Trend analysis of Housing Price Index using Time Series

Abhiram Vadlamani, Niyati Srivastava, Parth Mahajan

December 2023

1 Introduction and Motivation

1.1 Abstract

Housing market in the United States of America has always been the most buzzing area of interest for investors, economists, the general public and the government. The country has witnessed the most volatile economic situations arising due to the mortgage industry. There has always been a need to predict the movement of Housing Price index(HPI) for the future. This stems from the need to buy houses, to invest and diversify portfolio or to predict consumer sentiment. Although there are a number of economic variables which influence the movement of HPI, we decided to use 30-year mortgage rates because of its high correlation and recent interesting trends driven by the Federal Reserve Bank's monetary policies.

This project aims to conduct a comprehensive time series analysis on the relationship between 30-year mortgage rates and the Housing Price Index (HPI) over a 30-year period, utilizing monthly data. The primary focus is on employing prewhitening techniques to identify and mitigate the impact of autocorrelation in the mortgage rate time series and covert the same into a white noise series. Subsequently, a transfer function model is applied to capture the dynamic relationship between prewhitened mortgage rates and the HPI. Additionally, the project seeks to model and analyze the residual noise to gain insights into unaccounted factors influencing the housing market. Lastly, we analysed the outliers that are present in our data to better understand their impact, i.e if the outliers are additive or innovational in nature, prior to forecasting the series.

1.2 Process Outline

The initial step involved Exploratory Data Analysis (EDA), where we examined the temporal patterns, trends, and potential seasonality in both the mortgage rates and HPI series. We plotted the histograms to better understand features such as each variables' distribution and skew. To complement the histograms plots, we finally employed boxplots to better understand key quantiles and the presence of outliers in our data.

Following the EDA, we prewhitened the mortgage rate time series to eliminate autocorrelation, a crucial step in understanding the true relationship between 30 year mortgage rates and housing prices.

After prewhitening, we applied a transfer function model to capture the dynamic interactions between the prewhitened 30 year mortgage rates and HPI.

The next step involved fitting a model to the residual noise, enabling a detailed analysis of any remaining unexplained variance.

Following the model fitting, we noticed that the data contained some outliers. Therefore, we decided to analyse these outliers to understand the impact they had on the HPI.

Subsequently, we employed forecasting techniques to predict the future trends in the HPI, incorporating the insights gained from the transfer function model.

The project concludes with a presentation of results, showcasing the effectiveness of prewhitening in revealing underlying patterns, the predictive power of the transfer function model, and insights derived from the residual noise analysis. This comprehensive approach provides valuable insights into the intricate dynamics between 30 year mortgage rates and housing prices.

2 Data Exploration

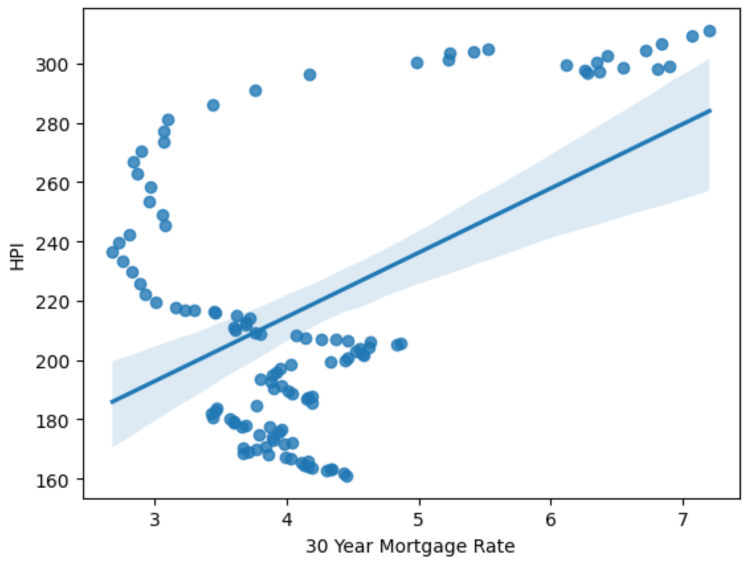

Figure 1: HPI vs 30 year mortgage rate Scatterplot

We see from the above scatterplot that there is a relationship between the HPI and the 30 year mortgage rates. While this relationship is positive, it is not linear but concave. However, with a correlation of 0.482, there is a linear trend that exists between our variables as well.

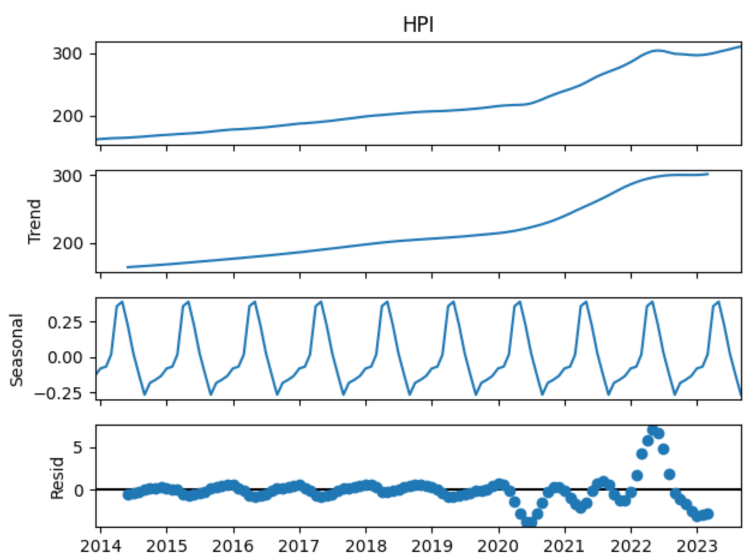

Figure 2: HPI Decomposition

We see in the above plot that there are seasonal components to our HPI series, with residuals oscillating around 0 until 2022, after which the HPI experienced some volatility.

Figure 3: 30 year mortgage rate Decomposition

Similar to the HPI, there exists a seasonal pattern in the 30 year mortgage rates as well. Furthermore, the rates exhibit an increased volatility post 2022 as well, which could suggest a relationship between the two variables under consideration, since they show similar trends during the same time intervals.

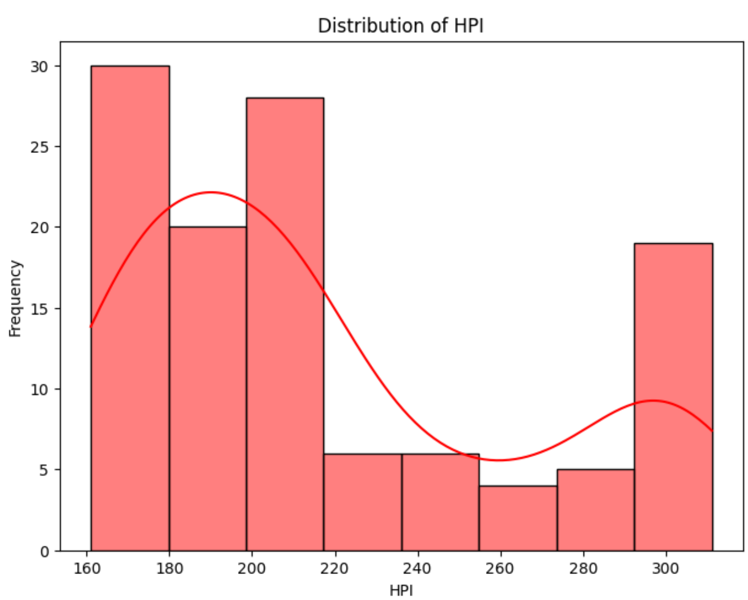

Figure 4: Distribution of HPI

We see from the histogram of the HPI that our data is right skewed, which means that we have a higher mass at relatively smaller values of the HPI.

