Volume 3 Project

Lillian Bartholomew, Connor Franckowiak, Jonathan Merrill,
Reece Robertson, Ben Summers

April 2, 2022

Abstract

We analyze Utah housing data in an attempt to determine if authorizing more housing will decrease the housing price. We use several models to attempt to predict this, including ARIMA, VARIMA, "ARIMA Forest" (a model of our own creation), and Kalman Filter. Several experiments were not successful, and those that were successful found that housing authorized has minimal impact on housing price. This may reflect reality, but more likely it is an indicator that our models are not working properly.

Problem Statement and Motivation

In recent years, the housing market in Utah has taken off like a rocket, striking fear in the hearts of the younger generation who worry if they will ever be able to afford a home. The goal of this project is to investigate what can be done by the state to make housing more affordable for those looking to buy. Specifically, it seems likely that if Utah authorized more housing than the price of housing on the market would decrease. This is a time series-focused project, as each quarter of housing costs and other variables will influence the future supply of housing.

Questions to Answer

As said above, we hypothesize that authorizing more houses will decease the price of the housing of the market. Thus, our primary question is, will changing how much new housing is authorized in Utah affect the median house price? However, we admit that there may be other economic factors more closely related to the price of housing. This leads to a secondary question of what economic factors can be influenced by the state to make houses more affordable?

Data

All of our datasets came from FRED (Federal Reserve Bank of St. Louis) which keeps track of a lot of economic indicators for the United States. Since we didn't know exactly which variables would be predictive of housing supply and price, we gathered many variables that we thought may be relevant. These include: household consumption spending, income inequality, active listings for Salt Lake or Utah Counties, median price per square foot, number of SNAP (food stamp) recipients in the state, median days a house is on the market, new building prices issued, GDP, population, housing price index, per capita median income, unemployment rate, educational attainment, net migration flow into Utah and Salt Lake Counties, the single family households, and housing authorized. Important features will be explained in more detail as they are used.

Many of these metrics were missing several years of data and many had yearly or monthly data as opposed to quarterly. For many of the variables (such as population), we extrapolated the data between years to create quarterly data (i.e., if the population was 100 one year and 200 the next, the second quarter would see a population value of 125, etc.). For other variables that were monthly, we summed or averaged the values over the quarter to make them quarterly. Additionally, to deal with missing data, we limited the number of years we examined in some of our models. Again, the more details will be given as we discuss each model and the data it used.

Methods and Results

Before we begin a discussion of the different models we used to answer our questions, we have one note to make. 2020 was an exceptional year in many regards, and each feature reflects this. Unemployment rate, for instance, spiked from 3% to 10% almost overnight, and then returned to normal over the next few months. As such, we decided not to include 2020 in the training data for our models.

ARIMA

Before we began our investigation of how other features affected the housing price, we thought that it would be useful to gain some understanding about the housing price considered in isolation. As such, we began our investigation into with a simple ARIMA model. We identified the `Housing Price Index` (`hpi`) as a feature that is useful in measuring affordability. The `hpi` measures the percentage of change in housing price from a specific start date. Thus, if the housing price index increases rapidly this may indicate a time in which housing is becoming unaffordable. Entering the market at a favorable time could make the difference in affordability for some. Predicting this feature with accuracy will be the focus on this and all subsequent models we employ.

We trained an ARIMA model on a one-step forward difference for the `hpi` data from 1997 to the end of 2015, then used the model to predict the difference through the start of 2020. In Figure 1 you see the result of this experiment, with the ground truth in black and the predictions in red. Observe that the model generally predicted the up-and-down trend of the data accurately at each step in this interval, although it certainly underestimated the degree of change each time and included a general downward trend after 2015 that was not observed in reality. Thus, this model gives a home buyer some idea of the future movement, however it cannot be trusted to reflect the full extent of reality. This limits its applicability to real scenarios.

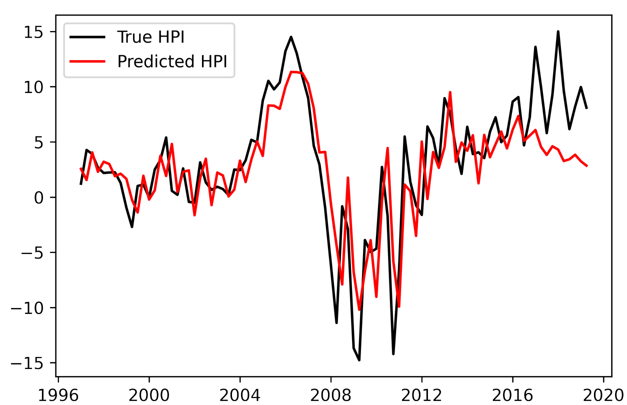
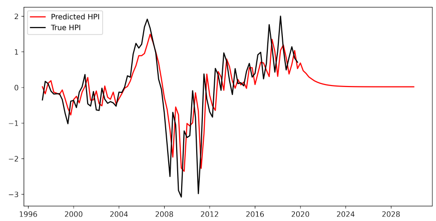


