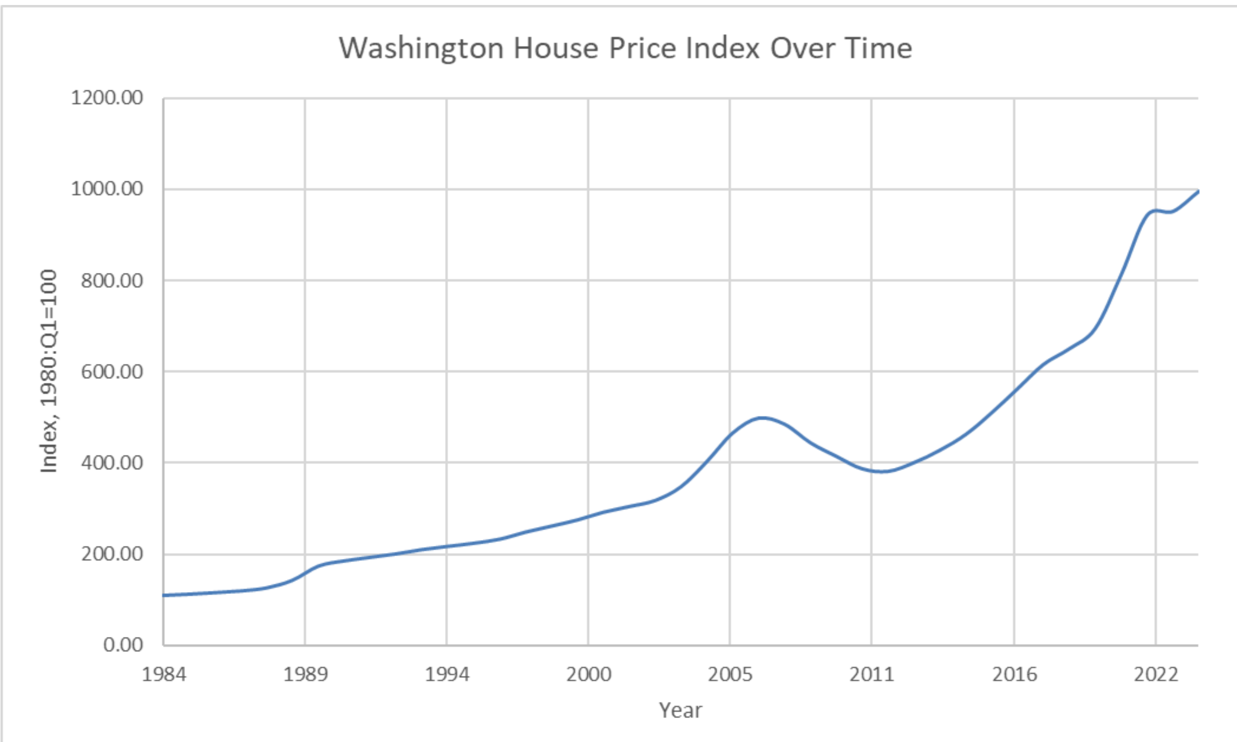
Forecasting the Housing Price Index for Washington State

Caden Brown

December 2025

Introduction

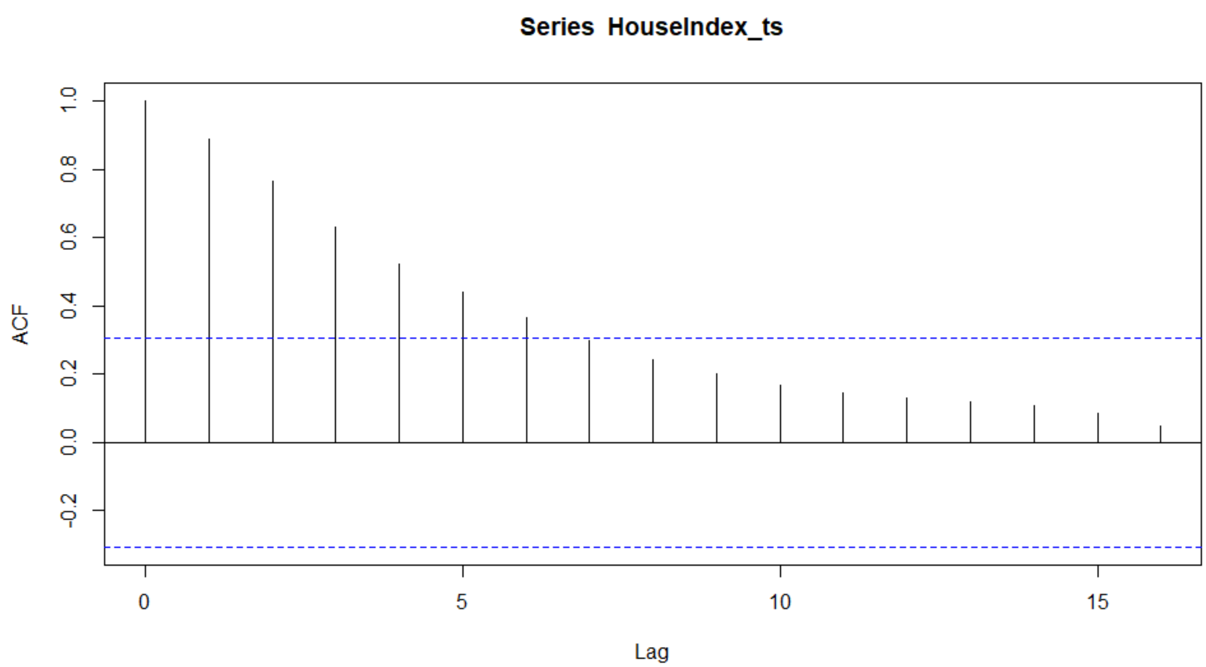
The time series I will be forecasting is the All-Transactions House Price Index in Washington. The index is estimated using sale prices and appraisal data of houses in Washington state. The data for this forecast is provided by the Federal Reserve Bank of St. Louis (FRED), sourced directly from the U.S. Federal Housing Finance Agency. The data used for this forecast ranges from 1984 to 2024, reported annually. There are no gaps in the time series; values for all time periods are accounted for.