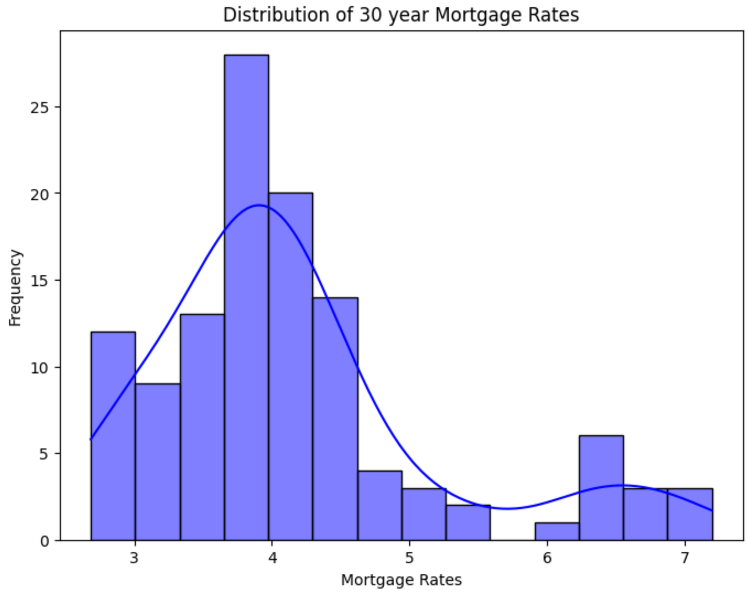

Figure 5: Distribution of 30 year mortgage rate

We observe a similar trend in the histogram of our 30 year mortgage rates as we did in the histogram of the HPI, in that our data is skewed to the right.

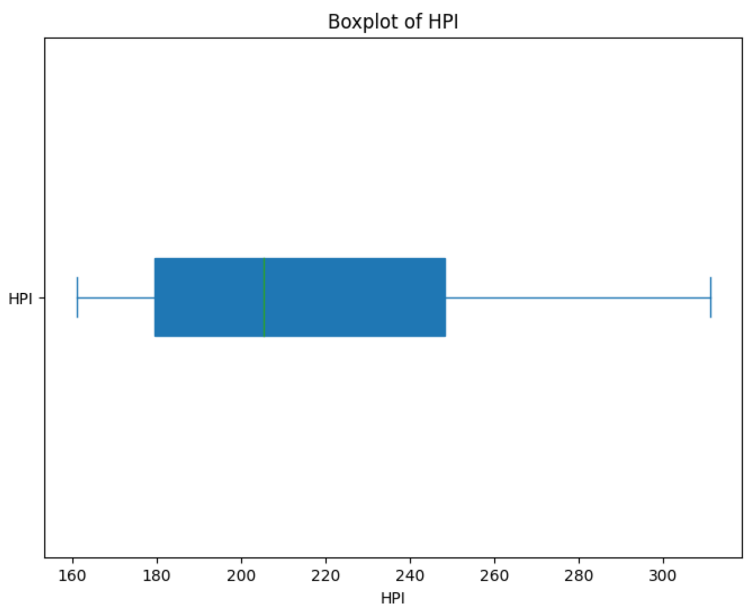

Figure 6: Boxplot of HPI

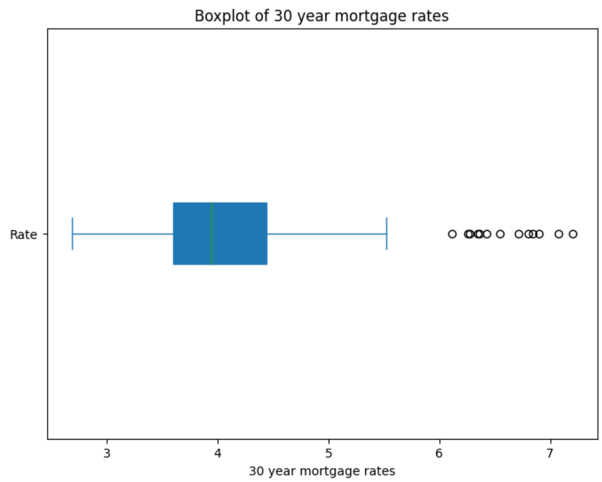

Figure 7: Boxplot of 30 year mortgage rates

The key insight from the boxplot of the 30 year mortgage rates is the presence of outliers. We take this into account and identify the impact of these outliers later in our analysis.

3 Applying Transfer Function Model

3.1 Pre-Whitening Independent Variable

Our study includes two key variables 30-Year Mortgage Rates and the Housing Price Index. We denote our X(independent) variable to be Rates and Y (Dependent) Variable to be HPI.

In order to fit the transfer function model as a cause and effect model to HPI, we first began by pre-whitening the X variable. The diagnostic plots of the rates variable are below.

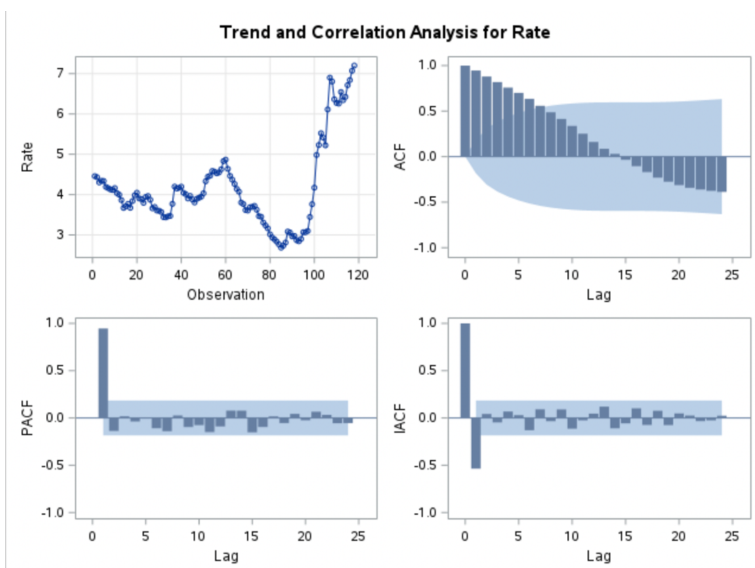

Figure 8: Data Diagnostics

An initial look at out data indicates a slowly decreasing Auto correlation Function which points towards non-stationarity and the need for differencing. By looking at the data it makes most sense to test for the single mean ADF. The results are shown below.

Dickey-Fuller Unit Root Tests							
Type	Lags	Rho	Pr < Rho	Tau	Pr < Tau	F	Pr > F
Zero Mean	0	0.7609	0.8658	1.54	0.9693		
Single Mean	0	2.3969	0.9988	1.15	0.9977	1.51	0.6872
Trend	0	0.6272	0.9981	0.29	0.9984	4.09	0.3599

Figure 9: Unit Root Test

We get a P value of > 0.99 clearly indicating the need for differencing.

We try differencing once and look at the result.