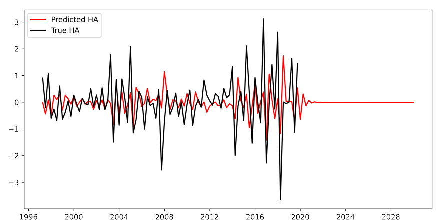
Figure 1: ARIMA Predictions on Change in `hpi`

VARIMA

Following our implementation of an ARIMA method, we created a VARIMA model. This model predicted change in the `hpi` using both the `hpi` itself and the number of `housing authorized` feature. We trained the model on both features, using data from 1997 to the end of 2019. We then predicted values from 2020 through 2030. The results are shown in Figure 2.



(a) VARIMA Predictions on Change in `hpi`

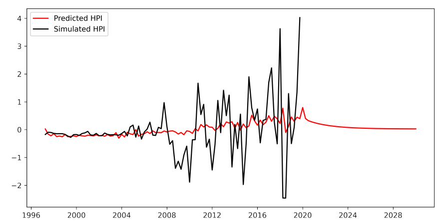


(b) ARIMA Predictions on Change in `housing authorized`

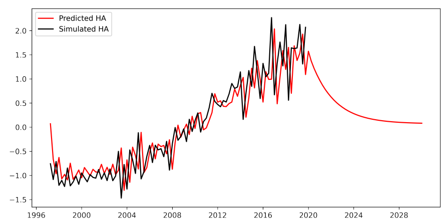
Figure 2: VARIMA Model on True Data

Afterwards, we artificially altered the `housing authorized` data to steadily increase the amount of `housing authorized` over time. To avoid assuming that altering `housing authorized` has no effect on the `hpi`, we randomly perturbed the `hpi` data proportional to the change in `housing authorized`. To do this, we computed the effect of `housing authorized` on `hpi` at each time step, then randomly selected one of these values for each time step in which we altered the `housing authorized`. We then multiplied that perturbation by the amount we increased the housing authorized for that time step, and added the product to the `hpi`. We then repeated the experiment on this adjusted data, the results of which are given in Figure 3.

We see that drastically increasing the amount of housing authorized has little effect on the `hpi`. In both cases the predicted `hpi` tended to 0 over time. To be certain of our result, we repeated this experiment with a one-hundred fold increase in the artificial adjustment of the housing authorized and saw similar results. This can mean one of two things. First, it could be that changing the housing authorized really does have a negligible impact on the `hpi`. Second, this could be indicative that this model does not accurately reflect reality. The truth is probably somewhere between these two



(a) VARIMA Predictions on Change in `hpi`



(b) ARIMA Predictions on Change in `housing authorized`

Figure 3: VARIMA Model on Artificially Adjusted Data

statements, as it is implausible that changing the housing authorized has so little an effect on the `hpi`, however it may not have as large an impact as we originally supposed.

ARIMA "Forests"

ARIMA is a time series method that can be helpful in understanding underlying data, so we wondered if it might be possible to combine ARIMA with some of the methods we learned last semester. By so doing we created a new model which we have not learned about in class. An "ARIMA Forest," as we call it, creates an ARIMA model to predict future values for each feature, then uses a Random Forest Regressor on those predicted feature values to estimate values of the target variable.

The results of this were unsatisfactory, but provide for some other interesting applications. Similar to the VARIMA model, we can artificially change the data for any given feature and predict the effect this has upon the model. As before, we change `housing authorized` to see the effect this might have upon the `hpi`. However, one advantage of the ARIMA Forest method is that because each feature is independently predicted, we need not introduce random perturbations back into other features when changing the `housing authorized`.

The result of this can be seen in Figures 4 and 5. Unfortunately, it seems that even when we widely vary `housing authorized`, from normalized values 1 to -1, almost no change is produced.