Over time, the House Price Index in Washington has an upwards trend. On average, the index exhibits a constant, positive trend upwards over each year. However, between the years of 2005 and 2011, there is a steep increase that deviates from the general trend followed by a steep decrease. The steep increase can be attributed to the housing bubble, which peaked in 2007. When there was greater access to purchasing a home through subprime mortgages, housing prices became much more inflated. The subsequent decrease in 2008 was the collapse of these markets as mortgage delinquencies skyrocketed from the subprime mortgages. As foreclosures rose and government restrictions decreased the inflated access to purchasing a home, housing prices fell significantly.

Univariate Model

The graph below models the autocorrelation function (ACF) of the housing price index. As we can see, there is a smooth decay of effects. There are no irregularities or spikes throughout the function. In this instance, it would be best to use an autoregressive process to forecast into the future.

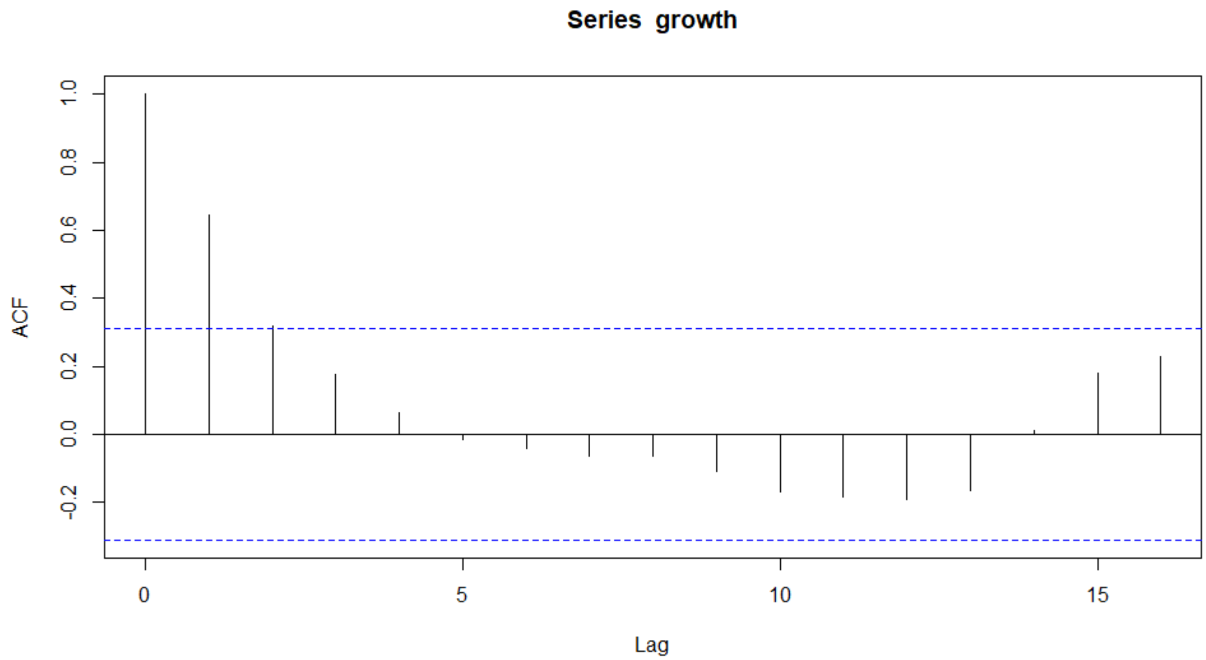


I conducted an Augmented Dickey-Fuller (ADF) test with one lag on the time series for the housing price index to check if the time series is non-stationary. The results gave a test statistic of 0.737 with a 5% level critical value of -2.93. Because the test statistic is more than the critical value, we cannot reject that the time series is non-stationary. However, after taking the first difference of the data, the results give a test statistic of -2.977 with a 5% level critical value of -2.93. Using the differenced time series, we can reject that the time series is non-stationary. As a result of the ADF tests, I will be using the first difference of the time series in order to forecast. The first differenced time series represents the change in the housing price index from the previous year.