Autocorrelation Check for White Noise									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	35.44	6	<.0001	0.410	0.031	0.052	0.109	0.260	0.199
12	37.85	12	0.0002	0.012	0.050	0.014	-0.055	0.062	0.093
18	41.82	18	0.0012	0.073	0.088	0.083	0.038	0.046	0.073
24	46.50	24	0.0039	0.062	0.004	-0.073	-0.085	-0.097	-0.079

Figure 10: Autocorelation Check for White Noise

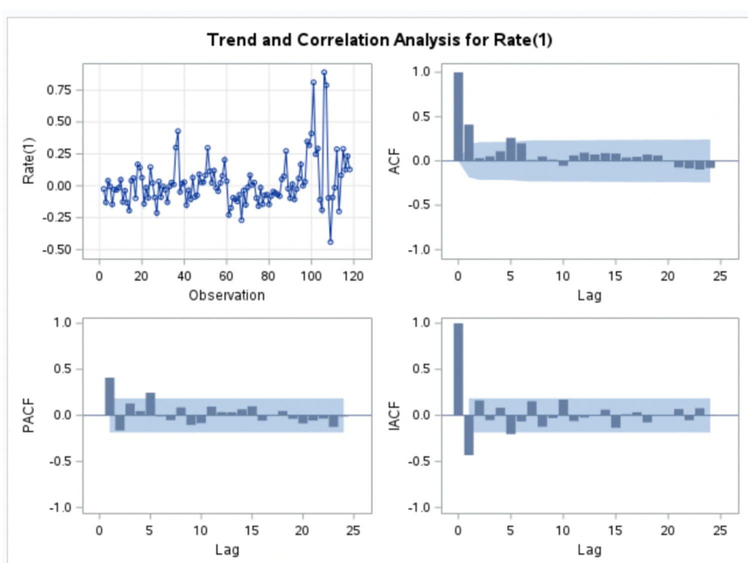

Figure 11: Diagnostics for Differenced Data

The differenced data appears to be stationary. Let's look at the Augmented Dickey Fuller test with zero mean for validation.

Augmented Dickey-Fuller Unit Root Tests							
Type	Lags	Rho	Pr < Rho	Tau	Pr < Tau	F	Pr > F
Zero Mean	0	-67.3452	<.0001	-6.84	<.0001		
	1	-91.8394	<.0001	-6.69	<.0001		
Single Mean	0	-68.3404	0.0010	-6.89	<.0001	23.75	0.0010
	1	-94.4738	0.0010	-6.78	<.0001	22.98	0.0010
Trend	0	-73.1484	0.0004	-7.21	<.0001	26.03	0.0010
	1	-107.272	0.0001	-7.19	<.0001	25.87	0.0010

Figure 12: ADF

The P value is less than 0.0001 and hence we reject the null hypothesis and do not difference further.

The ACF has a spike at 5 and so does the PACF. Let's test two models (5,1) and (5,5). First we look at the diagnostics for (5,1)

Conditional Least Squares Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MU	0.02417	0.03639	0.66	0.5080	0
MA1,1	-0.02277	0.38404	-0.06	0.9528	1
AR1,1	0.45852	0.37159	1.23	0.2199	1
AR1,2	-0.22993	0.20572	-1.12	0.2661	2
AR1,3	0.15768	0.12981	1.21	0.2271	3
AR1,4	-0.07231	0.11097	-0.65	0.5160	4
AR1,5	0.25496	0.09495	2.69	0.0084	5

Constant Estimate	0.01042
Variance Estimate	0.02979
Std Error Estimate	0.172599
AIC	-72.2738
SBC	-52.9386
Number of Residuals	117

Figure 13: AIC for (5,1)

Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	.	0	.	0.000	0.014	-0.021	0.029	0.009	0.030
12	5.63	6	0.4657	-0.091	0.046	-0.017	-0.172	0.012	0.032
18	7.97	12	0.7876	0.029	0.041	0.097	0.022	0.029	0.061
24	10.73	18	0.9055	0.056	0.047	-0.055	-0.006	-0.100	-0.021

Figure 14: Residual Diagnosis for (5,1) (A)

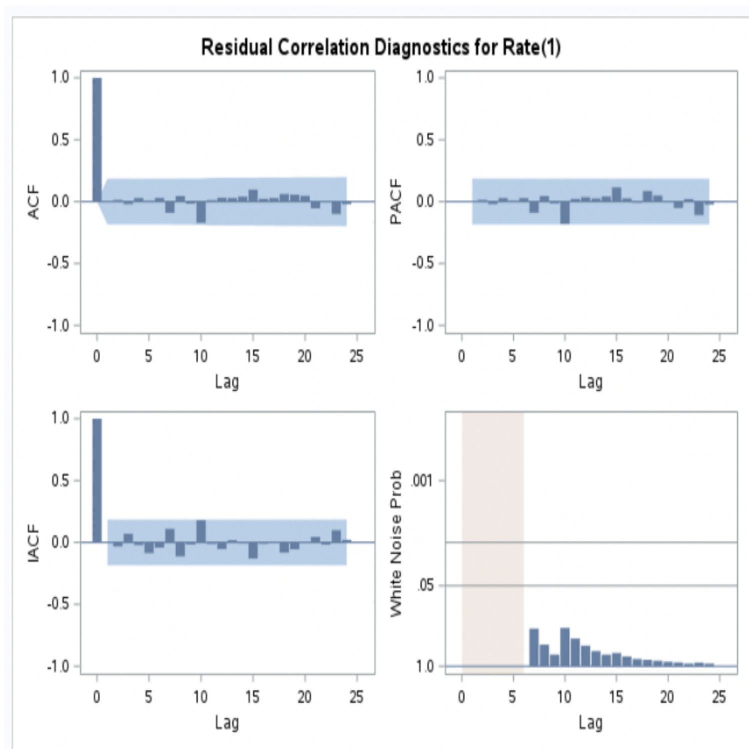

Figure 15: Residual Diagnosis for (5,1) (B)

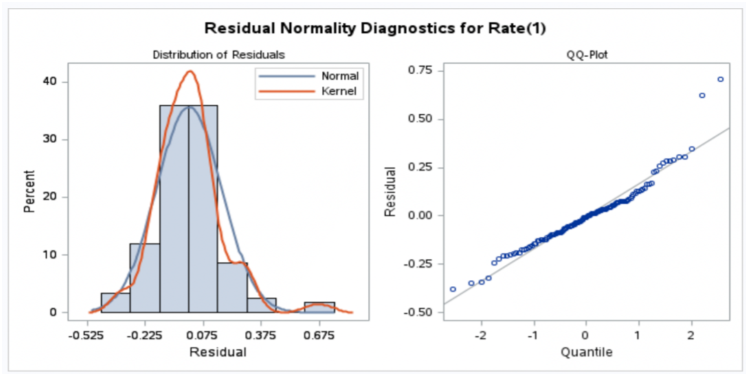

Figure 16: Residual Diagnosis for (5,1) (C)

The (5,1) model seems to be doing fairly well. The residuals indicate white noise. But the diagnosis for (5,5) model prove slightly better. AIC indicates a better fit and so does the residuals white noise test. Hence we move forward with this model. Below are the diagnostic plots for (5,5).

Conditional Least Squares Estimation

Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MU	0.02557	0.03326	0.77	0.4437	0
MA1,1	0.31920	0.20544	1.55	0.1232	1
MA1,2	-0.38400	0.20996	-1.83	0.0702	2
MA1,3	0.47089	0.17061	2.76	0.0068	3
MA1,4	-0.19295	0.20831	-0.93	0.3564	4
MA1,5	-0.48388	0.18301	-2.64	0.0094	5
AR1,1	0.80308	0.22083	3.64	0.0004	1
AR1,2	-0.70124	0.25450	-2.76	0.0069	2
AR1,3	0.77525	0.20738	3.74	0.0003	3
AR1,4	-0.52165	0.23483	-2.22	0.0285	4
AR1,5	0.06538	0.19032	0.34	0.7319	5

Constant Estimate	0.014811
Variance Estimate	0.028426
Std Error Estimate	0.168601
AIC	-74.0917
SBC	-43.7078
Number of Residuals	117

* AIC and SBC do not include log determinant.

