Identifying the Relationship Between Macroeconomic Signals and the California Housing Market Surrounding Economic Downturns in the 21st Century

Malcolm Hsu[1]
QAC320
Wesleyan University

Abstract

Both the 2008 Financial Crisis and the COVID-19 Pandemic had extreme implications on both the U.S. economy as a whole, specifically the housing market. Using California housing market data from Zillow and macroeconomic data collected from various sources, the aim is to examine the connection between the country-level economy and the state-level housing market. Monthly time series data will be used to reestablish the connection between macroeconomic signals and fluctuations in housing prices, fit appropriate time series models, and forecast the housing market.

[1] Hsu: mhsu01@wesleyan.edu. Mathematics & Economics, Wesleyan University, 2025

1 Introduction

The housing market plays a critical role in the stability of the economy, extending to financial policy making and consumer behavior. Fluctuations in the housing market influence broader macroeconomic indicators, and thus understanding the connections between trends in the housing market and shocks in the economy is crucial during times of economic crises. Two of the largest crises, the 2008 Financial Crisis and the COVID-19 Pandemic, demonstrated the extent to which economic downturns affect housing markets. The 2008 crisis infamously caused a collapse in housing prices due financial instability and mortgage defaults, while the 2020 recession resulted in mixed outcomes due to both low mortgage rates and changes in consumer demand. From 2006 to 2011, housing prices dropped roughly 30 percent in the United States, accounting for eight million dollars lost in equity total. Additionally, at the peak of the crisis, 25 percent of homeowners owed more on their mortgages than the value of their property (Rohe et al., 2013). In California, the focus of this study, after the COVID-19 recession, one in ten home owners fell behind in their mortgage payments, with one in six renters falling behind in their rent (Little Hoover Commission, 2021). Understanding how the housing market evolves surrounding economic crises is critical in predicting how future shocks may impact the market.

Studies on the 2008 Financial Crisis examine the contributing factors to the crisis, including the negative shock of mortgages, over-optimism about returns resulting in a surge in asset prices, increased liquidity in financial markets, and excess investment. The combination of these factors resulted in an incorrect expectation of price expectations, opening a gap between mortgage standards and the demand for housing, resulting in the boom-bust phenomenon of housing prices and downturns in the financial market (Duca et al., 2010). When turning to the COVID-19 pandemic, housing market behavior was oppositely affected by a change in consumer behavior, housing supply shortages, and federal funds rate (Balemi et al., 2021). Furthermore, during the pandemic between the months of February and April, the housing market experienced major shutdowns due to implementation of stay at home orders. Once lifted, the housing market surged, increasing by a total of 23.1% by September 2020 (Wang, 2021).

Despite substantial literature on housing market fluctuations during both crises, current research rarely attempts to model market behavior in a predictive framework, mainly focusing on the immediate impacts the 2008 crisis and the 2020 pandemic had on the economic and housing markets. Additionally, the current literature is limited to the observation of specific economic and housing variables, mainly the relationship between home prices, mortgage rates, etc. The inclusion of broader macroeconomic indicators–such as GDP indexes, the consumer price index, the consumer confidence index, population, and unemployment measures–alongside additional housing market measures will enable a more precise analysis of how exactly the housing market both affects and is dependent on economic fluctuations.

The goals of this analysis include 1) reestablishing the connection between the housing market and fluctuations in the economy, specifically during the 2008 Financial Crisis and the COVID-19 Pandemic; 2) understanding how the housing market recovers from such economic recessions; 3) forecasting the future expectations of the housing market based on the present economic and housing markets; and 4) instituting the relationship between broader macroeconomic indicators and the housing market. That is, this study will examine to what extent the housing market is affected and affects the economy.

2 Data

2.1 Sample

Housing market data was collected from Zillow's research database, which represents all homes in the United States. For comparative analysis, the sample was set to four California counties, namely Santa Clara, Marin, Shasta, and Sutter, with a total of 252 observations from January 1, 2004 to December 31, 2024. Data on "all" homes is representative of the 35th to 65th percentile range, inclusive of single-family residences, condos, and coops. Some housing measures, such as housing inventory, listing counts, days on the market, and new construction metrics, are limited to 84 observations between January 1, 2018 to December 31, 2024. Sales counts is limited to 203 observations between February 1, 2008 to December 31, 2024. Different analyses will be run for the corresponding time periods. All other macroeconomic variables are representative of the United States and contain 252 observations from January 1, 2004 to December 31, 2024.