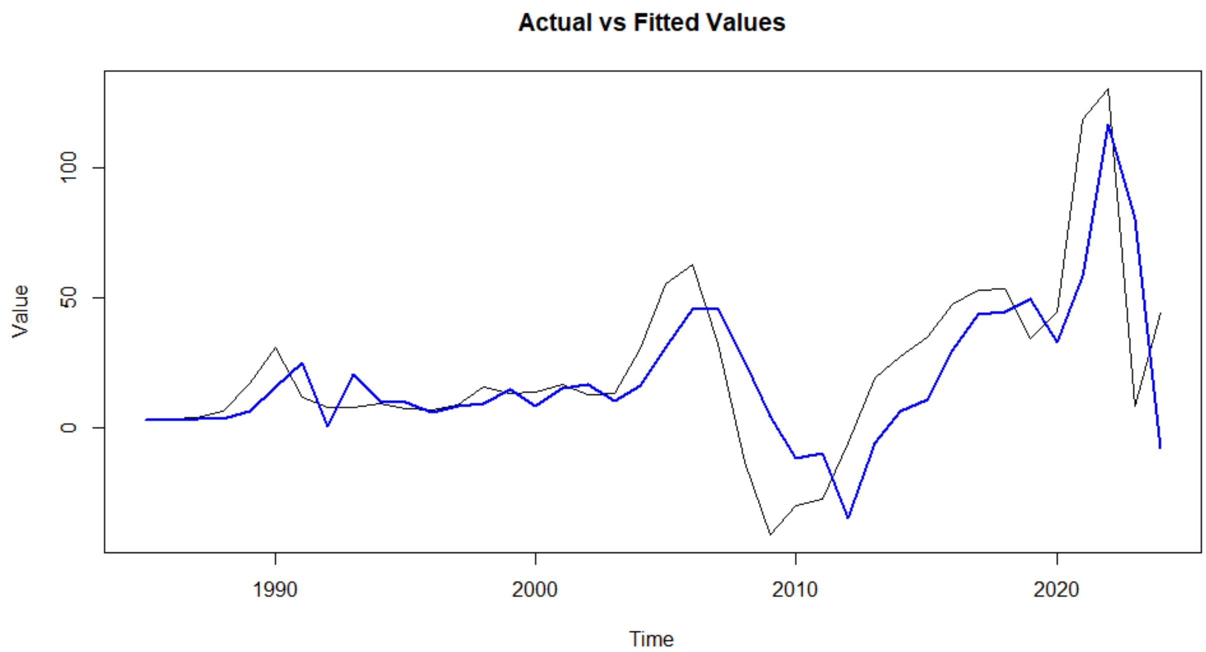
I estimated a SARIMA model based on the first differenced time series. Using *auto.arima* in R, the order model selected was (0,1,0) with an AIC of 370.94. However, after manually adjusting the lags of AR and MA, we can see that some of these models yield lower AICs than the automatic model.

ARIMA(p,d,q)	AIC
auto.arima: (0,1,0)	370.94
(0,1,1)	372.78
(1,1,0)	372.82
(1,1,1)	**369.79**
(2,1,0)	**363.86**
(0,1,2)	**364.38**

Based on these results, the most suitable model for the differenced time series would be an ARIMA(2,1,0) process. This result is also consistent with the graph of the original ACF and the ACF of the differenced time series, as shown below.

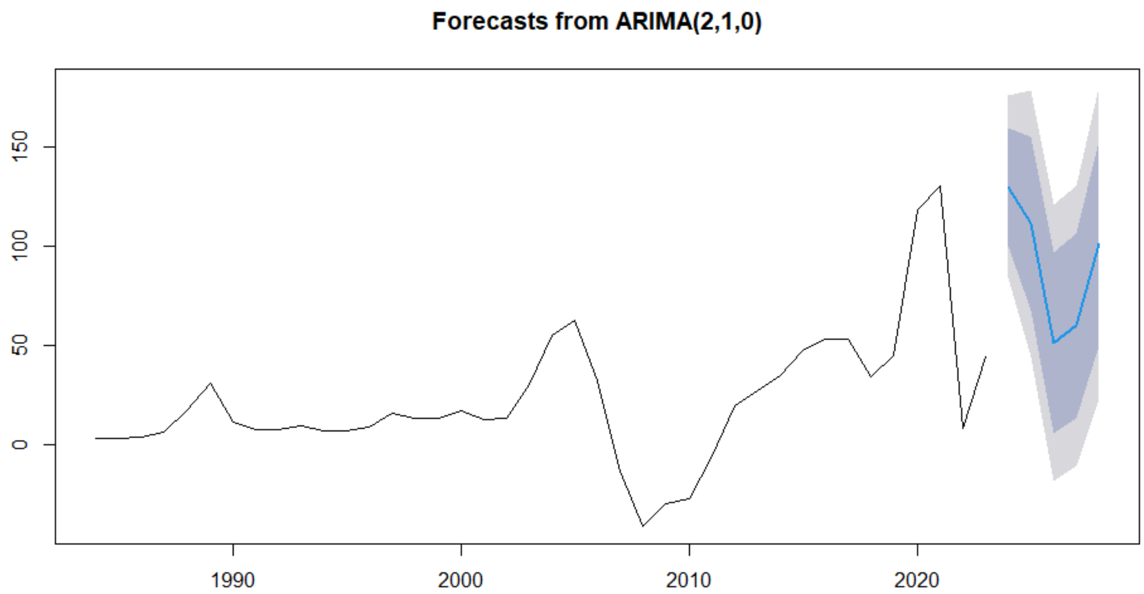


The graph below compares the fitted forecasts to the actual data based on the ARIMA (2,1,0) model. The blue line shows the one step ahead fitted values from the ARIMA(2,1,0) model, while the black line shows the actual change in the housing price index. The fitted series forecasts the overall level and turning points of the data reasonably well, although it smooths out some of the sharper movements, especially around the large boom/bust periods. The pattern is consistent with the model of the differenced series and suggests the ARIMA(2,1,0) captures the main dynamics of the housing index.



