Project Final Report

Ryan Liu, Onching Lin

ABSTRACT

Home ownership is a fundamental part of financial stability and usually a very important part of an adult's life, and a purchase that many people save up for. Traditionally, we have believed that home ownership rate is dependent on a few things, the sales price of the houses, the interest rates at which people can borrow for this purchase, and the amount of wealth a median person has to purchase the house. For this assignment, we are provided with the home ownership rate as well as interest rates, median sales price of houses, and the real GDP of the US.

Recently, however, a new factor has come into play. The Covid-19 pandemic has affected all aspect of US life, and as people adapt to the new normal, the housing market has changed as well. Covid has caused an exodus from the cities, and allowed people to work from home more, causing changes in house purchasing behavior, so we want to add this factor into the predictions to see if it can account for the spikes we've been seeing in the housing market.

For this project, we will use an ARIMA model to understand how home ownership has changed over time. We've decided that this method from investigating the ACF and PACF graphs before and after fitting the model.

1 DATA

For obtaining the data, we were able to get the Covid infection we were able to get weekly data of infections, deaths and vaccinations since the start of the pandemic in the United States from the CDC. We aggregate this information into quarterly data for our analysis. We also got monthly interest rate from the assignment and we will convert it to quarterly. For the other two data sets, the home ownership rate and the real GDP, we were also provided data at a quarterly level.

To clean and transform the data, we need to change the frequency of the data to be quarterly for all our information, and there are some missing or outliers in the covid data that we take care of using OpenRefine. We used Python pandas for cleaning, scraping, and filtering data.

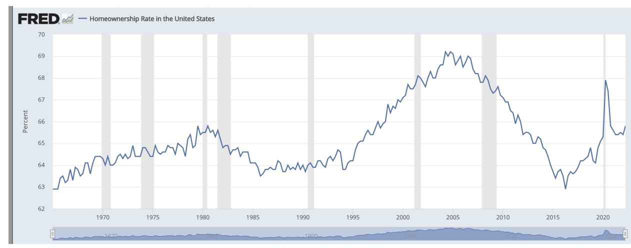


Figure 1: Quarterly Homeownership Rate

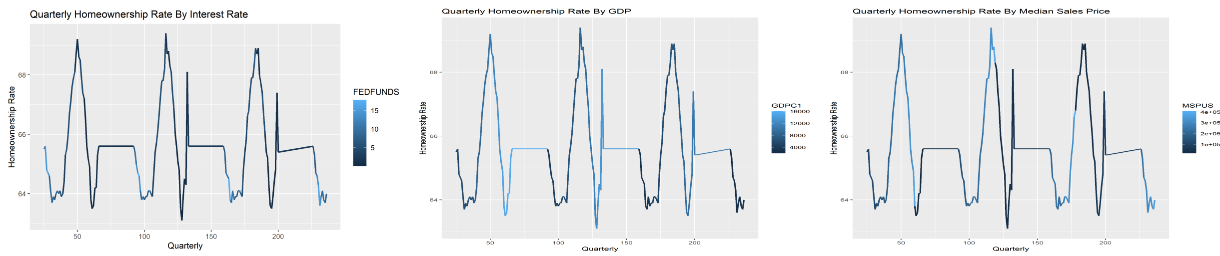
We can see from the 5 of the data that there seems to be an upward trend and a lot of seasonality in the data. It also look like the variance could have increased recently, and perhaps this is something we should model as well with a GARCH Model, but we will first try the ARIMA model and see what the residuals look like.

2 CORRELATION WITH FACTORS

We visualize the correlation between the factors with color as seen in Figure 2. Based on the plots, there is trend and seasonality for each factor. As the interest rate increases, the home-ownership rates starts to significantly decrease. There's also a relationship between GDP and home ownership rate. People start buying houses when GDP reaches its peak and then start slowing down. Oddly, the median sales price of the houses has a negative correlation with the home ownership.