2.2 Measures

2.2.1 2008 Financial Crisis & COVID-19 Pandemic

A variable representing times of financial recessions was created using a dummy variable indicating times of crises, specifically from December 2007 to June 2009 to capture the 2008 Financial Crisis and February 2020 to April 2020 to capture the COVID-19 Pandemic.

2.2.2 Home Value

Housing market dynamics was assessed using the housing data collected by Zillow, specifically the Zillow Home Value Index (ZHVI). The ZHVI is calculated using an aggregate of Zestimates, which is Zillow's measure for assigning a value to each home. Furthermore, this data is smoothed and seasonally adjusted. The intention of the ZHVI is to accurately capture the price of a typical home by area, whether that be nationwide or city-level. The ZHVI also achieves three main ideas: timeliness, comprehensiveness, and visibility. Timeliness is achieved by the publication of monthly data on the last day of each month. This is an advantage over other housing market evaluators as the others publish data with significant lags between months. Next, the ZHVI encompasses all homes, inclusive of those that are newly constructed or homes that are vacant/not been listed for long periods of time. This allows for a more complete sample size of homes. Finally, the ZHVI is the most actionable, meaning that instead of only state or city-level data, more specific subsets are available, including zip code and neighborhood level samples.

2.2.3 CCI

The consumer confidence index (CCI) is a normalized confidence indicator from the OECD which indicates future expectations of household consumption and savings. The CCI is measured by a set of questions regarding a household's financial expectations, attitude towards the current economic situation, unemployment, and ability to save. The indicator is measured as an adjusted index with a long-term average of 100. A CCI above 100 indicates a boost in consumer confidence, where households are more likely to consume rather than save. On the other hand, values below 100 indicate a negative attitude towards the economic situation, resulting in a common bias to save rather than consume. The observed values range from 96.62451 to 101.18340.

2.2.4 CPI

The consumer price index (CPI) is used as a measure of inflation. Specifically, two baskets of goods are used, one which includes all items in the U.S., and the other which includes shelter. Both are collective of all urban consumers with the total basket being seasonally adjusted and shelter not seasonally adjusted. CPI is given in index values, set relative to a base period. The base period index is calculated by an average between 1982 and 1984 and is given the index value 100. Every sequential year measures the change relative to that average. To express as a percent change instead of index point change, the difference between sequential index points is divided by the most recent year's value and multiplied by 100. The observed values range from 186.3 to 317.603.

2.2.5 Mortgage Rates

30-year fixed rate mortgage averages are collected via the Primary Mortgage Market Survey conducted by Freddie Mac. The survey collects responses from different lender types: credit unions, commercial banks, and mortgage lending companies. Survey results are collected from lenders' applications and sent to Freddie Mac every time a borrower applies for a mortgage. This data is collected on a weekly basis and is aggregated to a monthly average. Furthermore, the units of these mortgage rates are percentages and not seasonally adjusted. The observed mortgage rates range from 2.65 percent to 7.79 percent.

2.2.6 Brave-Butters-Kelley Real GDP

GDP growth and economic activity expectations are measured by the Brave-Butters-Kelly Indexes (BBKI), which was formulated by economists at the Federal Reserve Bank of Chicago researchers at Indiana University with the goal of creating the most comprehensive measure of economic activity. The BBKI utilizes 490 measures of economic activity to generate three indexes that explain the current state of the economy, future expectations, and monthly GDP growth estimates. The first index, measured in standard deviation units, called the coincident index, measures the current economy, and it solves the problem of having a substantially lagged measure of GDP. With the coincident index, researchers are able to identify shifts in economic activity earlier. Next, the leading index is a component of the coincident index that signals the trajectory of economic activity in the future, typically the most accurate between six and seven months into the future. Lastly, monthly GDP is composed of three components: trend, which models the long-run average of real GDP growth, cycle, which models tendencies of business sectors, and irregular, which accounts for anything not captured by the trend and cycle components. The BBK real GDP is measured as an annualized percent change from the preceding time period and is seasonally adjusted. The observed values range from -72.458 percent to 44.987.