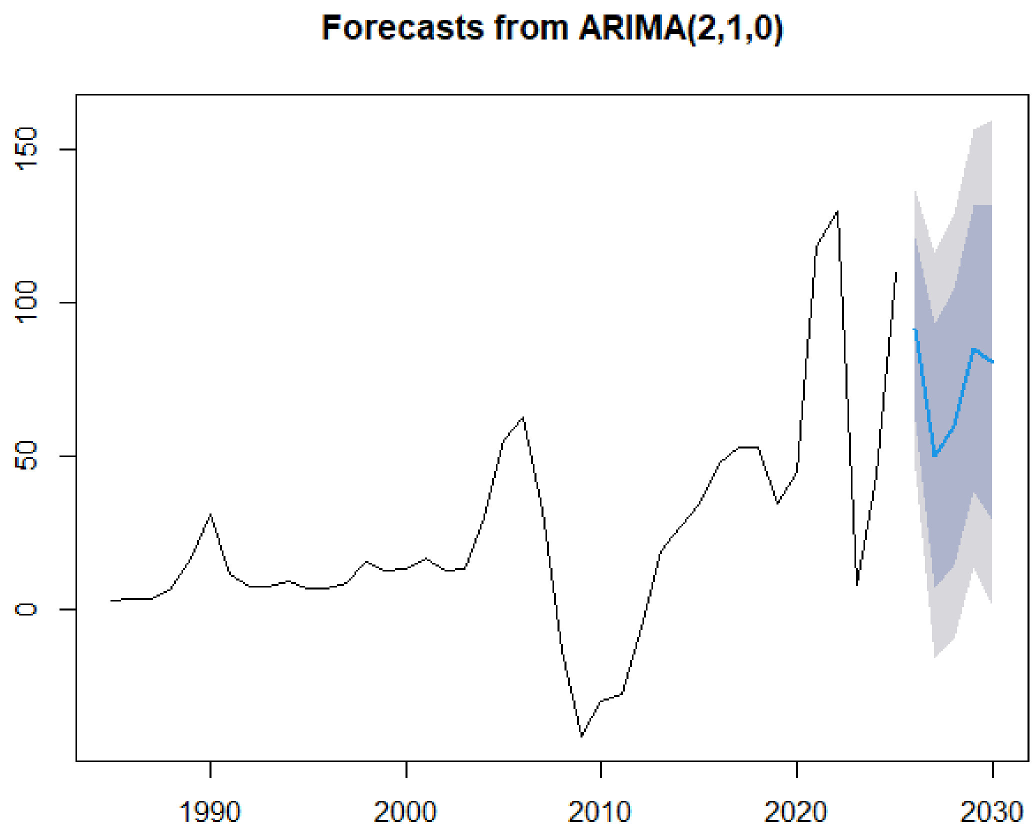
Forecasts from the Univariate Model

From the forecasted results below, the 95% confidence intervals are fairly wide. This indicates that there is a substantial amount of uncertainty, which increases every year forward. Even the one year ahead forecast for 2025 spans from about 84 to 176. This is a range of nearly 92 index points, which is large relative to the scale and historical variability of the housing price index. The wide interval suggests that the model considers both a very large positive change and a much smaller increase as possible outcomes. Since these intervals refer to annual growth, the wide range reflects substantial uncertainty about the direction and magnitude of year to year movements in the housing price index. As a result, the univariate forecast is not reliable enough to be useful, as the model cannot confidently distinguish between periods of strong growth and much weaker performance.