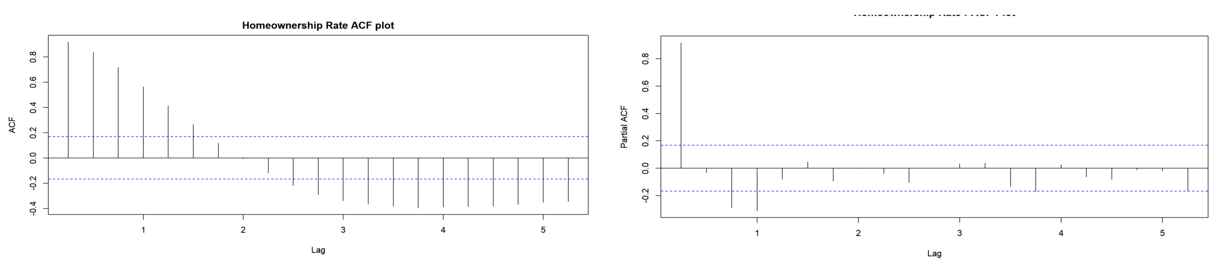
3 MODEL SELECTION

One of the questions we are asked to answer is whether or not the recent trend in home ownership is an anomaly or is it expected. This is why we're adding in the pandemic information to see if the recent increase correlates with the pandemic information; and if so, we can probably say that the recent trends will not continue. It will also let us know if there are other factors that affect home ownership besides the traditional ideas, i.e. the many things that were affected by the pandemic like the increase of remote work and the change in income distribution.



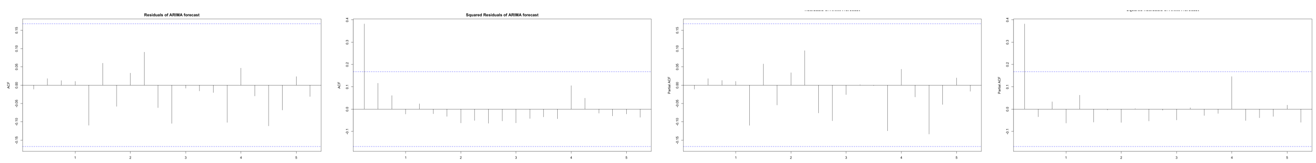
(a) (b) (c)

Figure 2: (a) Quarterly Homeownership Rate By Interest Rate (b) Quarterly Homeownership Rate By GDP (c) Quarterly Homeownership Rate By Median Sales Price Sold



(a) (b)

Figure 3: (a) Home Ownership ACF (b) HomeOwnership PACF



(a) (b) (c) (d)

Figure 4: (a) Residuals ACF (b) Residuals Squared ACF (c) Residuals PACF (d) Residuals Squared PACF

From extensive exploratory data analysis, we decided to proceed the following approaches in predicting homeownership rate using Time Series Analysis:

1. ARIMA
2. ARIMA-GARCH
3. VAR

ARIMA is the most basic but powerful auto-regressive time-series modeling approach that uses past values.

GARCH is a general variant of ARCH, which is used for modeling volatility of shocks to the series. GARCH allows for lagged conditional variance, and is generally used to model volatility and uncertainty over time, and it tries to better map the squared residuals. We thought that this would be a good model to try with the homeownership rates as there can be volatility in the housing market as we saw in the financial crisis, but in general we believe that homeownership and housing prices are more stable than GDP and other financial indicators. We use ARIMA GARCH to acocunt for both the lag residuals and the lagged squared residuals.

Finally, we tried a VAR (Vector Autoregressive Model) which looks at all the different datasets we have. VAR is a multivariate time series model that relates current observations of a variable with past values of itself and past observations of other variables in the system. VAR models differ from uni-variate auto-regressive models because they allow feedback to occur between the variables in the model.

4 HOME-OWNERSHIP RATE ACF AND PACF PLOTS

We can see from the Figure 3 that there looks to be a trend in the data from the ACF plot. Then, with the ACF plot, it looks like we should use a AR model since ther is some correlation in the lag as some values are beyond the prediciton band.

