Analyzing Homeownership in the United States

Group 3: Ashley Stroup, Marissa Adams, Timothy Powers, and Steven Cruz

Summary:

Homeownership rates have been viewed as how economically wealthy a country is; therefore, many studies have been performed to understand what factors affect the homeownership rate in the United States. Interest rate, median home prices, and the GDP are considered to be contributors to the homeownership rate, as well as the unemployment rate and new housing unit permits. By analyzing the homeownership rate from 1980 through 2020 on a quarterly basis, predictions are produced for 2021-22 and compared back to actual data for an ARIMA, ARMA-GARCH, ARIMAX, and VAR model. The VAR model, which uses all the given factors, performs best over the other models as it is a multivariate time series. This model can be used to predict short-term rates relatively well if all factors have been accounted for.

Introduction:

Factors such as the interest rate (Fed funds rate), median home prices, and the gross domestic product (GDP) are thought to affect the homeownership rate, given that both the interest rate and home price are the fundamental components of a mortgage. GDP is the metric that depicts the overall health of the economy and thereby the homebuyer. Furthermore, a number of other economic and housing-specific indicators, such as the unemployment rate and new home starts/permits may singularly or when paired with other factors signal changes in the direction of the homeownership rate. Additional environmental factors include economic policy aimed at broadening homeownership among diversified or marginalized populations and communities, the most notable of which have come about in the early 2000s, and which ultimately perpetuated the 2008 financial crisis by lowering industry-wide requirements for mortgages.

Peculiar to the past year in the homeownership rate, the combined effects of meager interest rates on mortgages as well as the broad professional transition to remote working incentivized homebuyers to purchase earlier than they may have otherwise planned and in metros or suburbs geographically distinct from their workplaces. This group intends to examine if recent rates of homeownership fall within the range of reasonable expectation, whether short or long-term rates can be more practically predicted, which time series characteristics contribute most to homeownership predictability, and whether there are further external factors to those already mentioned that may also assist in predicting homeownership. The group anticipates that all three factors (and potentially others) contribute to the homeownership rate and that the ARIMAX methodology may be the best performer for short-term predictions. However, long-term predictability may be more difficult given that political policy, which can be arbitrary, plays a role in changes in the series.

Analysis:

Time Series Characteristics and Predictability

This section aims to understand if there are specific characteristics of the homeownership rate time series that contribute most to its predictability. For the data sets provided, Figure 1 shows the time series plots. It is interesting to note that as the interest rate peaked in the early 1980s, the homeownership rate begins its decline. With the interest rate at significant lows, the homeownership rate begins to climb. However, there seem to be no significant dips in the median home sale price or GDP. Both factors have an increasing trend over time. Given this potential relationship between interest and homeownership rates, a trend analysis was performed on both data sets as seen in Figure 2. This shows that the spline model is the closest approximation; however, there is still a substantial amount of variability that is not captured by the current trend model. Additionally, variance appears to be non-constant as well. This indicates that the data is not stationary.

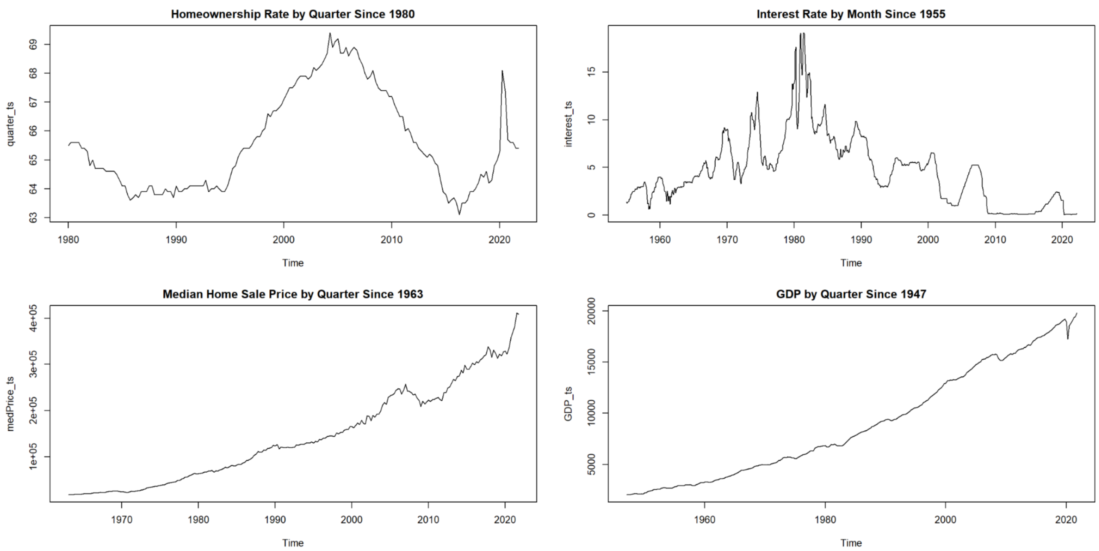


Figure 1. Time Series Plots of Provided Data Sets.