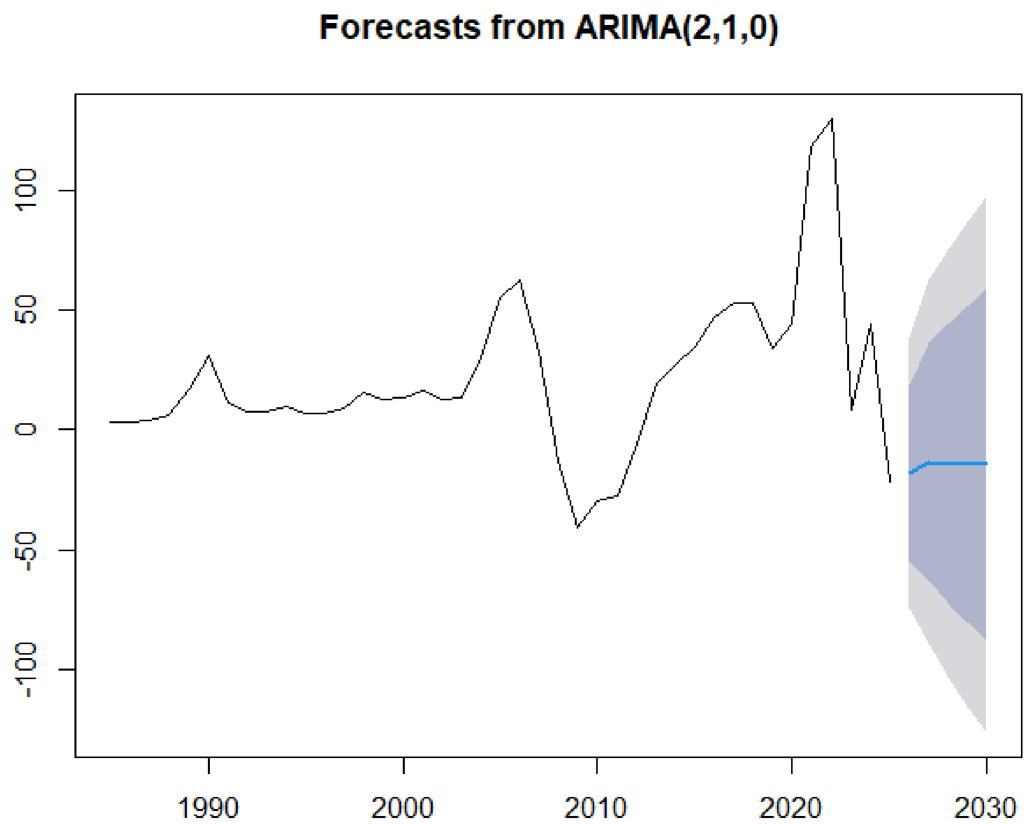


Forecasted Period	Point Forecast	Lower 95%	Upper 95%
2025	129.86	83.99	175.72
2026	111.44	44.28	178.60
2027	51.39	-18.11	120.89
2028	59.82	-10.99	130.62
2029	101.52	22.72	180.31

To expand on this forecast, we are going to add a new value to the end of the time series that is equal to the very last period plus two standard deviations. This experiment will be used to understand how a positive shock will impact the forecast. Now, running the same SARIMA model for five years forward, we have the result below. Comparing the positive shock to the original forecast, it has its biggest impact on the first year forecast then starts to weaken. In the new forecast the first period is noticeably lower than in the baseline forecast. In the second period, there is also a noticeable difference. However, after the second forecasted period the paths begin to converge. Overall, the two standard deviation "shock" has meaningful effects on the forecast for about two years.



Alternatively, we can run the same exact experiment to understand how a negative shock will impact the forecast. Instead of adding two standard deviations, we will subtract two standard deviations. The result of subtracting two deviations is graphed below. Introducing a large negative shock results in a large downward shift in the forecast. The first forecasted value is much lower than the baseline which shows the model's sensitivity to the most recent data point. The second period ahead is also lower than in the baseline, but there was an increase from the first forecasted period. Similar to the earlier experiment however, after about two periods the effect begins to die out and converge toward the original trajectory.



Multivariate Model

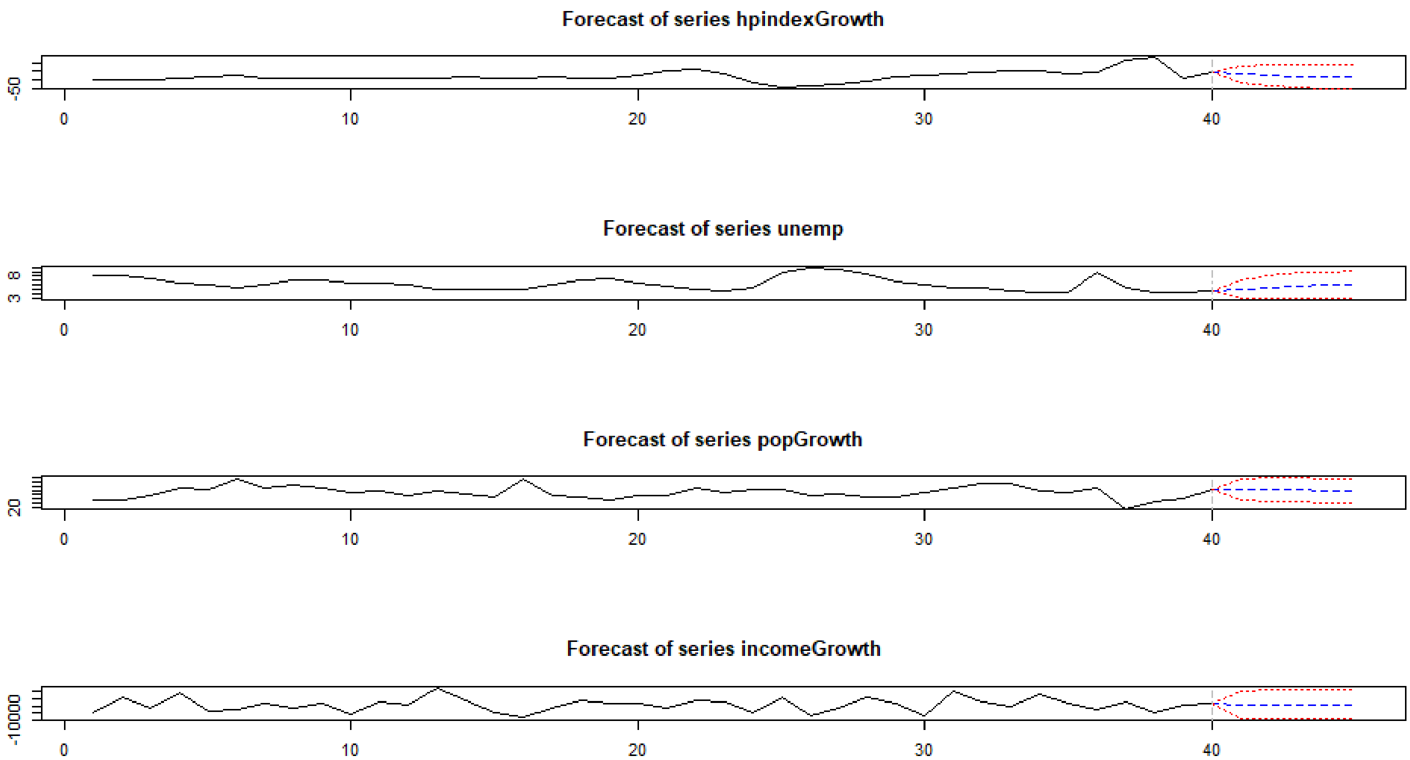
Building on from the univariate model, I will be adding three key variables to assist in creating a more accurate forecast of the housing price index. The three time series are Unemployment Rate in Washington, Resident Population in Washington, and Real Median Income in Washington. All time series are reported on a yearly basis and will be in the same time frame as the housing price index. The data for these variables have been sourced from FRED. Specifically, the unemployment rate is provided by the U.S. Bureau of Labor Statistics, and the resident population and real median household income are provided by the U.S. Census Bureau. Changes in unemployment directly affect household financial stability and the ability to purchase or maintain a home. Adding the unemployment rate makes it a useful indicator for the shifts in housing demand. Rising unemployment tends to weaken the market and can put downward pressure on housing prices. Population growth increases the number of potential homebuyers which raises demand for housing and influences pricing. Including population will help capture long term demographics that can drive sustained movements in the housing market. Real median household income reflects households' purchasing power and affordability in the housing market. Higher income levels generally support stronger housing demand and can cause price appreciation, while declining incomes can constrain the market.

