and home prices. To conduct this analysis, we focus on the SPY and HPI indices as proxies for stock market and housing market performance respectively. We begin by constructing three linear regression models to explore the relationship between these two indices. To account for the residuals in these models, we apply different SARIMA models to test their significance and the validity. After performing various statistical tests, two of the models pass the criteria and demonstrate strong forecasting power. Then we use the SPY forecasts to predict HPI. Both models generate results that closely align with actual data, providing valuable insights into the interconnectedness of stock market performance and home prices. This study contributes to a deeper understanding of how these two important economic indicators influence one another over time, offering practical implications for investors and policymakers alike.

1. Introduction

The U.S. Stock Market and U.S. House Price Index are two key indicators of U.S. economic health[1]. Financial market dynamics, particularly the distinction between bull and bear markets, serve as critical indicators of economic sentiment and potential macroeconomic trajectories. In a bull market, sustained stock price increases reflect investor optimism and anticipation of economic expansion. This positive sentiment correlates with enhanced consumer confidence, increased discretionary spending, and potential GDP growth. Conversely, a bear market represents a period of stock price decline, signaling economic uncertainty and reduced market confidence. Such conditions typically trigger a contraction in investment strategies, decreased consumer spending, and a more risk-averse economic landscape.

Similarly, housing prices often mirror these economic patterns. During periods of strong consumer spending, housing demand typically escalates, consequently driving property valuations upward. On the other hand, economic contractions lead to decrease consumer spending, manifesting in reduced housing demand and price depreciation. Therefore, housing prices and the broader economy also move in tandem.

Given the similar significance of the stock market and housing prices in economic performance, they maintain important differences. On one hand, the stock market is volatile and quick, reacting to new information instantly. On the other hand, housing prices are slow. They move only as quickly as people buy and sell houses. This substantial difference in metrics motivates our research, which seeks to address two central questions: How do these similar markets—each exhibiting vastly different volatilities and reaction speeds—correlate? If a correlation exists, can the faster-reacting stock market be leveraged to forecast movements in the slower housing market?

---

[1]The scope of this research is limited to the United States, unless otherwise specified, all references to the stock market, housing prices, and economy should be understood as referring to the U.S.

2. Methodology

Framework

We begin with an exploratory data analysis of stock market and home price index data, examining key trends and relationships. Data transformation might be conducted to facilitate further analysis. Next, an ARIMAX model will be identified and constructed to establish the relationship between stock market and home price indices. We assess the distribution of the cross-correlation function between the stock market and home price series, implementing a pre-whitening process if necessary. After identifying the optimal predictive lags, an ordinary least squares regression model is fitted to the data. The model will be further examined to determine whether ARIMA error components should be incorporated. Finally, the model will be evaluated and adjusted to improve its predictive accuracy. Additionally, we also test three different forecasting methods, specifically exponential smoothing, SARIMA model, and OLR with SARIMA errors (ARIMAX), to determine the model effectiveness.

Data

Due to the immense size of the stock market and virtual inconceivability as a dataset, we chose the S&P 500 to represent the general market performance. S&P500 tracks the stock performance of 500 of the largest public companies by market capitalization. It includes approximately 80% of the total market capitalization of U.S. public companies. We use SPY, or SPDR S&P 500 ETF Trust, as the S&P 500 index.

The housing price data comes from the S&P CoreLogic Case-Shiller U.S. Home Price Index (HPI), a measurement that tracks and weights the repeat sales of single-family houses to generate the seasonally adjusted home price index. The Case-Shiller index is used by the Federal Reserve to evaluate housing inflation.

The stock market and housing price data are monthly sets covering 60 months from September 2020 to August 2024. For SPY, we calculate the average daily closing price of each month, whereas HPI data is already reported on a monthly basis