2.2.7 Population

Population is measured by the U.S. Bureau of Economic Analysis and captures resident population in the United States, also accounting for armed forces overseas. It is estimated monthly in thousands of people, ranging from 292,046 to 341,320.

2.2.8 Unemployment

The unemployment rate in both the United States is measured with the Labor Force Statistics from the Current Population Survey. The data is seasonally adjusted and is collected across all industries and occupations for all ethnicity origins, races, sexes, educational levels, and for those 16 years and older. The rate is presented as a percentage with its lowest and highest levels being 3.4 and 14.8 percent respectively. Furthermore, the unemployment rate in California is also used, and collected by the Local Area Unemployment Statistics survey. This data is also seasonally adjusted and measured as percentages. The observed values range from 3.8 percent to 16.1 percent.

2.3 Stationarity of Zillow Data

In order to proceed with time series analysis, the data must be stationary. Using both the ADF and KPSS tests, tests were performed on all housing related variables, and differenced when necessary. Both first and second differences were taken on the Zillow home data, and used when appropriate.

2.4 Decomposition of Santa Clara County Zillow Data

In an attempt to understand the trends of Santa Clara County housing prices, the home value data is decomposed into its trend, seasonal, and random parts, as seen in Figure 14. The trend decomposition aligns extremely closely with the observed data, with a yearly seasonal trend, and deviations from zero after 2020 in the random decomposition.

3 Time Series Models for Santa Clara County

3.1 Naive Models

First, the three naive models are used to model housing prices in Santa Clara county. Figure 1 displays each of the three models with the Santa Clara county home value data. The mean and naive models were not visually accurate, while the drift model did a somewhat decent job modeling the expected future home values. Each of the models was forecasted one year into the future.

3.2 ARIMA

Using the auto arima function, ARIMA models were generated using both the first and second differenced data. With the seasonal parameter set to TRUE, approximation and stepwise FALSE, ARIMA(5,1,0) and ARIMA(5,2,0) models were produced. Both ARIMA models yielded a 4766.453 AIC. Figure 2 shows the residual plot for the ARIMA(5,1,0), and Figure 3 shows the coefficients of each of the five AR terms. The first, third, and fourth AR terms are significant at the 0.1 percent level, with the fifth AR term significant at the five percent level. The ARIMA(5,1,0) and ARIMA(5,2,0) are as follows:

$$\phi(B)(1 - B)y_t = \epsilon_t \quad (3.1)$$

$$\phi(B)(1 - B)^2 y_t = \epsilon_t \tag{3.2}$$

3.3 SARIMAX

Building on the ARIMA models in the previous section, seven macroeconomic variables were included to build an SARIMAX model: GDP, California unemployment rate, consumer confidence index, consumer price index using all goods, consumer price index of only shelter, mortgage rates, and population. Using the auto arima function, the best model with an AIC of 4912.293 was a SARIMAX(2,0,1)(0,0,1)[12] as follows:

$$\phi(B)y_t = \theta(B)\Theta(B^{12})\epsilon_t + \sum_{i=1}^{7} \beta_i X_{i,t} \tag{3.3}$$

This model differs from the previous ARIMA models, using 2 AR terms, 1 MA term, and 1 seasonal MA term. Figure 15 displays the coefficients of the SARIMAX model, where all regular AR and MA terms, CCI, and population are significant, with the seasonal MA and other macroeconomic coefficients being statistically insignificant.