Due to this model being multivariate, before constructing a vector autogression (VAR), I checked whether the new variables in the time series were stationary using the ADF test. The test statistic for unemployment was -3.33, population was -1.05, and income was -1.28 all with a 5% critical value of -2.93. Due to population and income not passing the test, I took the first difference of both variables. Running the ADF test again, the test statistics of the first differences of population and income were -3.43 and -5.19 respectively. As a result of this, the forecast will have the first difference of the housing price index, population, and income (all can be interpreted as growth), while unemployment remains as its original value. Running the VAR, I got the following correlation matrix.

The results from the correlation matrix below show that the housing price index growth is most strongly related to changes in unemployment with a negative correlation of -0.363. This indicates that higher employment is associated with slower growth of the housing price index. Alternatively, lower employment aligns with stronger growth of the housing price index. The correlations between the housing price index growth and both population growth (-0.012) and income growth (0.052) are extremely small. This suggests there is little to no linear relationship on a year to year basis between the growth of the housing price index and these two variables.

Correlation Matrix	hpindexGrowth	unemp	popGrowth	incomeGrowth
hpindexGrowth	1.000	-0.363	-0.012	0.052
unemp	-0.363	1.000	-0.062	-0.180
popGrowth	-0.012	-0.062	1.000	-0.116
incomeGrowth	0.052	-0.180	-0.116	1.000

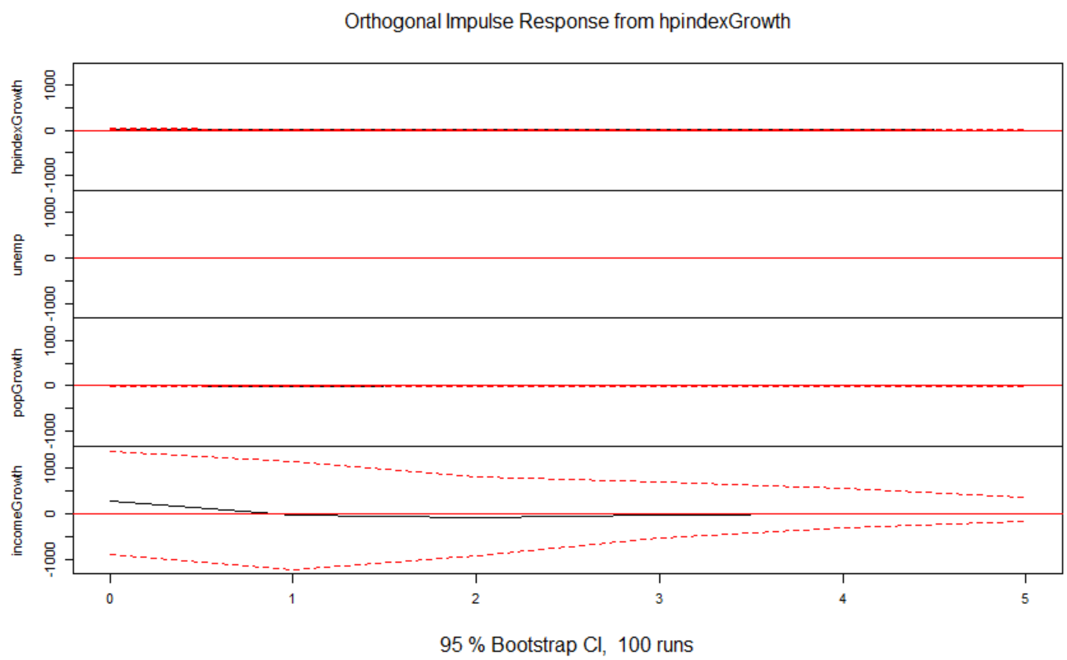
The plot and table below show the forecast for each variable from the VAR model, along with the specific values of the point forecast for the growth of the housing price index. The VAR model produces much smaller forecasts of the housing price index, both in terms of the point estimates and the confidence intervals. The SARIMA forecast suggested very large swings in growth over the next five years, indicating that the univariate model suggests strong persistence and volatility from previous periods. In contrast, the VAR model generated much less volatility in the forecast, with growth of the housing price index gradually declining from about 33 in 2025 to 22 in 2029. The confidence intervals are much narrower in magnitude as well. Overall, the VAR forecasts appear to be much more conservative and plausible.