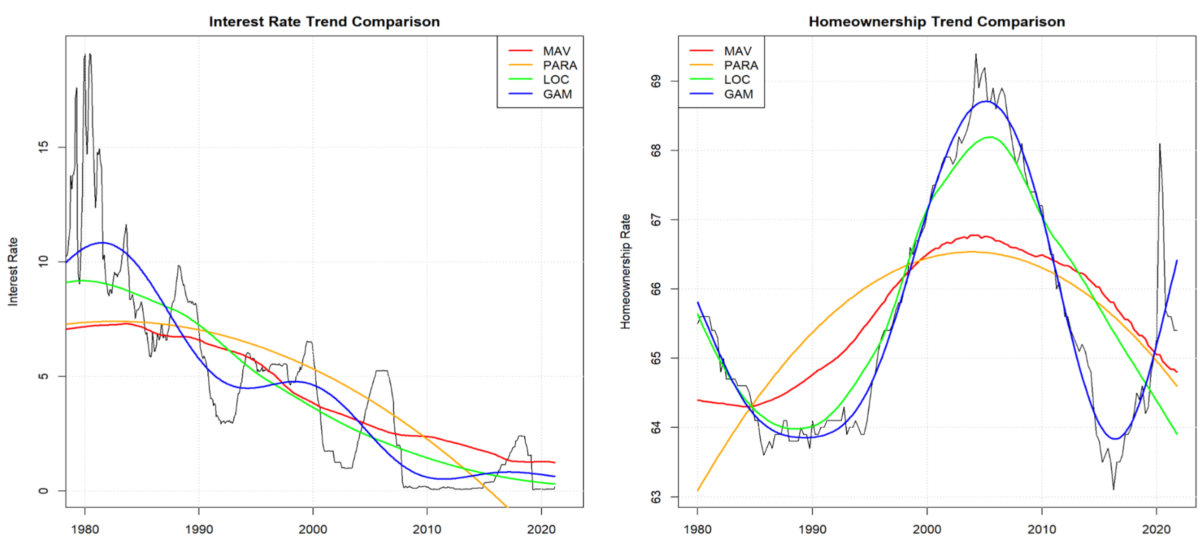


Figure 2. Interest Rate and Homeownership Trend Estimation Comparison

Furthermore, when examining the ACF plots (Figure 3) of the homeownership rate, GDP, median house sales, and interest rate, all factors appear to have slowly declining lags. These slowly declining lags indicate non-stationarity, which may be remedied by differencing. Additionally, the ACF plot of the homeownership rate appears cyclical, intermittently spiking, positively and negatively. There do not appear to be clear signs of seasonality in the ACF plots. Further, examining the PACF plots of the four series, none of the sets have slowly declining lags. Instead, the spikes cut off quickly, seemingly at 1 or 2 lags, depending on the series as shown in

Figure 4. Overall, from the analysis of the collection of trend plots and ACF plots, it appears that trend and variance look to contribute the most to the predictability of the time series.