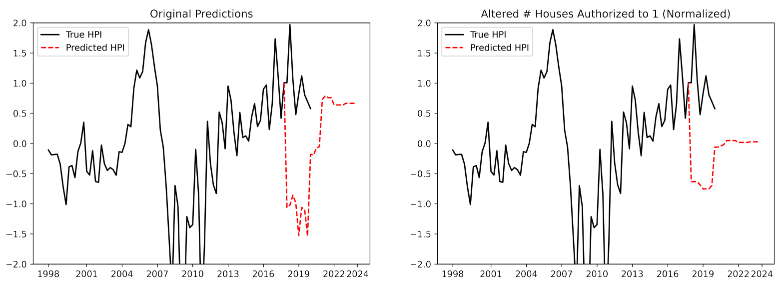


Figure 4: ARIMA Forest Predictions of `hpi` with and without increasing `housing authorized`

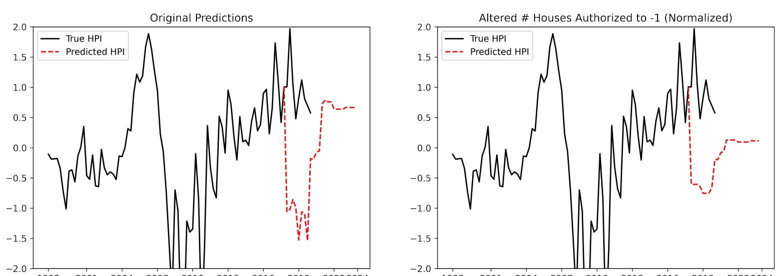


Figure 5: ARIMA Forest Predictions of `hpi` with and without decreasing `housing authorized`

As you can see, there seems to be some impact from the original to altered predictions, but it's more likely that this is a result of forcing `housing authorized` to be a constant vector.

Kalman Filter

In our foray into possible techniques, we considered using a Hidden Markov Model to analyze this data. However, we realized that Hidden Markov Models are used for a discrete state space, which doesn't have very much application to this issue. We turned our attention instead to methods used for continuous state spaces, including Kalman and Particle Filters. In order to use these methods, however, it would be necessary to specify an applicable model to avoid model misspecification (see Volume 3, Chapter 13.3.2). To estimate a model, we used Ordinary Least Squares to estimate how the features predict the next quarter's `hpi`. We engineered some of the features by adding squares, cubes, quartics, sines, cosines, tangents, and pairwise multiplication of each of the features to access a wider range of functions. We then pared down the input features to eliminate the seemingly insignificant features included. As a result, we found that the most important features were `per_capita_personal_income` × `current_hpi`, `current_hpi` × `household_consumption_index`, and `current_hpi` × `population`

In our assessments of other models, such as Random Forests, we found the feature importance of SNAP benefits on our model to be incredibly high. Hence, to avoid co-linearity, we specifically eliminated the SNAP benefits in this OLS/Kalman Filter analysis.

From the OLS analysis, it appears that `per_capita_personal_income` × `current_hpi` is the highest indicator of the next quarter's `hpi`. This is interesting, yet it makes sense, since this indicates that the `hpi` increase is proportional to the income of the residents in that area.

With this model, we attempted to implement a Kalman Filter. Figure 6 illustrates our results:

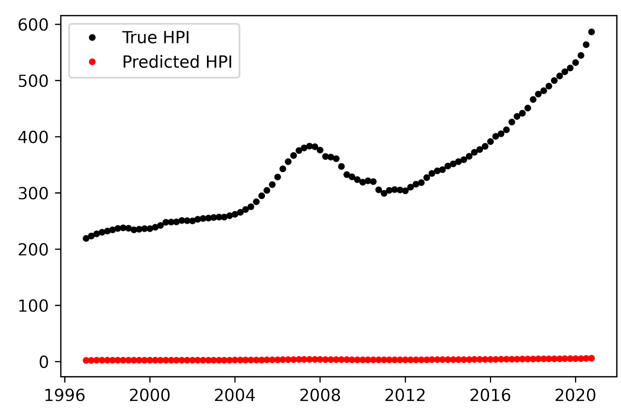


Figure 6: Kalman Filter Predictions Compared to Observed Data

In the end, the Kalman Filter didn't predict the `hpi` correctly. Because there is no physical law we can appeal to that determines HPI, and because we weren't able to find a clear governing equation, we were unable to specify a model for the Kalman Filter to use in its prediction.

Although the Kalman Filter didn't work with the model we constructed, we still were able to glean valuable insights (as discussed above).

Ethical Considerations

In our analysis, we identified some potential correlations between the housing price index and various features we tracked (such as SNAP benefit distributions, approved building permits, etc). In observing these correlations, care must be taken to avoid believing in a causal relationship between those two features. At first glimpse, our analyses might imply that the government should reduce SNAP benefits in order to bring down the housing price index; however, there is no guarantee (and we believe it is highly unlikely) that this would achieve the desired effect. Instead, this would likely only increase the poverty in the area. As such, further experimentation is required before any conclusions of causality are drawn.