Exploratory Data Analysis
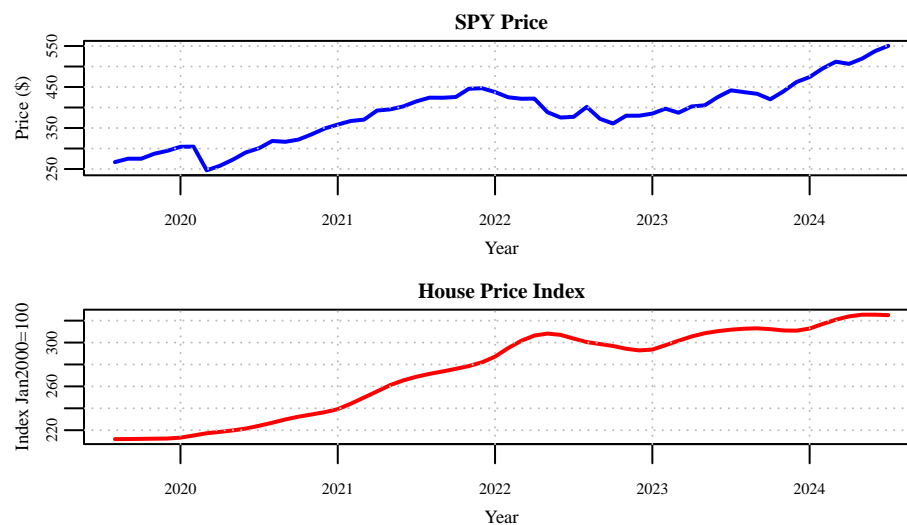
Both SPY and HPI exhibit a similar upward trend overall. SPY experiences a slight decline at the beginning of 2022, while HPI follows with a comparable drop three to five months later. These price decreases can be attributed to the Federal Reserve's interest rate hike policy. The time lag gives us preliminary knowledge about the potential time lag between the two series. Additionally, the HPI data may indicate potential signs of annual seasonality.

Due to the presence of strong trends, we decide to remove the trends in both SPY and HPI by conducting

first-degree differencing ([Plots](#)). This helps us focus on the fluctuation of monthly difference in the series. The potential issues of heteroscedasticity and other variance-related properties will be easier to identify. The study will use the differenced SPY and differenced HPI from now on.

3. Result

Identify Possible Predictors from Lead/Lag Relationship

CCF: differenced SPY vs differenced HPI
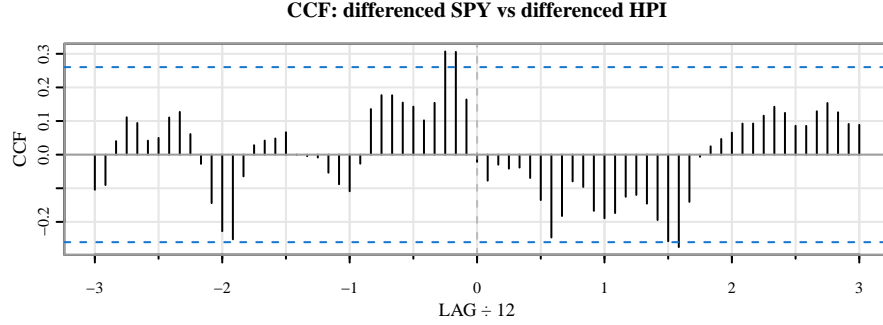
We perform a cross-correlation function (CCF) between the differenced SPY and the differenced HPI. The result indicate that the lag 2 and lag 3 are significant, suggesting these two lags are potential predictors. However, the significant lag observed at +19 will be disregarded due to the lead/lag relationship, in which HPI leads SPY and its considerable temporal distance. Using SPY lag -2 and -3 as predictors, the resulting ordinary linear regression model is as follows:

$$\nabla \hat{HPI}_t \;=\; \beta_0 \;+\; \beta_1 \nabla SPY_{t-2} \;+\; \beta_2 \nabla SPY_{t-3}$$

Running the regression gives the following results:

```
##               Estimate Std. Error  t value      Pr(>|t|)
## (Intercept) 1.60598593 0.32325667 4.968145 7.428040e-06
## spylag2     0.04729885 0.02045140 2.312744 2.464996e-02
## spylag3     0.04725241 0.02049805 2.305215 2.509944e-02
```

The p-value for the intercept, lag 2, and lag 3 are all below 0.05, indicating statistical significance. This confirms the validity of the model for predictive purposes. Using the coefficients estimated by R, the resulting equation is as follows:

$$\nabla \hat{HPI}_t \;=\; 1.6056 \;+\; 0.04723 \nabla SPY_{t-2} \;+\; 0.04725 \nabla SPY_{t-3}$$