Figure 17: AIC for (5,5)

Autocorrelation Check of Residuals

To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	.	0	.	0.007	-0.005	-0.004	0.000	-0.012	0.027
12	1.51	2	0.4695	-0.034	0.026	-0.028	-0.048	0.076	0.003
18	3.06	8	0.9306	0.059	0.038	0.023	0.037	0.045	0.049
24	5.63	14	0.9749	0.062	0.061	-0.049	-0.007	-0.087	-0.010

Figure 18: Residual Diagnostic for (5,5) A

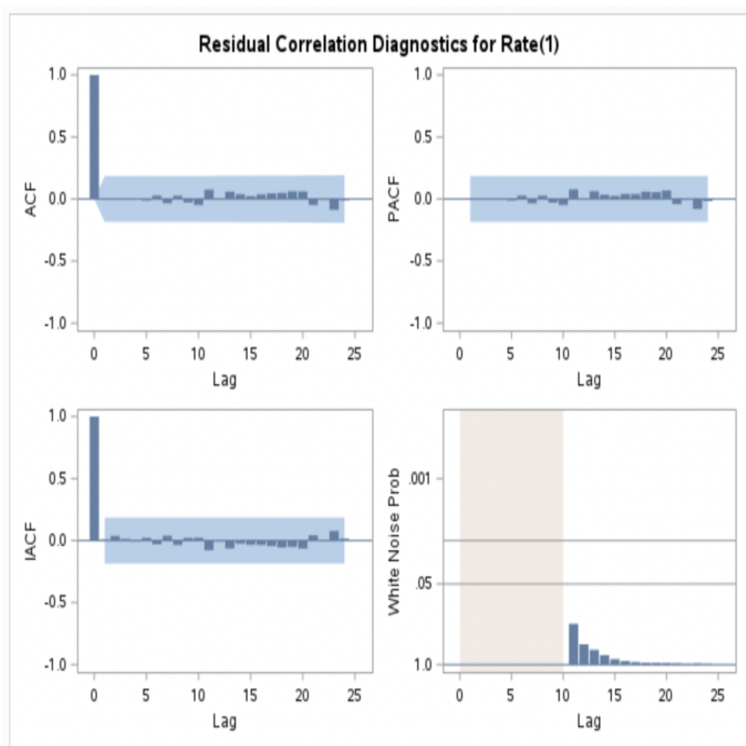

Figure 19: Residual Diagnostic for (5,5) B

3.2 Checking Cross Correlation

Once the X variables has been converted to white noise, we began identifying the transfer model. In theory, the rate variable has huge impact on housing price index. As the rates are increased, the demand for mortgages falls and in line with demand supply theory the house price index also falls. We try to convert this theory to practice by identifying the model parameters.

The cross correlation plot between the differenced HPI and differenced rates is shown below:

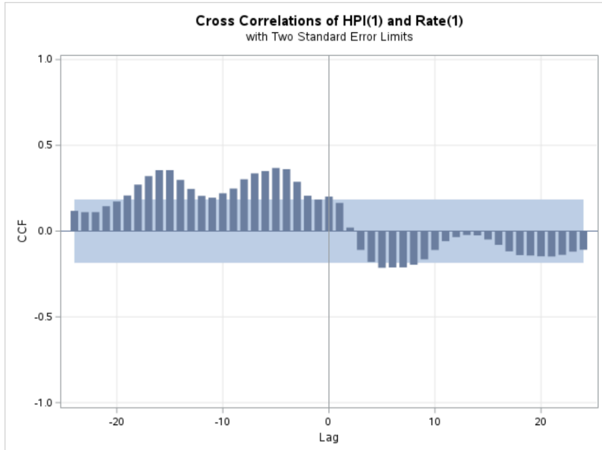

Figure 20: Cross correlation plots between differenced HPI and Rates

We observe significant values at negative lags and acknowledge there is feedback in the system. Therefore we first pre-whiten Y (HPI) variable and then proceed with cross correlation

3.3 Pre-Whitening Dependent variable(HPI series)

The diagnostic plots for the HPI Variable are as follows:

Augmented Dickey-Fuller Unit Root Tests

Type	Lags	Rho	Pr < Rho	Tau	Pr < Tau	F	Pr > F
Zero Mean	0	0.7153	0.8559	10.92	0.9999		
	1	0.7266	0.8584	1.29	0.9498		
	2	0.7008	0.8526	2.79	0.9987		
Single Mean	0	1.2325	0.9917	3.96	0.9999	62.04	0.0010
	1	-0.2251	0.9439	-0.10	0.9456	0.94	0.8311
	2	0.7787	0.9837	0.85	0.9945	3.87	0.0965
Trend	0	-0.7231	0.9907	-0.74	0.9673	10.29	0.0010
	1	-14.9341	0.1687	-2.89	0.1701	4.60	0.2586
	2	-3.8388	0.8928	-1.44	0.8437	2.03	0.7713

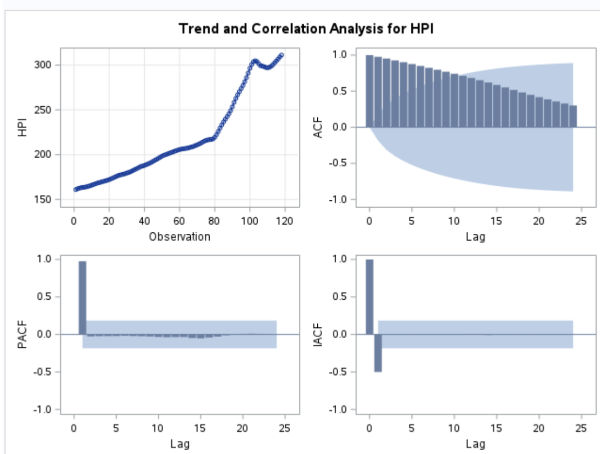

Figure 21: Trend diagnostics and stationarity test for HPI

A clear trend is observed in the plot of HPI. The slowly decaying ACF plots signal a need for differencing. There is also a significant lag at 1 in the PACF plot.

The Augmented Dickey Fuller test also fails to reject the null hypothesis of non-stationarity for trend case at lag 0 with a p-value of $0.9907 > 0.05$ for 95% confidence.

We fit the ARIMA(1,1,0) model to HPI and see the residual diagnostics:

Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	4.01	4	0.4043	0.039	0.015	-0.011	-0.070	0.020	-0.158
12	11.21	10	0.3413	0.143	0.057	0.062	0.164	-0.032	0.024
18	23.14	16	0.1101	-0.201	0.006	-0.063	0.099	0.181	-0.022
24	27.47	22	0.1941	0.040	-0.072	-0.018	-0.096	-0.111	-0.030

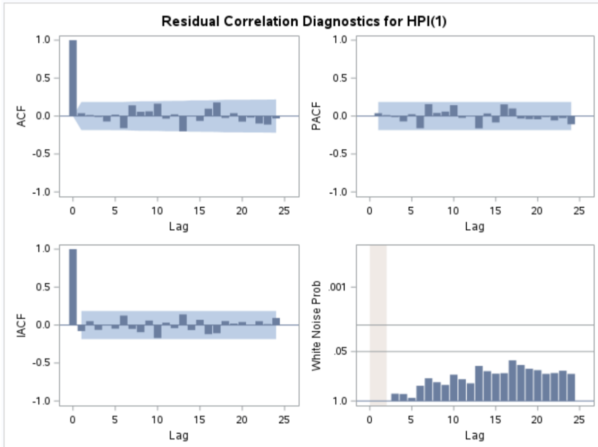

Figure 22: Residual diagnostics for HPI after ARIMA(1,1,0) model

This model was chosen as a good fit as the ACF/PACF plots imitate white noise and the Ljung Box test is not rejected at a p-value greater than 0.05.