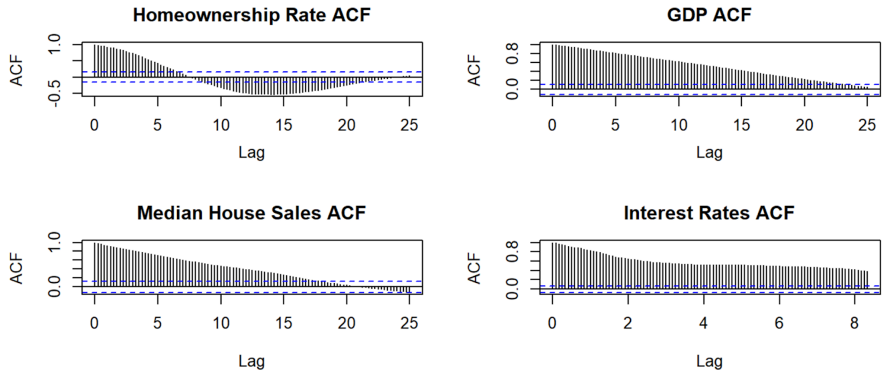


Figure 3. ACF All Datasets

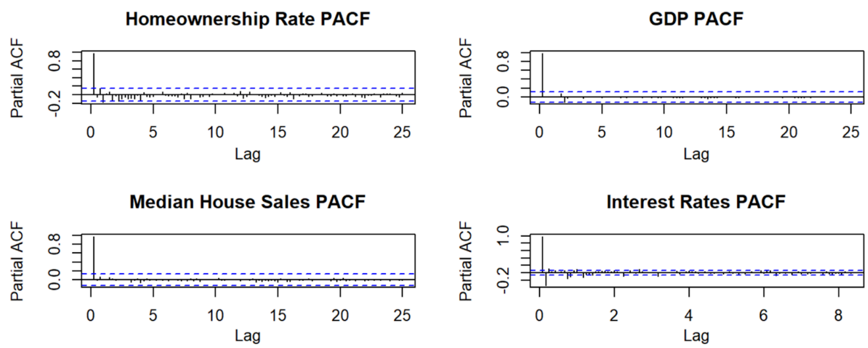


Figure 4. PACF All Datasets

External/Exogenous Factors in Predicting Homeownership

The following modeling sections explore whether or not external or exogenous factors help in predicting homeownership and if so, what these factors are. The timeframe for training the following models is 1980-2020 on a quarterly basis, with the testing timeframe being the four quarters of 2021. The three most recent quarters of 2022 are included as visual reference but the analysis exercise includes forecasts for the full year of 2022.

ARIMA Modeling:

From the exploratory analysis, the data has been shown to be non-stationary. ARIMA is a good modeling approach for non-stationary processes because the trend can be removed by differencing. This methodology also provides a more precise fit than a splines regression alone, which only really gives an indication of the trend. An ARIMA analysis was performed on the homeownership rate with the first step of finding the orders for the best-performing model on the basis of their AICc. As shown in Figure 5, the best model that has the lowest AICc score of 74.04 is an ARIMA(1, 1, 6) process. The ARIMA(1, 1, 6) ACF residual plot shows all spikes under the significance lines. The histogram and QQ plots also show the residuals having an approximately normal distribution.

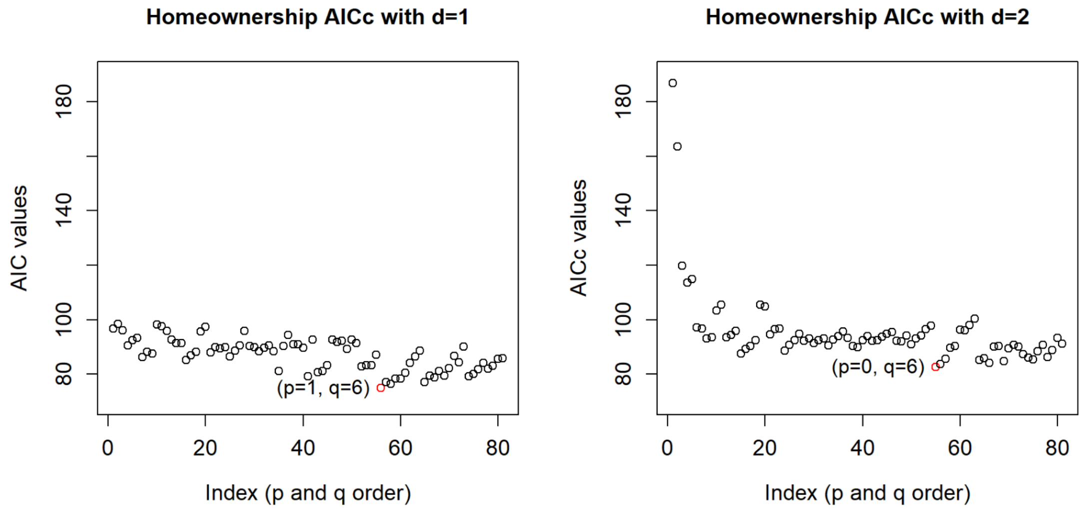


Figure 5. Homeownership AICc Values

ARIMA Forecasting:

Using the model chosen from above, forecasted values are obtained for all of 2021 as well as future values for 2022. Observed 2022 values are included along with the forecast as reference points. As shown in Figure 6, the forecasted data for 2021 is higher than the actual values. The 95% prediction interval also encapsulates none of the observed values except one.

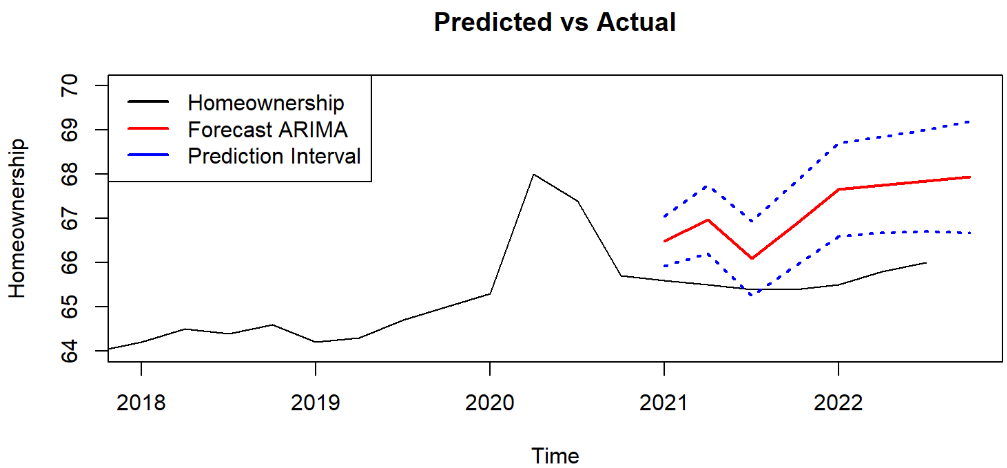


Figure 6. Forecasting ARIMA

ARMA-GARCH Modeling:

Along with trend, variance appears to be an important characteristic representing the homeownership rate over time contributing most to the predictability of the time series. As seen in the ARIMA modeling section, the ARIMA model does not factor in the variability or shocks that occur within the system. However, a GARCH model does. It is common to model both the mean and variance data of the time series. Therefore, an ARMA-GARCH model is appropriate to analyze and forecast the data. It also may perform better than an ARIMA approach alone. The analysis initially fits orders of the ARIMA component of the analysis with p=1 and q=6, which then allows for the GARCH order to be tuned. Once the initial GARCH order is selected, the ARMA orders are refined. This refinement selects p and q equal to 7 and 6, respectively. The GARCH order is finalized with new ARMA orders; however, the GARCH orders remain consistent. The final model selected is an ARMA-GARCH(7,6)x(1,1). Also, this model reflects undifferenced data. The residual analysis shows that the residuals are uncorrelated. This is seen in Figure 7 with the ACF and PACF of the residuals and squared residuals. The plots show that there are no significant lags and thus appear to resemble white noise.

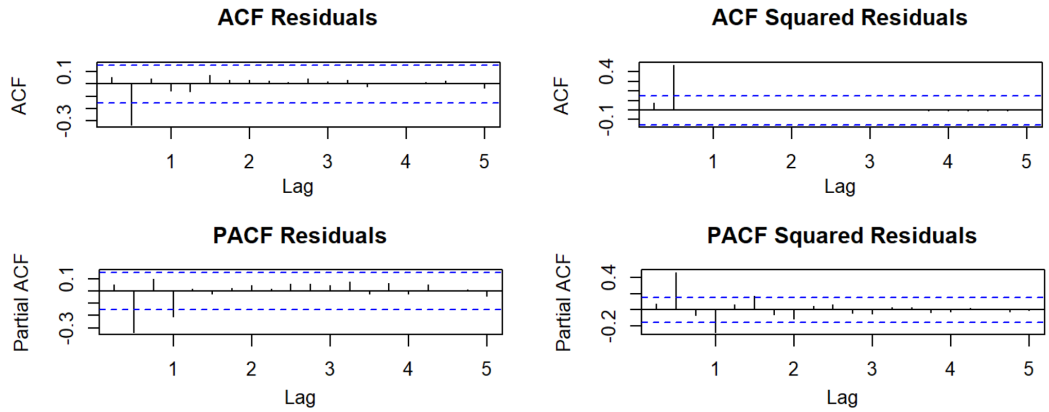


Figure 7. ARMA-GARCH Residual Analysis

ARMA-GARCH Forecasting:

Using the model selected in the previous section, the next eight quarters of data are forecasted. The ARMA-GARCH definitely predicts more accurately than the ARIMA model as shown in Figure 8. It follows the observed data much more closely, with most actual time points occurring within the prediction interval. The final forecasted time point is a substantial deviation from the prior progression; however, this final time point is unrealized. While it may be tempting to say such a forecast is unlikely, it is similar to the change between Q1 and Q2 of 2020.

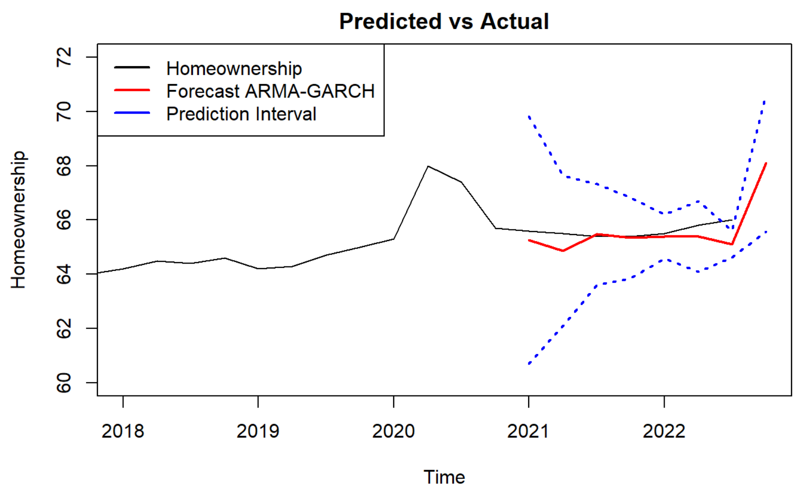


Figure 8. ARMA-GARCH Forecasting

ARIMAX Modeling:

This section continues the exploration of the previous sections as well as looks into external factors to aid in predictions. As with the ARIMA and ARMA-GARCH above, each only takes into account one univariate time series. However, factors such as the interest rate, median home prices, and the GDP are thought to be contributors. Additional economic series were analyzed that might influence the homeownership rate, including the unemployment rate, the inflation rate, and new unit housing starts and permits. In order to address these additional variables, an ARIMAX process is employed for analysis.