3

Fit Regression with SARIMA Errors

ACF and PACF of the Residuals from OLR Model
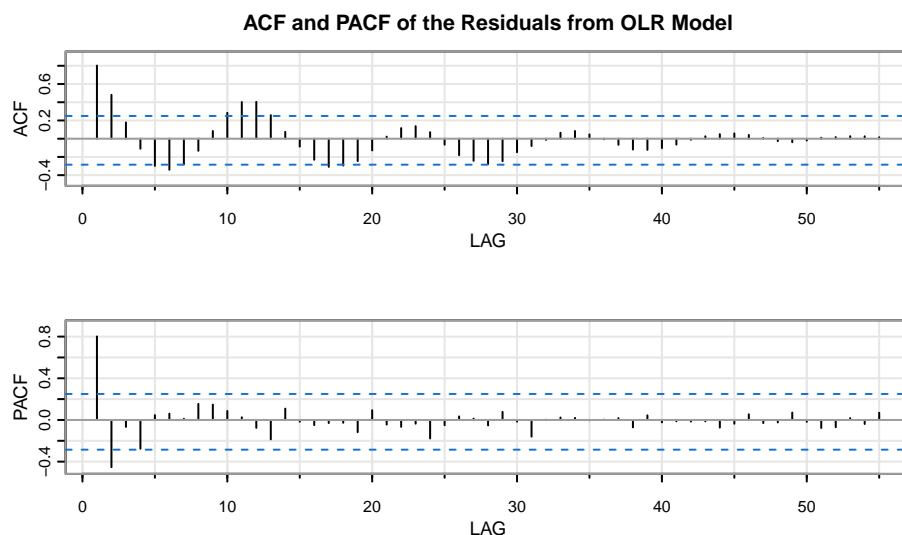
The ACF and PACF of the residuals from the OLS model are plotted to assess the potential presence of ARIMA errors. The plots indicate that the residuals do not represent a white noise series, suggesting the need for a SARIMA error model. The ACF plot shows signs of seasonality at lag 12, corresponding to a yearly cycle, while the PACF plot exhibits two significant lags before tapering off. These patterns suggest an AR(2) or ARMA model for the non-seasonal component, with SAR(1) potentially required to account for the seasonal effect. In the next part, we will test three models: two models will use the lags 2 and lag 3 as the predictors, while one model will use the lag 2 as the sole predictor

Model 1:
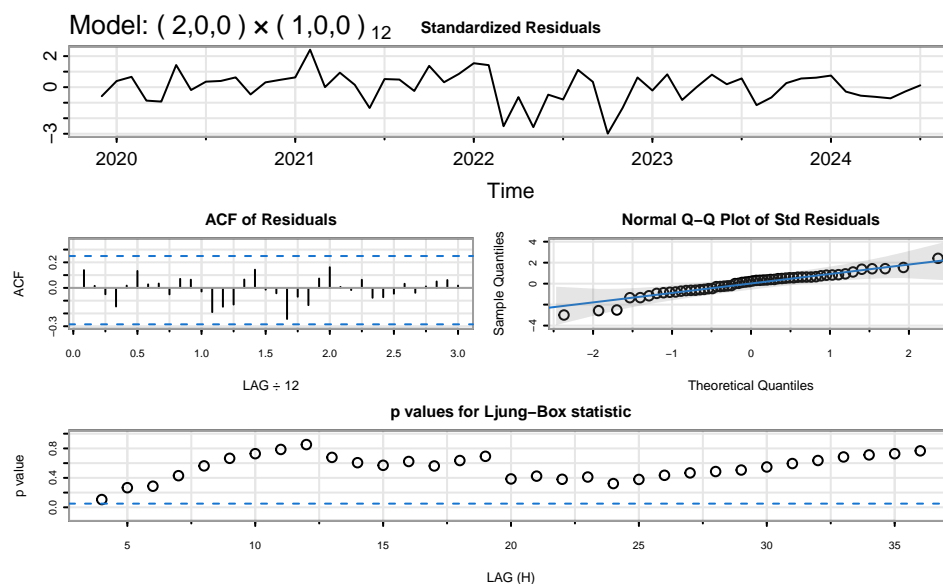


```
##          Estimate     SE t.value p.value
## ar1        1.5451 0.0928 16.6412  0.0000
## ar2       -0.6999 0.0893 -7.8358  0.0000
## sar1       0.5623 0.1102  5.1028  0.0000
## intercept  1.6058 1.0349  1.5516  0.1271
```

4

```
## df$spylag2    0.0088 0.0044  2.0183  0.0489
## df$spylag3    0.0065 0.0046  1.3941  0.1695
```

$$\nabla H\hat{P}I = 1.606 + 0.009\nabla SPY_{t-2} + 0.007\nabla SPY_{t-3} + \epsilon_t$$

$$(1 - 0.562B^{12})(1 - 1.545B + 0.700B^2)\epsilon_t = w_t$$

$$w_t \sim N(0, 0.431)$$

The diagnostic plots look satisfactory. The standardized residuals plot exhibits a pattern consistent with i.i.d. errors. There are no significant lags in the ACF of the residuals. The Q-Q plot, while showing a few heavy tail data points, suggests that the residuals follow a normal distribution. Furthermore, all lags in the Ljung-Box test are significant. However, when examining the coefficients, the intercept, lag 2, and lag 3 are found to be insignificant, even though the SARIMA error terms fitting well. As a result, Model 1 is deemed invalid.