5 RESIDUALS FROM ARIMA MODEL

We fit an ARIMA plot to the data and then checked the residuals and squared residuals of the data. It looks like all the values are within the prediction band so it suggests that the ARIMA model already takes care of a lot of the autocorrelation. However, the fit can still be improved, and thats why we used a ARIMA GARCH model

6 MODEL RESULTS

Model	MAPE	PM
ARIMA	0.018	17.47
ARIMA-GARCH	0.0068	2.527
VAR	2.223	27.62
VAR Restricted	1.706	15.92

We saw that the ARIMA GARCH model predicted the best in terms of MAPE and PM. It wasn't surprising that ARIMA GARCH would do better than ARIMA, but it was surprising that ARIMA GARCH predicted better than VAR.

When we look at the VAR results, we saw that the only significant variables for predicting homeownership rates were lagged homeownership rates. This was surprising as we thought that different interest rates and GDP measures would be good contextual information in prediction, but the VAR model tells us this isn't the case.

We looked at the Granger Causality, and the none of the other variables we looked at were helpful in predicting homeownership rates, but rather homeownership rates were helpful in predicting GDP, FEDFUNDRATES etc. While this doesn't show actual causation, its an interesting theory as to why VAR perhaps doesn't perform as well in prediction since there is more noise its overfitting on potentially.

7 EXTERNAL FACTORS

We wanted to also look at how different external factors could have effected home-ownership rates. To do this, we used other external factors as regressors, which includes work from home prevalence, COVID infections, housing market regulation, interest rate, real GDP and the median sales price of houses sold. Using these regressors, we forecast changes in the real estate market under different scenarios. As a supporting documents, we used data from Redfin/Zillow for housing prices to later compare with the home-ownership rate.

There are also good visualizations of how people have migrated and how housing prices have changed, but there are no visualizations that show the combination of the two factors. While Zillow can show the historical price changes, our approach will depict how "Work from home" culture and Covid have impacted the home-ownership rate and housing prices.

The forecast data show them how their return of investment will look like. With the data visualization through maps or graphs, it will be easier to identify trends, patterns, and outliers within large data sets. Also, since the pandemic is ongoing and the housing market is still adapting to it, with more companies requiring people to come back to the office, any new analysis will be able to look into more interesting trends.

8 RESULTS FROM EXTERNAL FACTOR ANALYSIS

According to research, 37 percent of jobs, a share concentrated among high paying jobs can perform entirely remotely. This could also be the reason why they would buy houses in less competitive areas. Based on our studies, we found that states with highest increase in house pricing are the southern states such as Georgia, Florida, North Carolina, and Tennessee. In comparison, the larger states like New York and California showed less increase in house pricing. The impact on house pricing is paralleled with home-ownership rate. This concludes that working from home has impacted the real estate market and Covid pandemic has indirectly impacted the home-ownership rate.

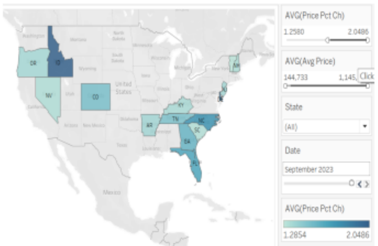


Figure 5: States with High Increase In House Pricing

Our Output from other ARIMA model using house pricing as dependent variable predicted that there was higher prices increase in the southern states. This also represent the increase of home-ownership rate. Therefore, we can conclude a trend of people migrating to the southern states and states around the larger states to avoid competitions.

During the Covid-19 pandemic, the monthly supply of homes for sale has decreased to historically low levels. Home-ownership rate growth has increased substantially during the pandemic as well as post-pandemic. We have a few hypothesis of factors that impacted the home-ownership rate; The significant drop in interest rates in the last year and a half can be attributed to actions taken by the Federal Reserve in responses to Covid-19. The inequality income among citizens, home-owners became at risk of defaulting on their mortgage payments. In an effort to avoid widespread foreclosures, the government implemented polices intended to decrease interest rates to make housing more affordable.

9 CONCLUSION

From our ARIMA, ARIMA-GARCH and VAR models, we found that the ARIMA-GARCH model did the best in predictions based on the MAPE and PM values. It performed signficantly better than the alternatives, and showed that it is important to pick good regressors in VAR.

We also looked at other external factors that could have given us more predictive power, specifically accounting for COVID and its effect in recent years. We saw some correlation with covid and homeownership, as well as with the home prices. It shows us that demand really did move over the last few years and that effect wasn't captured by the other macro economic factors we were looking at.