Another possible ethical consideration is how these results could affect people's decision making. Because we are dealing with data and results that have financial implications, we have to consider how the results might cause panicked buying or selling. For example, if the model predicted sharp decrease in housing price index, there could be a panic among people worried about losing their investment on their house. While this scenario seems far-fetched (housing is not an asset that a person is likely to sell on a whim) we advise all those looking to use our model to make predictions not to consider the value of the predictions (which are demonstrably inaccurate), but rather consider the trends when making decisions.

Conclusion

In conclusion, some of our models indicated that increasing the housing authorized had a negligible impact on the housing price index. This may reflect reality, however, it would be dishonest to say our models were sufficient to predict the nature of the housing market (in spite our best efforts to the contrary). While we may attribute this failure to many factors, such as lack of sufficient or relevant data, improper implementation of algorithms, or the difficult balance between causal and correlational relationships, it is more likely that the data is simply too chaotic to predict. For example, even in our very limited years of data between 1997 and 2020, we still had a housing market crash in 2008. Perhaps in future research, one could account for shock factors such as a market crash by using more data or methods outside the scope of this class.

In the end, we found that modeling the future of housing prices is difficult, but it seems the only thing in life more certain than death and taxes is that housing prices will continue to increase unless something is done.

References

[1] U.S. Bureau of Economic Analysis, Personal Consumption Expenditures: Total for Utah [UTPCE], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/UTPCE, March 19, 2022.

[2] U.S. Census Bureau, Income Inequality in Salt Lake County, UT [2020RATIO049035], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/2020RATIO049035, March 19, 2022.

[3] Realtor.com, Housing Inventory: Active Listing Count Year-Over-Year in Utah County, UT [ACTLISCOUYY49049], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/ACTLISCOUYY49049, March 19, 2022.

[4] Realtor.com, Housing Inventory: Median Listing Price per Square Feet Month-Over-Month in Utah County, UT [MEDLISPRIPERSQUFEEMM49049], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/MEDLISPRIPERSQUFEEMM49049, March 19, 2022.

[5] U.S. Census Bureau, SNAP Benefits Recipients in Utah [BRUT49M647NCEN], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/BRUT49M647NCEN, March 19, 2022.

[6] Realtor.com, Housing Inventory: Median Days on Market in Salt Lake City, UT (CBSA) [MEDDAYONMAR41620], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/MEDDAYONMAR41620, March 19, 2022.

[7] Realtor.com, Housing Inventory: Active Listing Count in Salt Lake County, UT [ACTLISCOU49035], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/ACTLISCOU49035, March 19, 2022.

[8] U.S. Census Bureau, New Private Housing Units Authorized by Building Permits for Utah [UTBPPRIVSA], re-

trieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/UTBPPRIVSA, March 19, 2022.

[9] U.S. Census Bureau, Estimate of Median Household Income for Salt Lake County, UT [MHIUT49035A052NCEN], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/MHIUT49035A052NCEN, March 19, 2022.

[10] U.S. Bureau of Economic Analysis, Gross Domestic Product: All Industry Total in Utah [UTNQGSP], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/UTNQGSP, March 19, 2022.

[11] U.S. Census Bureau, Resident Population in Utah [UTPOP], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/UTPOP, March 19, 2022.

[12] U.S. Federal Housing Finance Agency, All-Transactions House Price Index for Utah [UTSTHPI], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/UTSTHPI, March 19, 2022.

[13] U.S. Bureau of Economic Analysis and Federal Reserve Bank of St. Louis, Per Capita Personal Income in Utah [UTPCPI], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/UTPCPI, March 19, 2022.

[14] U.S. Bureau of Labor Statistics, Unemployment Rate in Utah [UTUR], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/UTUR, March 19, 2022.

[15] U.S. Census Bureau, Rental Vacancy Rate for Utah [UTRVAC], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/UTRVAC, March 19, 2022.

[16] U.S. Census Bureau, Real Median Household Income in Utah [MEHOINUSUTA672N], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/MEHOINUSUTA672N, March 19, 2022.

[17] U.S. Census Bureau, Bachelor's Degree or Higher for Utah [GCT1502UT], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/GCT1502UT, March 19, 2022.

[18] U.S. Census Bureau, Net County-to-County Migration Flow (5-year estimate) for Utah County, UT [NETMIGNACS049049], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/NETMIGNACS049049, March 19, 2022.

[19] U.S. Census Bureau, Single-Parent Households with Children as a Percentage of Households with Children (5-year estimate) in Salt Lake County, UT [S1101SPHOUSE049035], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/S1101SPHOUSE049035, March 19, 2022.