3.4 TAR

The best time series specific model is the TAR model, which makes sense as the four regime shifts due to the 2008 Financial Crisis and the COVID-19 Pandemic had major impacts on the housing market. To ensure that a TAR model is necessary, ACF and PACF plots and Tsay's test are utilized. Both ACF and PACF plots showed patterns experiencing instability, which hints at different behaviors between regimes. Tsay's test resulted in a small and significant p-value, which indicates that a TAR model is suitable. The AIC of the TAR model was 4094.64, good for the best time series model. The TAR model is as follows and is summarized in Figure 29:

$$y_t = \begin{cases} \text{Low Regime: } \phi_1(L)y_{t-1} + \epsilon_t & Z_t < -1043 \\ \text{High Regime: } \phi_2(L)y_{t-1} + \epsilon_t & Z_t \geq -1043 \end{cases} \tag{3.4}$$

The low regime model can be represented by an AR(1) model with a statistically significant first lag at the 0.1 percent level as follows:

$$santaclara_t = -306.016 + 0.855 santaclara_{t-1} + \epsilon_t \tag{3.5}$$

The high regime model is an AR(12) model with the first through fourth and seventh lagged terms significant at the 0.1 percent level, the fifth and eighth significant at the five percent level, and the sixth significant at the 10 percent level as follows:

$$santaclara_t = 441.364 + \sum_{i=1}^{12} \phi_i santaclara_{t-i} + \epsilon_t \qquad (3.6)$$

Figure 24 shows the differenced Santa Clara time series along with its residuals. The ACF and PACF plots are shown in Figure 25 and 26, respectively. Next, Figure 27 shows the regime switching plot, with the black lines representing observations in the low regime and red lines representing those in the high regime. Finally, the threshold variable plot of the manual model is shown in Figure 28.

4 Regression Analysis

4.1 Linear Regression

I start my regression analysis with a simple linear regression of seven macroeconomic variables on home value. These macroeconomic variables are the same as the ones used in the aforementioned ARIMAX model. Date is also included in the linear model to be consistent with the rest of the regression models. For the remainder of this section, the macroeconomic variables will be defined as:

$X_{1,t} = bbkgdp_t,\ X_{2,t} = caunemp_t,\ X_{3,t} = cci_t,\ X_{4,t} = cpiall_t,\ X_{5,t} = cpishelter_t,\ X_{6,t} = mrates_t,\ X_{7,t} = population_t$

The first linear regression is modeled as follows:

$$santaclara_t = \beta_0 + \sum_{i=1}^{7} \beta_i X_{i,t} + \gamma t + \epsilon_t \qquad (4.1)$$

where $santaclara_t$ represents the average home value in Santa Clara County, β_0 denotes the intercept, β_i the coefficients on the macroeconomic variables, γ the coefficient on time t, and error term ϵ_t.

Equation (4.1) produced an R-squared of 0.9799, which is extremely strong. As shown in Figure 8, all macroeconomic variables except GDP and CCI were significant at the 0.1 percent variable.

4.2 Polynomial Regression

Next, I continue with a polynomial regression, where time squared t^2 is added to Equation (4.1) to get:

$$santaclara_t = \beta_0 + \sum_{i=1}^{7} \beta_i X_{i,t} + \gamma_1 t + \gamma_2 t^2 + \epsilon_t \qquad (4.2)$$

where the variables and coefficients are consistent with Equation (4.1). This polynomial regression produced an R-squared of 0.9801, slightly higher than Equation (4.1), indicating that the inclusion of a non-linear time term explains more of the variance present in Santa Clara County home values. Surprisingly, as shown in Figure 9, the squared time term is not statistically significant, and consistent with the findings in Equation (4.1), GDP and CCI are the only variables not statistically significant at the 0.1 percent level. The higher R-squared contradicts the significance of the second degree time term, though this phenomenon could be explained by the explanatory power of other macroeconomic variables in conjunction with the two new time variables. Furthermore, the squared time term has a negative coefficient, which does not align with intuition, so no significance of this term can be accepted.

4.3 Fourier Regression

Expanding on the inclusion of a nonlinear time term, a Fourier regression is included as follows:

$$santaclara_t = \beta_0 + \sum_{i=1}^{7} \beta_i X_{i,t} + \gamma_1 cos(\frac{2\pi t}{12}) + \gamma_2 sin(\frac{2\pi t}{12}) + \gamma_3 cos(\frac{4\pi t}{12}) + \gamma_4 sin(\frac{4\pi t}{12}) + \epsilon_t \tag{4.3}$$

where now γ_i are the coefficients on the Fourier terms, with the other variables and coefficients consistent with the two previous models. Equation (4.3) produced a lower R-squared value at 0.9596, which even though is lower than Equations (4.1) and (4.2), still displayed strong modeling power. As shown in Figure 10, none of the Fourier coefficients are significant, while GDP and CCI continue to be insignificant. Additionally, CPI of all goods is now only significant at the one percent level.