Model 2:
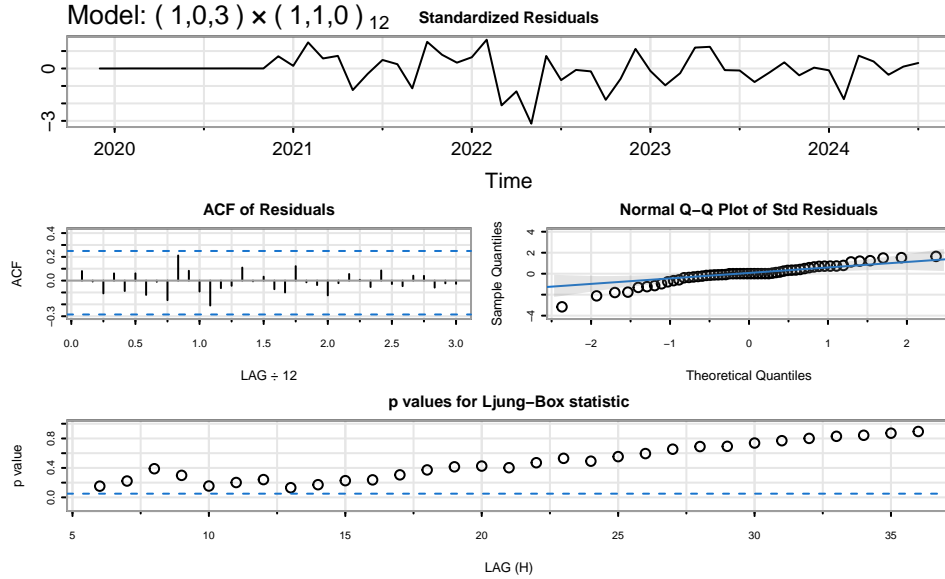


```
##            Estimate     SE t.value p.value
## ar1          0.7171 0.1086  6.6052  0.0000
## ma1          0.7310 0.1381  5.2950  0.0000
## ma2          1.0379 0.2173  4.7756  0.0000
## ma3          0.6748 0.1732  3.8967  0.0004
## sar1        -0.3861 0.1441 -2.6783  0.0110
## df$spylag2   0.0174 0.0024  7.2409  0.0000
## df$spylag3   0.0074 0.0033  2.2399  0.0312
```

$$\nabla H\hat{P}I = 0.0174\nabla SPY_{t-2} + 0.007\nabla SPY_{t-3} + \epsilon_t$$

$$(1 + 0.386B^{12})(1 - 0.717B)\nabla_{12}\epsilon_t = (1 + 0.731B + 1.038B^2 + 0.675B^3)w_t$$

$$w_t \sim N(0, 0.372)$$

Using a SARIMA(1, 0, 3)x(1, 1, 0)$_{12}$ error model, the standardized residuals generally exhibit an i.i.d. pattern. There are no significant lags observed in the ACF of the residual plot. However, a heavy tail is evident in the Q-Q plot, particularly in the lower tail. The Ljung-Box statistics are all significant. Examining

the coefficients, all SARIMA error terms have p-values close to 0. Additionally, the lag 2 and lag 3 terms are significant. Despite the presence of heavy tails in the normal Q-Q plot, the model may still be valid enough for consideration.