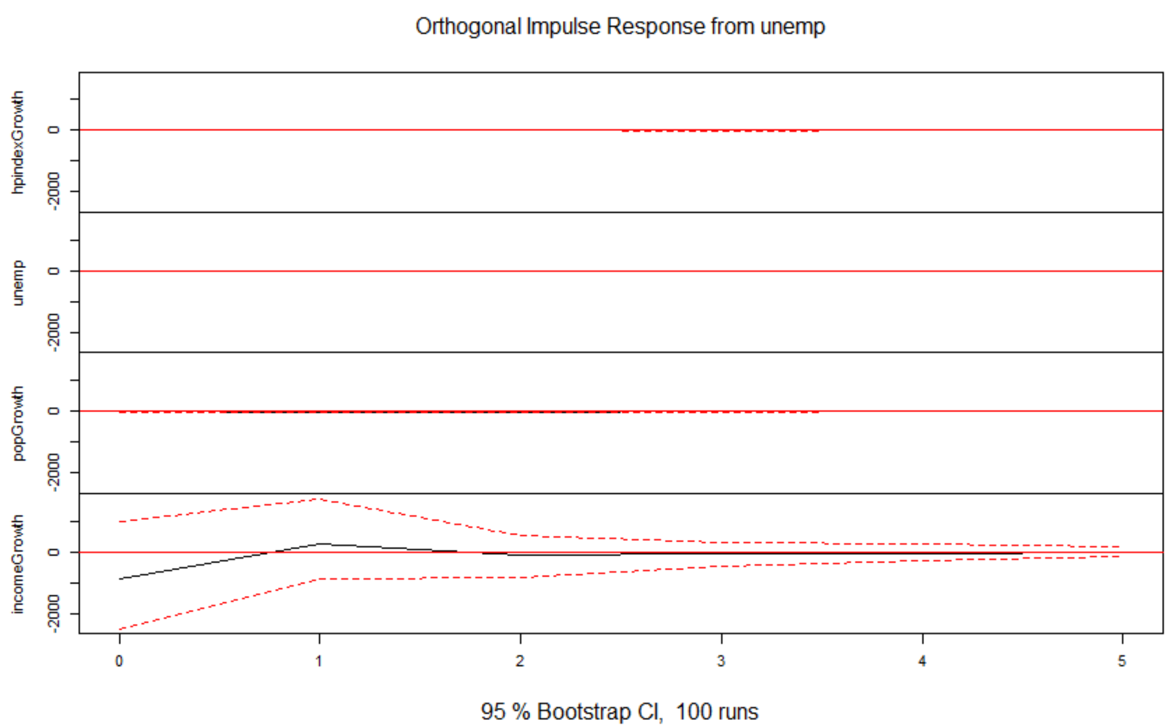


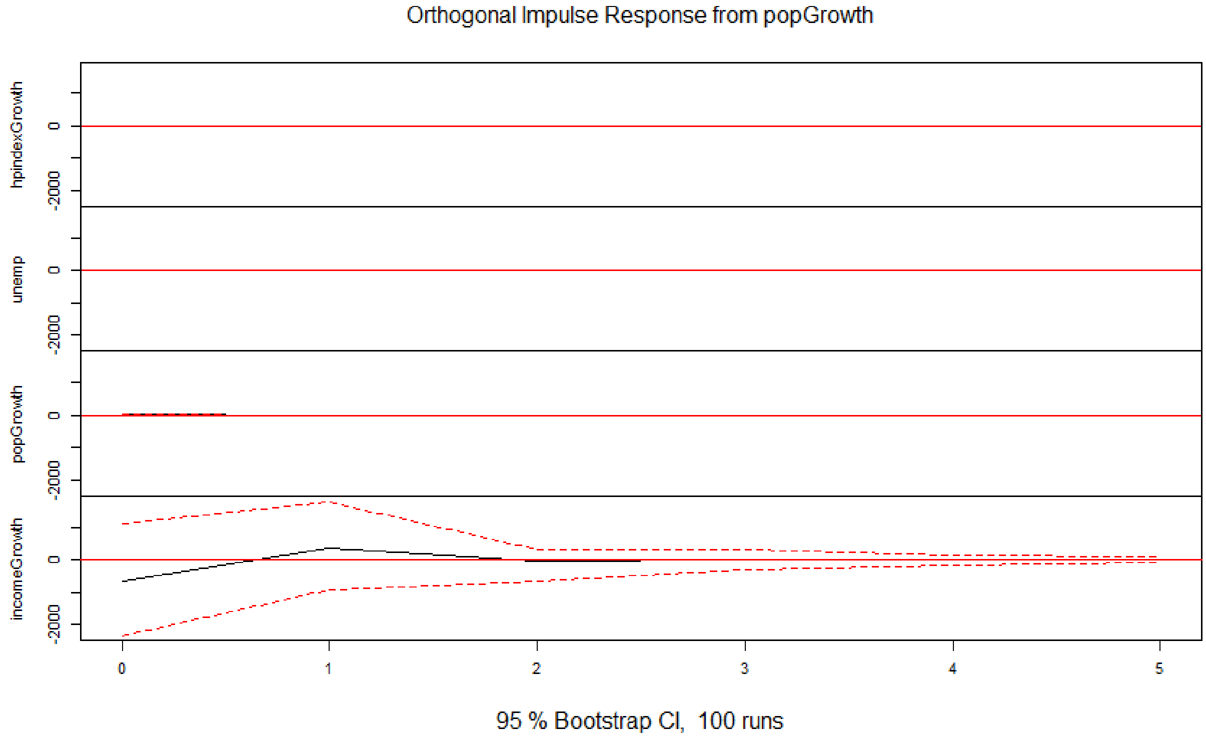
Point Forecast of hpindexGrowth

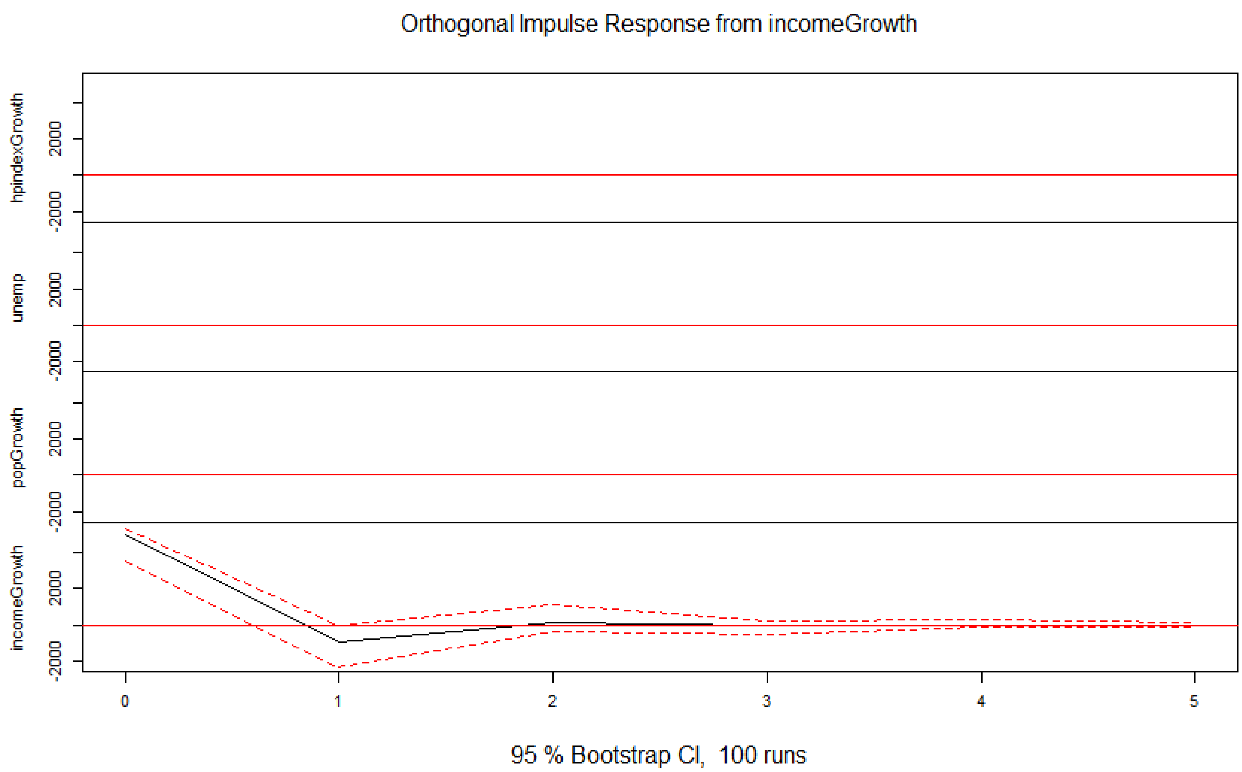
Forecasted Period	Point Forecast	Lower 95%	Upper 95%
2025	32.76	-16.76	82.28
2026	25.62	-37.41	88.65
2027	22.62	-44.72	89.96
2028	21.71	-46.74	90.15
2029	21.75	-46.90	90.41

Below are the Impulse Response Functions (IRF) of each time series. As shown in our correlation matrix, the unemployment rate has the greatest effect on the growth of the housing price index. Unemployment does not seem to be affected by any time series. Population growth is affected most by the growth of the housing price index. Income growth is affected most by unemployment.






Orthogonal Impulse Response from popGrowth
hpindexGrowth
unemp
popGrowth
incomeGrowth
0
-2000
0
1
2
3
4
5
95 % Bootstrap CI, 100 runs


Orthogonal Impulse Response from incomeGrowth
hpindexGrowth
unemp
popGrowth
incomeGrowth
2000
-2000
0
1
2
3
4
5
95 % Bootstrap CI, 100 runs