4.4 Splines Regressions

To demonstrate the effects of the 2008 Recession and the COVID-19 Pandemic, three different spline regressions are used. There are four splines to each model: signaling the start and end of each recession, specifically December 2007, June 2009, February 2020, and April 2020.

4.4.1 Linear Spline

First, a linear spline regression with degree one is implemented by the following:

$$santaclara_t = \beta_0 + \sum_{i=1}^{7} \beta_i X_{i,t} + \gamma t + \sum_{j=1}^{4} \delta_j (t - k_j) + \epsilon_t \tag{4.4}$$

where $\sum_{j=1}^{4} \delta_j (t - k_j)$ represents the sum of the four basis functions and the remaining variables and coefficients are the same as the previous equations. Equation (4.4) produced the smallest R-squared of the three spline regressions at 0.9835, though the best regression so far. Figure 11 displays the coefficients and results from Equation (4.4), where each basis function is significant at the 0.1 percent significance level. Furthermore, CPI with shelter only loses significance while GDP gains significance at the five percent level. CCI maintains no significance.

4.4.2 Quadratic Spline

Next, quadratic splines are used with the same variables as Equation (4) as follows:

$$santaclara_t = \beta_0 + \sum_{i=1}^{7} \beta_i X_{i,t} + \gamma_1 t + \gamma_2 t^2 + \sum_{j=1}^{4} \delta_j (t - k_j)^2 + \epsilon_t \tag{4.5}$$

where $\sum_{j=1}^{4} \delta_j (t - k_j)^2$ represents the new squared basis functions, as opposed to linear as in Equation (4.4). Additionally, the squared t term is included, similar to Equation (4.2). Results in Figure 12 show that all basis functions are significant at the 0.1 percent level. CCI is now significant at the five percent level, and GDP and population are insignificant. Equation (4.5) outperforms Equation (4.4) with a 0.9871 R-squared.

4.4.3 Cubic Spline

Finally, cubic splines are implemented, along with a cubed t term:

$$santaclara_t = \beta_0 + \sum_{i=1}^{7} \beta_i X_{i,t} + \gamma_1 t + \gamma_2 t^2 + \gamma_3 t^3 + \sum_{j=1}^{4} \delta_j (t - k_j)^3 + \epsilon_t \tag{4.6}$$

where $\sum_{j=1}^{4} \delta_j (t - k_j)^3$ represents the cubic basis functions and t^3 the cubed time term. This equation performed the best out of all the regressions, yielding a 0.9897 R-squared value. In Figure 13, every coefficient with the exception of CPI (all) was statistically significant at some level, with GDP and CCI producing significance at the 10 percent level. Population had significance at the one percent level, and the rest of the coefficients, including the basis functions, had significance at the 0.1 percent level.

4.5 Regression with Recession Dummies

Finally, to observe the effects that the 2008 Recession and the COVID-19 Pandemic had on housing prices, the dummy variables for each of the two recessions are added to the linear and polynomial regressions only, since the fourier regression did not perform as well and the spline regressions already account for regime switches. Equations 4.1 and 4.2 are modified to obtain:

$$santaclara_t = \beta_0 + \sum_{i=1}^{7} \beta_i X_{i,t} + \gamma t + \delta_1 D_{crisis} + \delta_2 D_{covid} + \epsilon_t \tag{4.7}$$

$$santaclara_t = \beta_0 + \sum_{i=1}^{7} \beta_i X_{i,t} + \gamma_1 t + \gamma_2 t^2 + \delta_1 D_{crisis} + \delta_2 D_{covid} + \epsilon_t \tag{4.8}$$

Both of these regressions outperformed their counterparts without dummies, yielding an identical R-squared of 0.9802. Figures 35 and 36 show the summaries of each of the regressions. In Equation (4.7), CCI and both dummy variables are the only coefficients that are not statistically significant. GDP is significant at the 10 percent level while the rest are statistically significant at the 0.1 percent level. The signs of each of the dummy variables aligns with the expectations of the housing market during the recession and pandemic. Although there is no statistical significance, it is expected for house prices to increase during the crisis and for Santa Clara County, slightly decrease during the pandemic. Equation (4.8) produces similar results, where

CCI, the squared time term, and the dummy variables are insignificant, with GDP and CPI shelter yielding 10 and five percent significance, respectively. All other coefficients are significant at the 0.1 percent level. The coefficients on each of the dummy variables have the same sign as those in Equation (4.7), confirming the real-world findings (though statistically insignificant).