The parameter estimates are printed below and the analysis is continued with this model

Conditional Least Squares Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MU	1.18506	0.24749	4.79	<.0001	0
AR1,1	1.53104	0.07350	20.83	<.0001	1
AR1,2	-0.62543	0.07369	-8.49	<.0001	2

Constant Estimate	0.11186
Variance Estimate	0.140836
Std Error Estimate	0.375282
AIC	105.6543
SBC	113.9408
Number of Residuals	117

Figure 23: Parameter Estimates for HPI after ARIMA(1,1,0) model

3.4 Cross Correlation after pre-whitening

We look to find the parameter mapping for the transfer function model:

$$PreWhitenedHPI = \omega_s(B)B^b/\delta_r(B)PreWhitenedRATE + n(t)$$

The cross correlation is plotted between the pre-whitened 'Rate' variable and pre-whitened 'HPI' variable. This is achieved by considering the residuals from both the series post applying whitening filter. The Cross Correlation plot is as follows:

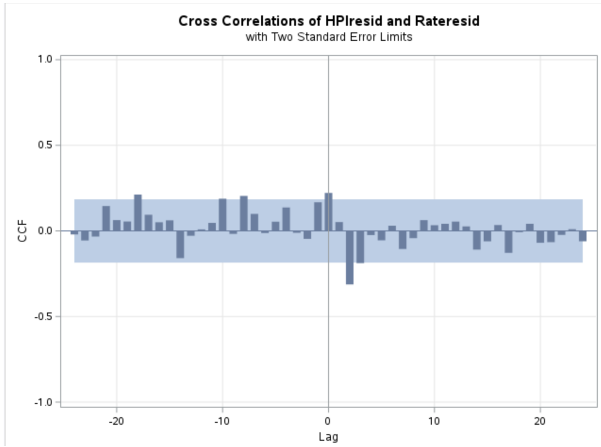

Figure 24: Crosscorrelation plot between pre-whitened HPI and Rates variable

We observe no significant values at negative lags and conclude there is no feedback in the system. Note: Although slight breach is witnessed at 1-2 negative lags they are grossly insignificant given the variables in question

The first significant spike is at lag=0. Therefore, we take 'b' = 0. This implies that there is no lag in the causal effect between Rates and HPI. The phenomenon is also seen in reality where a sudden drop or spike in mortgage rates leads to an impact on housing market and thus the housing price index

The pattern starts to die down after the 3rd lag and so we decide to put the 's' variable as 3 i.e the difference between 3 and 0.

Finally, we see a cutting off behavior in the correlation hence 'r' = 0 is viable. However, the pattern can also be construed as an sinusoidal decay which implies 'r' = 2.We try both models for transfer function filter and check diagnostics

3.5 Fitting overall model

Keeping in mind these transfer model parameters and above diagnostic plots we fit certain models.

The following two models are purely inspired by theoretical algorithm of first pre-whitening and then applying the filter

3.5.1 Model 1 with pre-whitened data and transfer model (b=0,s=3,r=0)

The model diagnostics is as follows:

Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	6.65	6	0.3541	-0.095	0.086	0.026	-0.101	0.046	-0.160
12	16.07	12	0.1881	0.188	0.069	0.102	0.153	0.004	0.035
18	30.56	18	0.0323	-0.205	0.081	-0.083	0.086	0.208	-0.049
24	33.16	24	0.1007	0.065	-0.039	-0.001	-0.050	-0.099	-0.003

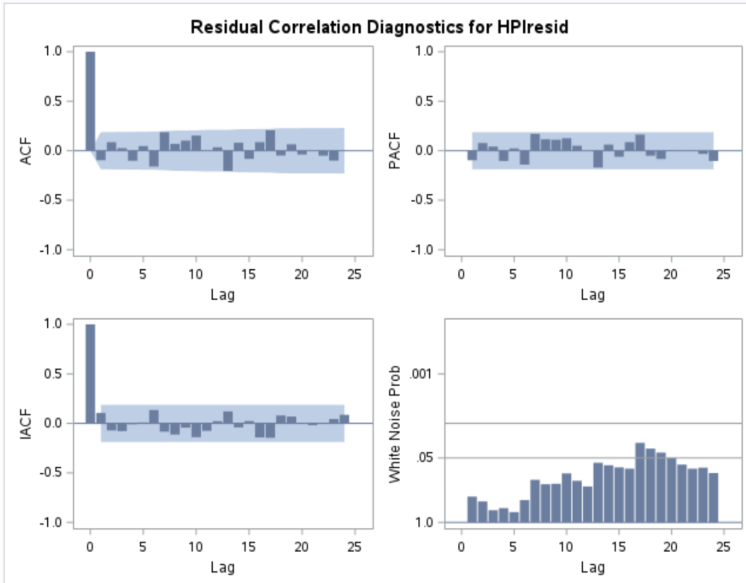

Figure 25: Model 1 Diagnostics

Clearly the Ljung-Box test has been failed as the p-value is less than 95% confidence statistic for lag 18 that implies the null hypothesis which states the residuals are white noise is rejected. We move forward to next model

3.5.2 Model 2 with pre-whitened data and transfer model (b=0,s=3,r=2)

The following model also rejects the the null hypothesis for Ljung Box test with a low p-value. Thus residuals are not white-noise.

The model diagnostics are as follows:

To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	9.57	6	0.1442	-0.069	0.133	0.076	-0.110	-0.024	-0.196
12	19.64	12	0.0743	0.207	0.044	0.081	0.167	-0.035	0.014
18	35.84	18	0.0074	-0.228	0.074	-0.115	0.062	0.206	-0.069
24	40.66	24	0.0181	0.064	-0.039	-0.049	-0.123	-0.103	0.002

Autocorrelation Check of Residuals

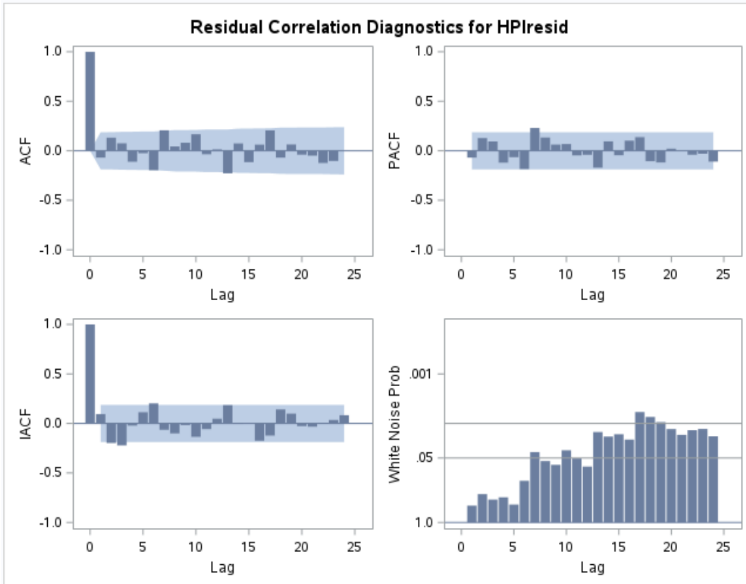

Figure 26: Model 2 Diagnostics

3.6 Model 3 with pre-whitened data, transfer model (b=0,s=3,r=1) and model fit to the noise

For this model we check the residual plots from Model1 and fit a model to the Noise term in our transfer function model.

Observing the residual plots we see negative significance levels at lags 6. We fit a factored MA model with q=6.

Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	2.21	5	0.8199	-0.055	0.094	0.045	-0.056	0.039	0.000
12	10.16	11	0.5165	0.138	0.095	0.083	0.162	0.048	0.009
18	25.71	17	0.0798	-0.195	0.119	-0.071	0.099	0.216	-0.051
24	27.46	23	0.2370	0.045	-0.024	-0.009	-0.031	-0.091	-0.013

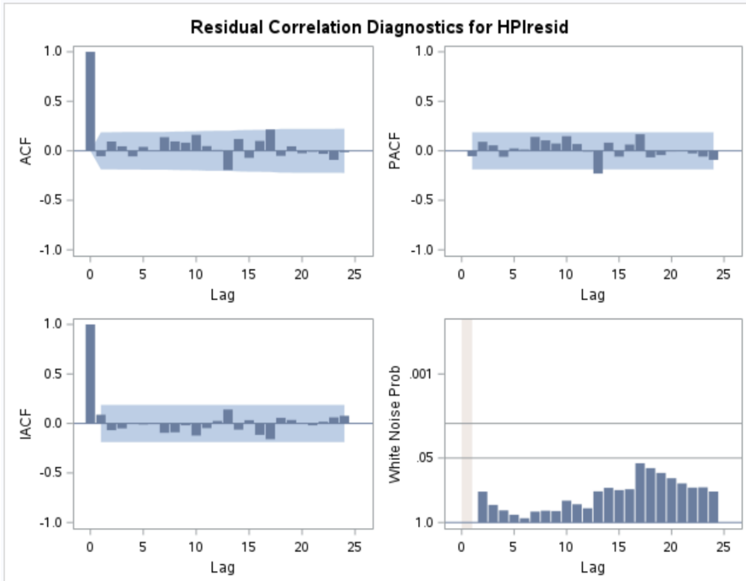

Figure 27: Model 3 diagnostics

This model fails to reject white noise for its residuals. The AIC and SBC values for the above 3 models are:

Table 1: Model comparison.

Model	AIC	SBC
1	97.44	111.12
2	93.70	110.12
3	96.16	112.57

Even though Model 3 has slightly higher AIC and BIC than Model 2, the latter rejects white noise for residuals for all lags.
Considering white noise test, AIC and SBC model checks we chose Model 3 as the champion model

4 Outlier Analysis

Before proceeding with the chosen model we also completed an outlier check

Outlier Detection Summary	
Maximum number searched	3
Number found	3
Significance used	0.01

Outlier Details				
Obs	Type	Estimate	Chi-Square	Approx Prob>ChiSq
103	Additive	-1.16793	23.22	<.0001
79	Shift	0.15316	23.72	<.0001
108	Additive	-1.04780	19.39	<.0001

Figure 28: Outlier Analysis

There are additive and shifting outliers present in out data set
We created three indicator variables to signify this outlier behavior:
AO: Additive outlier such that observation 103 is 1 and rest 0
AO1: Shifted outlier such that observation 108 is 1 and else 0
LS1: Shifted outlier such that observations beyond 79 are 1 and else 0

The differenced HPI Residual diagnostics were as follows:

Autocorrelation Check for White Noise									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	4.06	6	0.6687	0.038	0.014	-0.012	-0.072	0.019	-0.159
12	11.20	12	0.5122	0.142	0.057	0.062	0.163	-0.034	0.023
18	23.04	18	0.1890	-0.201	0.006	-0.064	0.097	0.179	-0.024
24	27.49	24	0.2822	0.038	-0.074	-0.020	-0.098	-0.112	-0.031

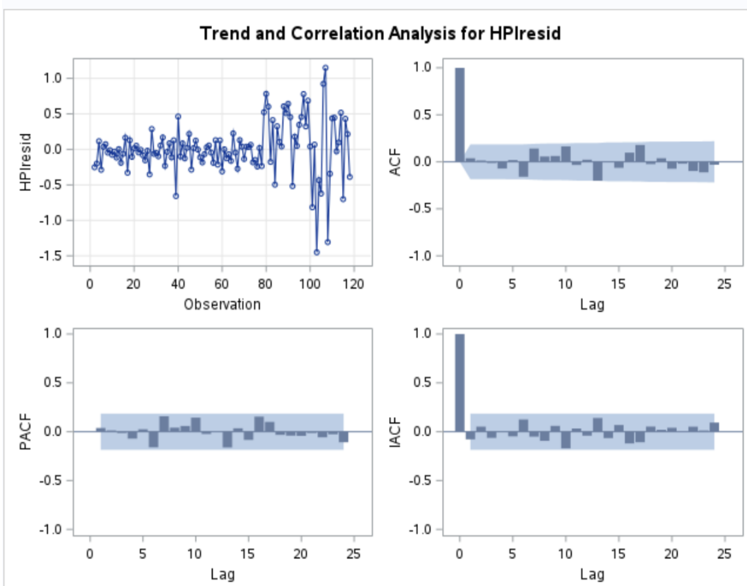

Figure 29: Pre-whitened HPI Model Residual Diagnostics

Based on negative spikes in PACF/ACF at lag 6 we decided to apply an MA model with just lag at 6. The model diagnostics were:

Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	13.49	5	0.0192	-0.182	-0.068	-0.007	-0.024	-0.272	0.014
12	19.70	11	0.0496	-0.028	0.141	-0.002	0.158	-0.054	-0.026
18	37.83	17	0.0026	-0.168	0.150	-0.145	-0.003	0.250	-0.033
24	40.72	23	0.0128	-0.026	0.034	-0.096	-0.029	0.001	-0.090

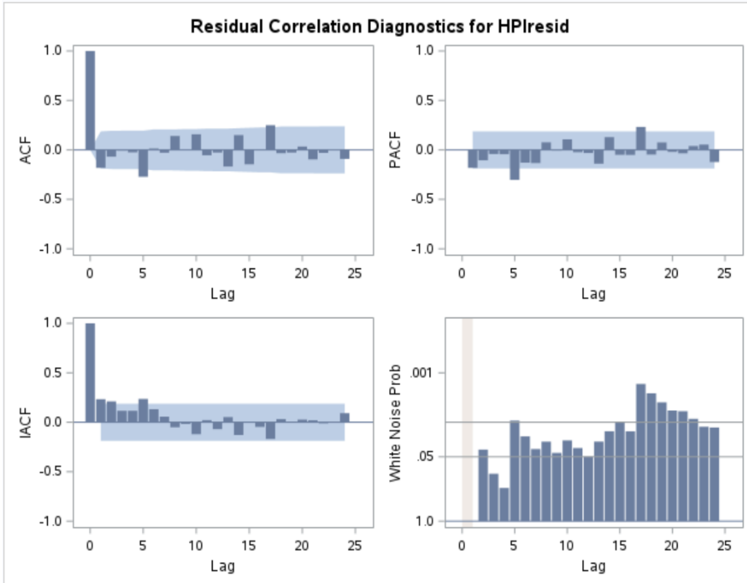

Figure 30: Outlier Model with q=(6) Residual Diagnostics

The white noise is rejected. We see a spike at lag 5 for PACF plot. We thus add AR terms until lag p=5. The model diagnostics are:

Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	.	0	.	0.006	-0.064	0.033	-0.034	0.020	0.002
12	4.82	6	0.5675	-0.088	0.077	0.038	0.043	-0.008	0.121
18	16.66	12	0.1630	-0.068	0.109	-0.120	-0.040	0.223	-0.075
24	24.33	18	0.1445	-0.138	-0.012	-0.085	-0.029	-0.022	-0.161