Initially, the original data sets are compared against their first-order differences to determine if this transformation is enough for stationarity among the time series. Differencing the various series to different degrees could ruin critical information and could cause difficulty in returning forecasted values (Figure 9a). Conversely, while the provided data sets are largely characterized by trend processes, this would allow for simple differencing to make them stationary. The additional data are highly seasonal, requiring detrending and deseasonalizing. In this case, applying a seasonal means model on a quarterly basis and differencing the series makes them stationary (Figure 9b).

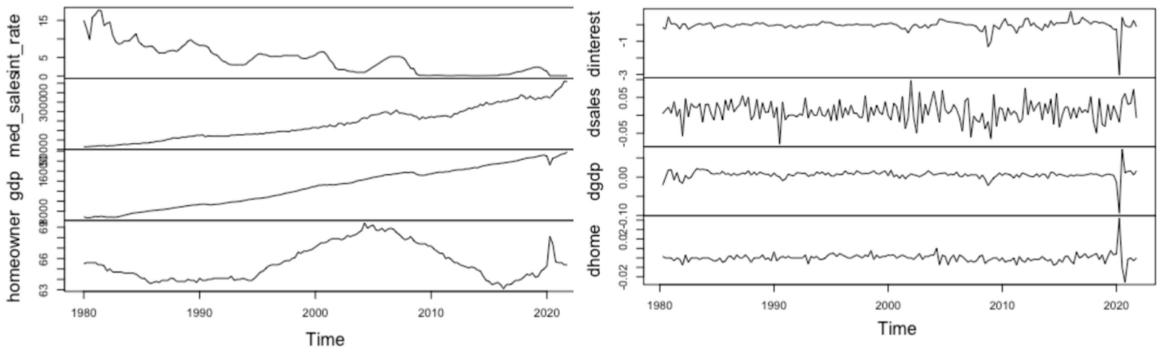


Figure 9(a). Provided Data Sets: Not Differenced vs Differenced

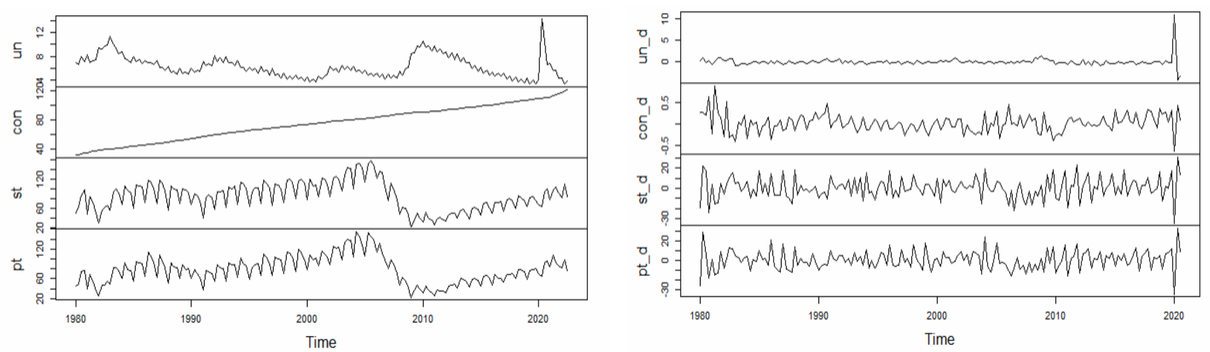


Figure 9(b). Additional Data Sets: Not Differenced vs Differenced/Deseasonalized

When examining the extraneous data sets, there appears to be a negative lagging relationship between the homeownership rate and the unemployment rate. Also, there are positive lagging relationships with new housing starts and permits. As explained above, the series are made stationary after detrending and deseasonalizing. Therefore, analysis suggests there is some value in assessing the relationship between the unemployment rate and new housing permits with the homeownership rate, alone and in conjunction with the provided data sets.

Iteratively looping over models only considering the provided time series (interest rate, GDP, and median sales price), the best ARIMAX model is found to have a fit of ARIMAX(3,1,3) with an AICc of 34.13. When considering additional data sets, specifically the unemployment rate and new housing permits, a seasonal ARIMAX model is employed to iterate over possible orders, finding those with the lowest AICc also usually having the highest orders. In this analysis, the two best models on the basis of low AICc include all provided time series with the unemployment rate and with both the unemployment rate and new housing permits, having an AICc of -38.50 and -39.58, respectively. However, the orders of both of these fits are particularly high, having S.ARIMAX(7,1,11)x(1,1,1) and (8,1,11)x(0,1,1), respectively. Furthermore, the performance metrics of both models are not comparable with the core of the analysis to this point. Instead, by removing the seasonal component, which does not improve the fit, and selecting more parsimonious models in line with earlier analysis, either fit performs much better having an AICc of -1.70 and -6.05, and ARIMAX(4,1,3) and (3,1,4), respectively.