Model 3 with only lag 2:
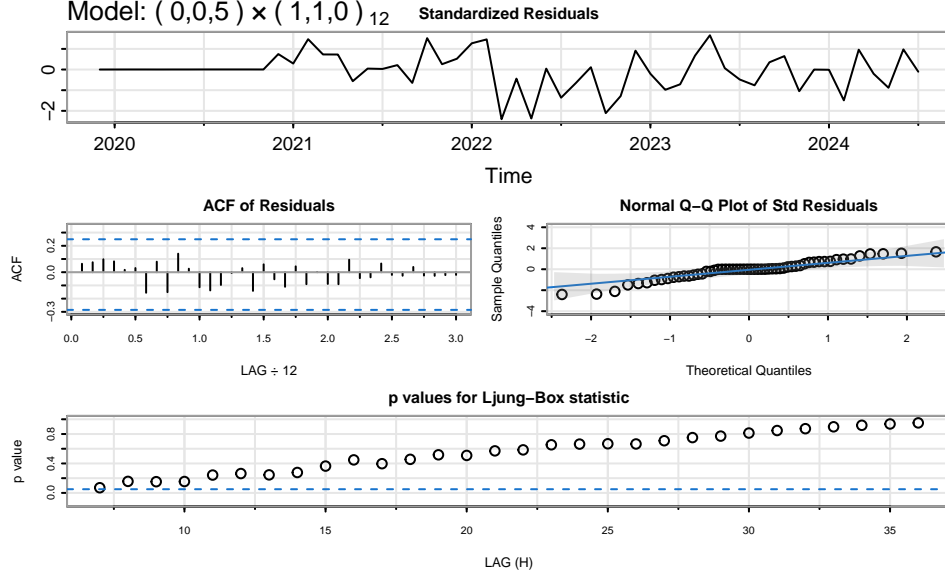

```
##       Estimate     SE t.value p.value
## ma1     1.4482 0.1549  9.3508  0.0000
## ma2     2.0669 0.3233  6.3926  0.0000
## ma3     1.9171 0.4821  3.9769  0.0003
## ma4     1.2522 0.4194  2.9860  0.0050
## ma5     0.5269 0.2168  2.4303  0.0201
## sar1   -0.3890 0.1415 -2.7489  0.0092
## xreg    0.0150 0.0034  4.4117  0.0001
```

$$\nabla\hat{HPI} = 0.015\nabla SPY_{t-2} + \epsilon_t$$

$$(1 + 0.389B^{12})\nabla_{12}\epsilon_t = (1 + 1.45B + 2.067B^2 + 1.917B^3 + 1.252B^4 + 0.527B^5)w_t$$

$$w_t \sim N(0, 0.391)$$

With the SARIMA$(0, 0, 5)$x$(1, 1, 0)_{12}$ error model, the standardized residuals show an i.i.d. pattern in general. There are no significant lags observed in the ACF of the residual plot. However, similar to the model 2, we observe a heavy tail present in the Q-Q plot, although it is less significant than the model 2. The Ljung-Box statistics are all significant. All of the coefficients in the model 3 have very low p-values, proven to be significant. Because of these characteristics, the model is valid enough for consideration.

Based on the results, the Model 2 or Model 3 will be used in subsequent forecasting.

Build a Predictive Model for Stock Market Data

SPY lags forecast HPI data in the exogenous linear regression with SARIMA errors model. However, when forecasting past two values, there will not be any future SPY data to make those forecasts. Therefore, SPY data must also be modeled to predict the HPI data further.

ARIMA and double exponential smoothing models will be compared to forecast SPY. To do so, first, the last 5 data points will be removed and used as test data, while the remaining points will be used as training data. Afterward, models will be fit and compared to the test set.

	MAE	RMSE
ARIMA	17.06	20.37
ETS	13.60	16.77

When comparing ARIMA(0, 1, 0) and ETS forecasting models, the ETS showed less error. As a result, the ETS method will be used for HPI forecasting.

Forecasting and Model Comparison

The following predictive HPI models will be compared:

- Exponential Smoothing
- SARIMA
- OLR with SARIMA Errors (ARIMAX)

To determine model effectiveness, the HPI and SPY data are be split into train and test sets as done before. The models then forecast the next five HPI values and compared with each other and the true values.

Model Forecasted Values

	TRUE VALUES	SARIMA	ETS	MODEL 2	MODEL 3
Mar 2024	**320.885**	321.305	319.179	319.619	320.743
Apr 2024	**323.859**	325.277	320.369	321.205	323.980
May 2024	**325.436**	328.351	321.462	321.511	326.207
Jun 2024	**325.461**	330.398	322.828	320.621	326.973
Jul 2024	**325.025**	331.712	324.755	319.168	326.669

Model Errors

	MAE	RMSE
ARIMA	3.275	3.994
ETS	2.415	2.753
MODEL 2	3.709	4.044
MODEL 3	0.838	1.060

The testing compared 5 fitted models: ARIMA(1, 0, 3)x(1, 1, 0)$_{12}$, Holt Winters triple exponential smoothed (ETS), Model 2, and Model 3. Model 2 did not perform that well when compared to the other models. On the other hand, Model 3 worked very well. It had small errors and almost exactly matched the true values during May and April. This is very intuitive because those two values are predicted using the true SPY values, not yet needing to use SPY forecasting.