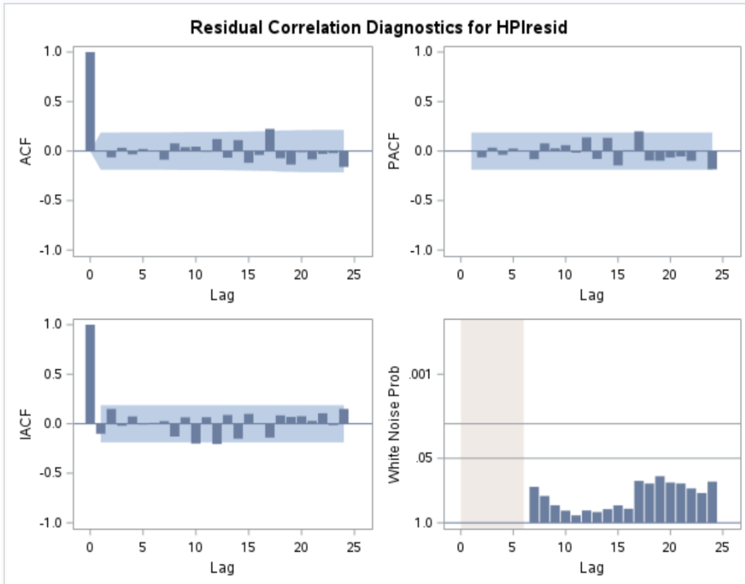

Figure 31: Outlier Model with p=5 and q=(6) Residual Diagnostics

The white noise is not rejected for residuals and the AIC and SBC for this model are found to be lowest of the 4. Additionally, no more outliers are found for this data. We term this model as Model4. The model comparison table is printed below:

Table 2: Model comparison including outlier model.

Model	AIC	SBC
1	97.44	111.12
2	93.70	110.12
3	96.16	112.57
4	33.72	72.04

We move forward with the outlier included model for forecasting and results

5 Forecasting

Using the parameters of Model 4 we forecast 10 months ahead HPI:

Forecasts for variable HPI				
Obs	Forecast	Std Error	95% Confidence Limits	
119	312.7875	0.4641	311.8779	313.6971
120	313.7713	1.2209	311.3783	316.1643
121	314.9192	2.2821	310.4464	319.3920
122	315.6596	3.4819	308.8351	322.4841
123	316.4550	4.8066	307.0342	325.8759

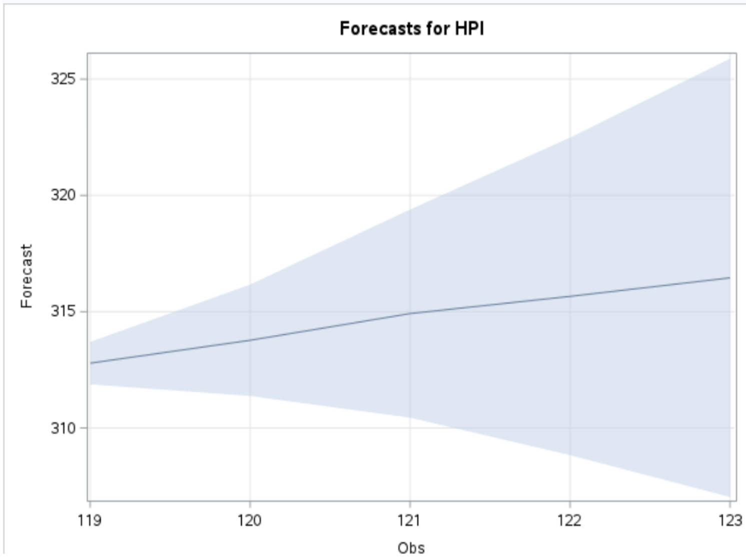

Figure 32: Housing Price Index Forecast for 10 months

An increasing trend of HPI is witnessed with falling rates as input. This seems in line with the recent trends in HPI. Although, the market was down for the last couple of months due to rate hikes by FED and overall low consumer sentiment, the recent quarter saw an increase in HPI to over 0.9% month over month.As the economy bounces back a further increase is expected

6 Results and Conclusion

The fitted model has the following parameter estimates and related statistics:

Conditional Least Squares Estimation							
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag	Variable	Shift
MU	-0.04808	0.0075210	-6.39	<.0001	0	HPIresid	0
MA1,1	0.56923	0.11016	5.17	<.0001	6	HPIresid	0
AR1,1	-0.34803	0.09874	-3.52	0.0006	1	HPIresid	0
AR1,2	-0.21914	0.10471	-2.09	0.0389	2	HPIresid	0
AR1,3	-0.16230	0.10338	-1.57	0.1196	3	HPIresid	0
AR1,4	-0.14923	0.10821	-1.38	0.1710	4	HPIresid	0
AR1,5	-0.45197	0.11047	-4.09	<.0001	5	HPIresid	0
NUM1	0.10904	0.15568	0.70	0.4853	0	Rateresid	0
NUM1,1	-0.03708	0.17584	-0.21	0.8334	1	Rateresid	0
NUM1,2	0.48263	0.14542	3.32	0.0013	3	Rateresid	0
DEN1,1	0.38916	0.20815	1.87	0.0645	1	Rateresid	0
NUM2	-1.66176	0.30669	-5.42	<.0001	0	AO	0
NUM3	-1.39595	0.28180	-4.95	<.0001	0	AO1	0
NUM4	0.26495	0.01662	15.94	<.0001	0	LS1	0

Figure 33: Conditional Least Square Estimates for Final Model

Although, some of the denominator and numerator terms seem to be insignificant there is still a causal effect seen with rate variable with low AIC and SBC values. Additionally, negative overall coefficient for numerator and denominator for Rates implies a slightly negative correlation between rates and HPI established earlier. It strengthens the argument in economics saying increasing mortgage rates will lead to a fall in housing price index due to the demand supply dynamics. Although, the relation between HPI and Rates are much complicated than a simple linear one as has been witnessed at various economic cycles in the country.

The outliers and noise fitted model parameters are significant implying the model is fit well. The dependency of HPI on its past values as well as the error terms leaves room for though about the randomness in its movement despite considering explanatory variables.

Overall, there can be more variables discovered having a much stronger correlation with HPI which can result in more significant parameter estimates. However, rates is still a contestable candidate for this position and serves well in predicting the trajectory of the housing market.

7 Appendix

7.1 Raw Data

Table 3: Raw Data

Date	HPI	30 year mortgage rates
12/01/2013	160.994	4.4575
01/01/2014	161.927	4.432
02/01/2014	162.526	4.3025
03/01/2014	163.086	4.3425

Continued on next page

Table 3 – continued from previous page

Date	HPI	30 year mortgage rates
04/01/2014	163.393	4.3375
05/01/2014	163.66	4.192
06/01/2014	164.062	4.1625
07/01/2014	164.577	4.13
08/01/2014	165.215	4.115
09/01/2014	165.905	4.1625
10/01/2014	166.642	4.036
11/01/2014	167.335	3.9975
12/01/2014	168.05	3.864
01/01/2015	168.634	3.67
02/01/2015	169.13	3.71
03/01/2015	169.8	3.77
04/01/2015	170.298	3.672
05/01/2015	170.881	3.84
06/01/2015	171.469	3.9825
07/01/2015	172.13	4.046
08/01/2015	172.937	3.905
09/01/2015	173.828	3.89
10/01/2015	174.792	3.796
11/01/2015	175.739	3.9425
12/01/2015	176.543	3.964
01/01/2016	177.274	3.8725
02/01/2016	177.649	3.66
03/01/2016	178.164	3.694
04/01/2016	178.767	3.605
05/01/2016	179.43	3.6
06/01/2016	180.082	3.568
07/01/2016	180.832	3.44
08/01/2016	181.852	3.435
09/01/2016	182.824	3.46
10/01/2016	183.751	3.47
11/01/2016	184.759	3.77
12/01/2016	185.722	4.198
01/01/2017	186.805	4.15
02/01/2017	187.315	4.1675
03/01/2017	187.993	4.196
04/01/2017	188.725	4.045
05/01/2017	189.617	4.01
06/01/2017	190.51	3.904
07/01/2017	191.452	3.9675
08/01/2017	192.665	3.88
09/01/2017	193.759	3.805
10/01/2017	194.804	3.895
11/01/2017	195.955	3.922
12/01/2017	197.172	3.95
01/01/2018	198.315	4.0325
02/01/2018	199.232	4.33
03/01/2018	199.966	4.444
04/01/2018	200.658	4.4675
05/01/2018	201.425	4.586
06/01/2018	202.233	4.57