Figure 10a depicts the residual analysis of the ARIMAx(3,1,3) for the provided data set, which shows no pattern. The ACF residuals show all spikes under the significance lines. The histogram and QQ plots show the residuals having an approximately normal distribution. When performing Box-Pierce and Ljung-Box tests, both p-values are greater than a 0.05 significance level, suggesting the null hypothesis of uncorrelated residuals not be rejected, indicating it is plausible the residuals are uncorrelated. Similarly, Figure 10b (left) depicts the residual analysis for the extended model using both additional variables: the unemployment rate and new housing permits. The primary difference between the two models can be seen in the modest left-skew of the residuals of the extended model. The p-values of the Box-Pierce and Ljung-Box tests are both close to zero, which suggests rejecting the null hypothesis of uncorrelated data. Economically speaking, new housing permits and the homeownership rate are likely to be correlated since one can make a case anecdotally for either reciprocating an increase or a decrease in the other. Removing the new housing permits variable, the given data and the unemployment rate create an ARIMAX(4,1,3) model, whose residual analysis is seen in Figure 10b on the right. The p-values for the Box-Pierce and Ljung-box tests are larger: just under 0.05. While not definitively above a significance level of 0.05, they are substantially larger than the extended model including new housing permits.

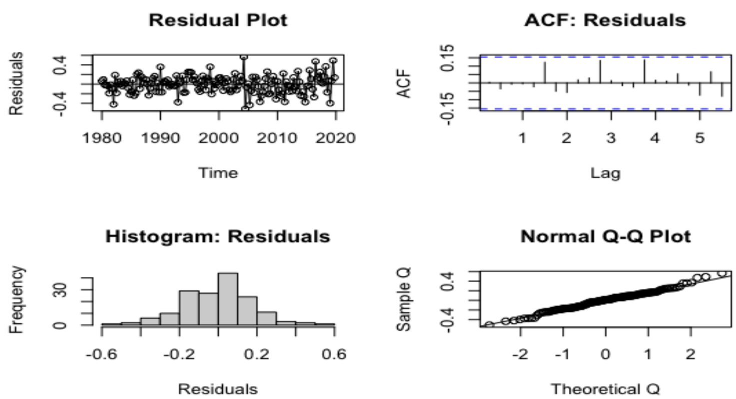


Figure 10(a). ARIMAX Residual Analysis (interest rate, med. home sales, GDP)

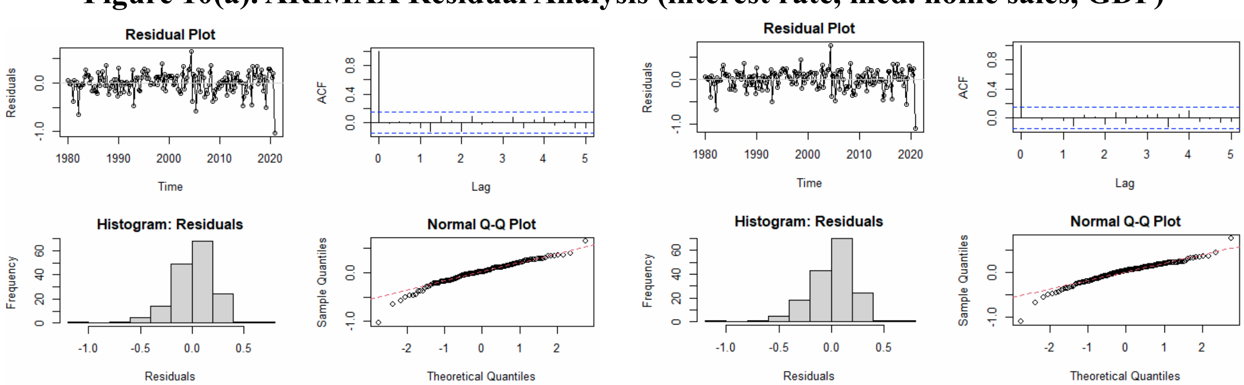


Figure 10(b). ARIMAX Residual Analysis (given data + unemployment rate, new housing permits)(left), ARIMAX Residual Analysis (given data + unemployment rate)(right)

ARIMAX Forecasting:

Given the ARIMAX(3,1,3) model selected from among the provided data sets, Figure 11a demonstrates its fit to predict the eight quarters of 2021-22. It is clear from both its performance metrics and its prediction that the exogenous series: the interest rate, median home sales, and GDP all contribute to creating a model that fits the data and performs much better than the simple, univariate ARIMA(1,2,9) model and marginally better than the ARMA-GARCH(7,6)x(1,1) model. Figure 11b (left) depicts the 2021-22 forecast of the ARIMAX(3,1,4) model, which uses the provided series, as well as the unemployment rate and new housing permits, as exogenous variables, and Figure 11b (right) depicts the 2021-22 forecast of the ARIMAX(4,1,3) using the provided data series in addition to just the unemployment rate.