4. Conclusion

The paper explores the lead-lag relationship between the S&P 500 stock index and Case-Shiller Home Price Index. Our results suggest that there is a strong connection between these two time series. Model 2 indicates that SPY leads HPI by two month lag and three month lag, while Model 3 indicates SPY leads HPI by exactly

two month. Different SARIMA errors terms are identified to construct the final model. In our forecasting tests, we found that the Model 3 demonstrates strong accuracy in short-term prediction, compared to other common time series predictive methods, which validates our research result.

Both models provide useful insight into the economic relationship between two indices. The stock market is widely known for the quick reaction that represents the market's current sentiment. On the other hand, the real estate market, with a low liquidity and transparency, takes longer time for price discovery. Only when a home sales transaction happens, the price becomes public. Our research result of two month lag fits the common real estate transaction time.

While limited research exists on the lead-lag relationship between the stock market index and home price index, our findings offer preliminary insights into this area of study. To further improve the research outcome, additional analysis should be conducted. Specifically, our study is based on monthly data, and future research could extend this to daily, weekly, or yearly datasets to capture a broader range of trends. Moreover, incorporating other exogenous factors, such as inflation rates, tax policies, and other macroeconomic variables, into the model could provide a more comprehensive understanding of the relationship by accounting for external economic influences.

5. Bibliography

GK. "S&P 500 SPY Daily Price History." Kaggle, 2024. https://www.kaggle.com/datasets/gkitchen/s-and-p-500-spy

Ismail, Mohd Tahir. "Modelling and Forecasting S&P 500 Stock Prices using Hybird ARIMA-GARCH Model." Journal of Physics Conference Series, 1366(1): 012130, 2019.

Muhammad, Adi Prayogo, Wahyu Wibowo, and Mike Prastuti. "Forecasting of Analytic Residential Price Index Using ARIMA Box-Jenkins." Proceedings of the American Institute of Physics Conference, vol. 2668, no. 1, 2023, p. 070019.

"S&P CoreLogic Case-Shiller U.S. National Home Price Index." Federal Reserve Bank of St. Louis, 2024. https://fred.stlouisfed.org/

6. Appendix

Differenced SPY and Differenced HPI plots
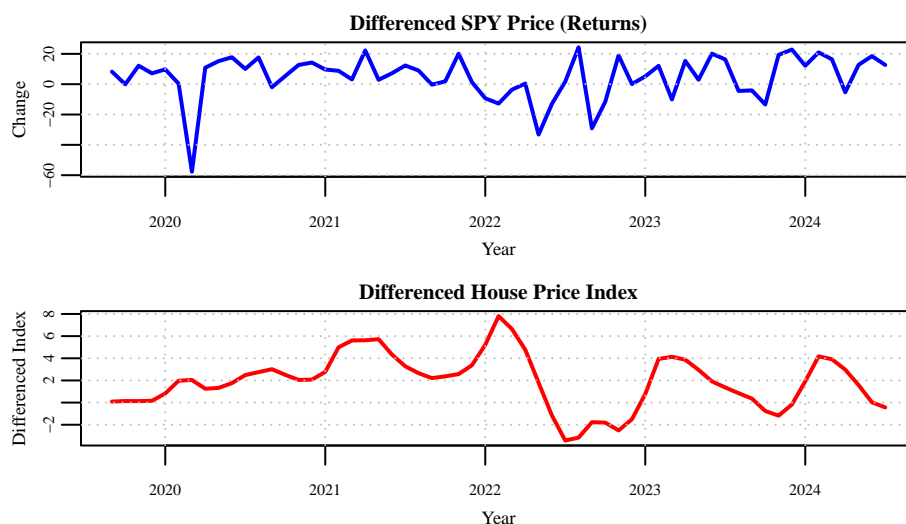
7. Contributions

Max Johnson: Model Building, Forecasting Test/Evaluation, Presentation (Intro, SPY Model, Forecasting), Report (Introduction, Methods, and Result).

Rory Fan: Data Collection, Exploratory Data Analysis, Presentation (Regression Model, Conclusion), Report (Abstract, Methods, Conclusion, Revision).