Continued on next page

Table 3 – continued from previous page

Date	HPI	30 year mortgage rates
07/01/2018	202.913	4.5275
08/01/2018	203.689	4.55
09/01/2018	204.349	4.6275
10/01/2018	205.112	4.83
11/01/2018	205.669	4.866
12/01/2018	206.156	4.6375
01/01/2019	206.539	4.464
02/01/2019	206.862	4.37
03/01/2019	207.066	4.265
04/01/2019	207.513	4.1425
05/01/2019	208.136	4.072
06/01/2019	208.649	3.8025
07/01/2019	209.287	3.765
08/01/2019	210.093	3.616
09/01/2019	210.904	3.605
10/01/2019	211.791	3.688
11/01/2019	212.787	3.695
12/01/2019	213.933	3.72
01/01/2020	214.994	3.624
02/01/2020	215.864	3.465
03/01/2020	216.402	3.45
04/01/2020	216.812	3.306
05/01/2020	216.98	3.2325
06/01/2020	217.616	3.1625
07/01/2020	219.378	3.016
08/01/2020	222.391	2.935
09/01/2020	225.837	2.89
10/01/2020	229.753	2.834
11/01/2020	233.208	2.765
12/01/2020	236.486	2.684
01/01/2021	239.56	2.735
02/01/2021	242.369	2.81
03/01/2021	245.465	3.0825
04/01/2021	249.07	3.06
05/01/2021	253.407	2.9625
06/01/2021	258.358	2.975
07/01/2021	262.82	2.868
08/01/2021	266.845	2.8425
09/01/2021	270.377	2.9
10/01/2021	273.725	3.0675
11/01/2021	277.21	3.0675
12/01/2021	281.342	3.098
01/01/2022	285.924	3.445
02/01/2022	291.153	3.7625
03/01/2022	296.445	4.172
04/01/2022	300.573	4.9825
05/01/2022	303.762	5.23
06/01/2022	304.724	5.522
07/01/2022	303.879	5.4125
08/01/2022	301.473	5.2225
09/01/2022	299.353	6.112

Continued on next page

Table 3 – continued from previous page

Date	HPI	30 year mortgage rates
10/01/2022	298.873	6.9
11/01/2022	298.269	6.805
12/01/2022	297.413	6.364
01/01/2023	297.03	6.2725
02/01/2023	297.537	6.2575
03/01/2023	298.637	6.544
04/01/2023	300.213	6.3425
05/01/2023	302.566	6.425
06/01/2023	304.593	6.714
07/01/2023	306.767	6.84
08/01/2023	309.155	7.072
09/01/2023	311.175	7.2

7.2 Codes

7.3 Python Code - EDA

```
import pandas as pd
import numpy as np
import seaborn as sb
import matplotlib.pyplot as plt

correlation = house_data['Rate'].corr(house_data['HPI'])
# Measuring the linear trend
print(f"The correlation between our interest rate and HPI is",round(correlation,3))

sb.regplot(x=house_data['Rate'], y=house_data['HPI'], data=house_data, fit_reg=True)
plt.xlabel("30 Year Mortgage Rate")
plt.ylabel('HPI')

from statsmodels.tsa.seasonal import seasonal_decompose
# Decompose our input variable into trend, seasonal, and residual components to identify u
result = seasonal_decompose(house_data['Rate'], model='additive', period=12)
result.plot()
plt.show()
# Decompose HPI into trend, seasonal, and residual components to identify underlying patte
result = seasonal_decompose(house_data['HPI'], model='additive', period=12)
result.plot()
plt.show()

plt.figure(figsize=(8, 6))
sb.histplot(house_data['Rate'], kde=True, color='blue')
plt.title('Distribution of 30 year Mortgage Rates')
plt.xlabel('Mortgage Rates')
plt.ylabel('Frequency')
plt.show()

plt.figure(figsize=(8, 6))
sb.histplot(house_data['HPI'], kde=True, color='red')
plt.title('Distribution of HPI')
plt.xlabel('HPI')
plt.ylabel('Frequency')
```

```
plt.show()

house_data['Rate'].plot(kind='box', vert=False, figsize=(8, 6), patch_artist=True)
plt.title('Boxplot of 30 year mortgage rates')
plt.xlabel('30 year mortgage rates')
plt.show()
house_data['HPI'].plot(kind='box', vert=False, figsize=(8, 6), patch_artist=True)
plt.title('Boxplot of HPI')
plt.xlabel('HPI')
plt.show()
```

7.3.1 Pre-whitening X Code

```
FILENAME REFF '/home/u63656218/housedata.xlsx';

PROC IMPORT DATAFILE=REFF
DBMS=XLSX
OUT=home;
GETNAMES=YES;
RUN;

proc contents data=home;
run;

proc arima data=home;
/* identify var= Rate; */
/* identify var= Rate(1); */
/* identify var= Rate(2); */

identify var= Rate(1) stationarity=(adf=1);

/* estimate p=5 q=1; */
estimate p=5 q=5; /*final estimated model


run;
```

7.3.2 Pre-whitening Y, Transfer model and Outlier Analysis Code

```
FILENAME REFF '/home/u49601401/sasuser.v94/housedata.xlsx';

PROC IMPORT DATAFILE=REFF
        DBMS=XLSX
        OUT=home2;
        GETNAMES=YES;
RUN;

proc contents data=home2;
run;

proc arima data=home2 out=rateresid;
identify var= Rate(1) stationarity=(adf=1);
estimate p=5 q=5; *prefered model;
```

```
forecast out=resid2(keep=residual);
run;

proc arima data=home2;
identify var=HPI stationarity=(adf=2);
run;

proc arima data=home2;
identify var=HPI(1);
run;

proc arima data=home2 out=model1;
identify var= HPI(1) ;
estimate p=2;
forecast out=resid1(keep=residual);
run;

data resid1;
  set resid1(rename=(residual=HPIresid) obs=118);
  Obs = _N_;
run;

data resid2;
  set resid2(rename=(residual=Rateresid) obs=118);
  Obs = _N_;
run;

data merged;
merge resid1 resid2;
by Obs;
run;

proc arima data=merged;
identify var=HPIresid;
run;

proc arima data=merged;
identify var= HPIresid crosscorr=(Rateresid);
*estimate input=(0$(1,3)/(1,0)rateresid);
*estimate input=(0$(1,3)/(1,2)rateresid);
*estimate q=(6) input=(0$(1,3)/(1,2)rateresid);
estimate q=(6) input=(0$(1,3)/(1,0)rateresid);
outlier alpha=0.01;
run;

*Outlier analysis;

data merged1;
set merged;
if _n_ = 103 then AO = 1;
else AO = 0.0;
if _n_= 108 then AO1=1;
else AO1=0;
if _n_ >= 79 then LS1=1;
```

```
else LS1=0;
run ;

proc arima data=merged1;
identify var= HPIresid crosscorr=(Rateresid AO AO1 LS1);
estimate q=(6) input=(0$(1,3)/(1,0)Rateresid AO AO1 LS1);
estimate q=(6) p=5 input=(0$(1,3)/(1,0)Rateresid AO AO1 LS1);
forecast lead=10 alpha=0.05;
outlier alpha=0.01;
run;
```