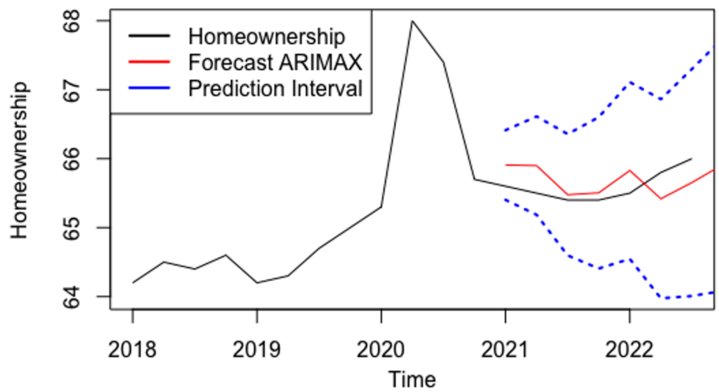
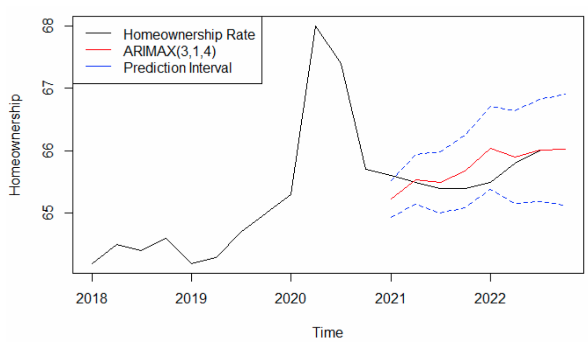


Figure 11(a). ARIMAX(3,1,3) Forecast



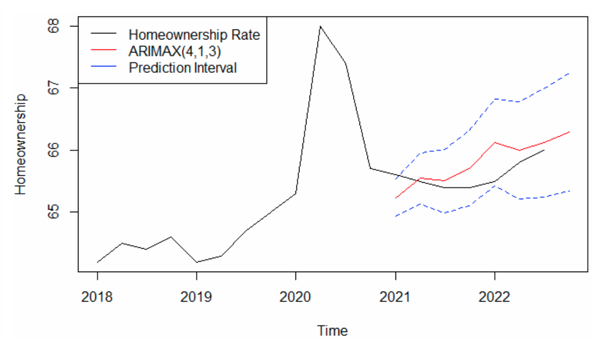


Figure 11(b). ARIMAX(3,1,4) & ARIMAX(4,1,3) Forecasts

VAR Modeling:

Further considering multivariate analysis for describing the homeownership rate, a VAR analysis demonstrates substantial forecasting power. Running the analysis on each combination of the provided data sets (interest rate, median home sales, and GDP), the combination of all as endogenous variables produces the most robust model thus far, with the unrestricted VAR model having marginally better performance over the restricted variant. As seen in Table 2, a model of order 2, having been selected according to the BIC as a model comparison criterion, reduces both the MAPE and PM as compared to the ARIMAX models. Among the contributing series, only the interest rate series fails the assumptions of the Wald test. Although counterintuitive, the interest rate does so for each VAR model it is built into. In the restricted model, the interest rate is the least incorporated variable among the various regressions. When including either the unemployment rate or the new housing permits series, the best endogenous model including the unemployment rate excludes the GDP series, while the best model including the new housing permits series includes all of the other variables, in addition to the unemployment rate. As seen in Table 2, the performance metrics for these models are comparable with the models produced by the earlier ARIMA and GARCH models. Unlike during the ARIMAX analysis, these two series do not demonstrate gains in performance.

VAR Forecasting:

Visually inspecting both VAR models in Figure 12, it is clear how well either performs over the 2021 test timeframe as 3 of the 4 time points assessed overlay almost identically, while the last time point deviates as the housing market begins to shift substantially near the start of 2022. Nonetheless, pulling the forecast forward, observed values remain within the prediction interval without it broadening by much. Although the model fits sync well during the test period, it is also clear that the model has difficulty anticipating seemingly unexpected shifts in the homeownership rate.

When comparing ARIMA, ARMA-GARCH, ARIMAX, and VAR via precision measures, the unrestricted VAR model with the interest rate, median home sales, and GDP has the best precision metrics (MAPE and PM) followed by ARIMAX with the interest rate, median home sales, GDP, unemployment rate, and new permits (Table 1, Table 2). The lesser performing models were ARIMA and ARMA-GARCH, which are univariate models and only model the homeownership rate itself. This indicates that additional variables do help in predicting the homeownership rate.

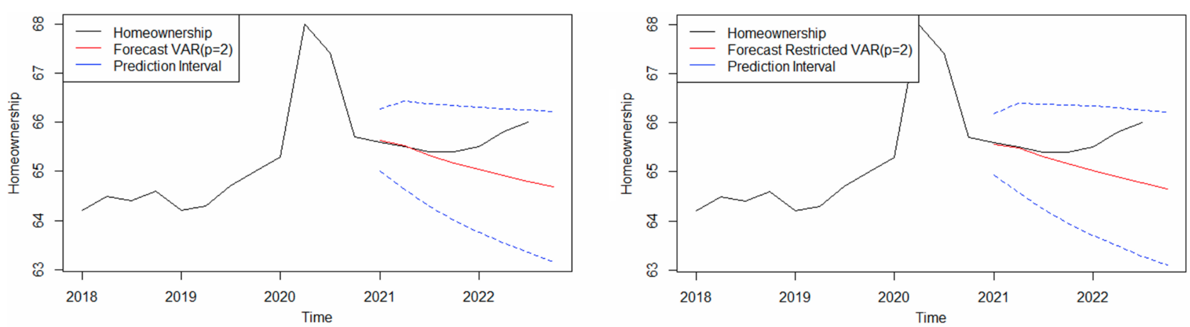


Figure 12. VAR(p=2) Unrestricted & Restricted Forecasts: All Provided Data Sets

Table 1. Precision Comparison for Test Data[1]

Precision Metric	ARIMA(1,2,9)	ARMA-GARCH (7,6)x(1,1)	ARIMAX (3,1,3) (Given Data)	ARIMAX (4,1,3) (Add'l Unemployment Rate)	ARIMAX (3,1,4) (Add'l Unemployment Rate & New Permits)
MAPE	0.0173	0.0043	0.0034	0.0032	0.0030
PM	204.148	19.201	9.845	8.848	8.217

1, Precision is measured over the 4 quarters of 2021, which represents the testing timeframe; 1980-2020 represents the training timeframe.

Table 2. VAR Model Precision Comparison for Test Data[1,2]

Precision Metric	VAR (p=2)	VAR(p=2) Restricted	VAR(p=2) (Interest Rate, Med. Home Sales & Unemployment Rate only)	VAR(p=2) Restricted (Interest Rate, Med. Home Sales & Unemployment Rate)	VAR(p=4) (Add'l Unemployment Rate & New Permits)
MAPE	0.0014	0.0015	0.0046	0.0025	0.0054
PM	2.109	2.434	14.076	3.938	25.775

1, Precision is measured over the 4 quarters of 2021, which represents the testing timeframe; 1980-2020 represents the training timeframe.
2, Multivariate models assumed to include Interest Rate, Median Home Sales, and GDP as exogenous inputs unless otherwise stated.

Time Series Predictions and Recent Trends in Homeownership

Based on the earlier analyses, the recent homeownership rates fall within the range of reasonable expectations. The ARIMAX and VAR models show that the 2021 rates fall within the 95% prediction intervals. This leads to the assumption that the rates can be more practically predicted on a short-term basis. For long-term predictability, this will be more difficult and less consistent. The ARMA-GARCH model can be helpful for predictions that experience variability due to policy changes. For instance, a lower Federal Funds rate has led to lower mortgage rates, which has consequently led to increased home sales and a higher homeownership rate, which is consistent with the above analysis.

Conversely, between March and November of 2022, the Federal Reserve aggressively increased the Federal Funds rate from roughly 0.25% to 3.75 - 4.00%, which has been the fastest and largest increase of rates in more than 10 years. The consequence of this monetary policy shift has been to increase mortgage rates, all within the context of a diminishing economic environment, which thus far has stalled both builder and homebuyer sentiment, and may very soon lead to diminishing homeownership over the coming years. The long-term predictability of the time series will likely be more reliable outside of significant macroeconomic policy changes. Also, it has been shown that there are further external factors that may assist in predicting changes in the homeownership rate. The ARIMAX and VAR models show, through the lowest MAPE as provided in Tables 1 and 2, that the unemployment rate and new permits may aid in predicting power.

Conclusion:

Being able to predict the homeownership rate can help with understanding a country's economy and stability. Characteristics such as trend and variance appeared to contribute most to the predictability of the time series. Both the ARIMAX and VAR models performed reasonably well with understanding short-term predictions. For future iterations, additional external factors could be pursued as well as exploring an ARIMAX-GARCH model. However, even with the additional external factors, policies can drastically change how potential future homeowners respond which affects the homeownership rate. Therefore, long-term predictions are not as efficient as short-term predictions.