5 Extended Models

5.1 Clustering

First, k-means clustering was used across all counties, accounting for housing prices. Using the elbow and gap statistic methods, three clusters were used. Figure 33 displays each of the clusters and the respective counties. Next, to identity the county that is most central to the California housing market, k-medoids clustering was used, and Shasta County was found to be the most central. Figure 34 shows each of the counties, with Shasta being the most central.

5.2 Cross-County Analysis

5.2.1 VAR

In order to determine causality between counties, I run a VAR model on house prices only, including all 35 counties in California. Using VARselect, the number of lags to be used was found to be 6. For Santa Clara County specifically, from the Granger Causality test, it Granger causes the other 34 counties, and further displays presence of instantaneous causality with the other counties.

5.2.2 Impulse Response Functions

Extending the VAR model and foreshadowing cross-county analysis, four impulse response functions are created to describe the dependencies of counties on a base county, in this case Santa Clara County. The four chosen counties are Los Angeles, San Francisco, Sutter, and Shasta. These counties were chosen based on both geographical and demographic characteristics. Los Angeles and San Francisco are both extremely urban and populated counties, while Sutter and Shasta are not.

Santa Clara County & Los Angeles County

Figure 16 shows the impulse response function from Santa Clara on Los Angeles. The 95% confidence interval displays a somewhat increasing trend across the first 11 periods, with the impulse response function following the same increasing trend but deviating far from the interval. This means that due to a one deviation shock in housing prices in Santa Clara County, home prices in Los Angeles County experience a positive response which is statistically significant at the five percent level. To ensure robustness and accuracy, Figure 20 shows the impulse response function from Los Angeles on Santa Clara. In this case, the effect is almost opposite, with both the confidence interval and impulse response function trending downwards for 10 periods. Similarly to Figure 16, results are statistically significant at the five percent level, this time from the 6th period onwards. These results display that a shock in Los Angeles housing prices causes a negative response from the Santa Clara County housing market. These impulse response functions being asymmetric implies that the two housing markets are not reciprocal and therefore reflect three different common housing aspects: 1) there is economic, geographical, or demographic hierarchy between the two counties, 2) substitution effects present between the two counties, and 3) investment or migration responses based on market conditions and/or constraints.

Santa Clara County & San Francisco County

Figure 17 and Figure 21 show the two impulse response functions between the two counties, with a negative effect from Santa Clara on San Francisco and a positive effect from San Francisco to Santa Clara. However, only the results in Figure 17 are statistically significant at the five percent level, displaying a negative response from an impulse in the Santa Clara County market. Figure 21 displays initial statistical significance with a negative response to a shock in the San Francisco market, though when trending upwards becomes insignificant. This is consistent with expectations in the real world, as both counties being geographically close causes competition between markets. When Santa Clara is used as the impulse, the response by the San Francisco market due to a positive shock in Santa Clara ends up negative. That is, when house prices in Santa Clara increase, one can expect a decrease in prices in San Francisco, a production of a competitive market. Since markets have additional unique characteristics, for example San Francisco being extremely urban, symmetrical responses are not expected, and thus the results from this set of impulse response functions is satisfactory.

Santa Clara County, Sutter County, & Shasta County

For Sutter and Shasta County, the expectation of effects of housing markets is that there may be minimal impact of these counties on larger, more economically developed ones such as Santa Clara. Figures 18 and 19 display the response functions due to a shock in the housing market of Santa Clara. Both show almost identical functions, being mainly statistically significant at the five percent level, and constantly increasing across the 12 month period. This indicates that shocks in the Santa Clara market have overall positive effects on both Sutter and Shasta County house prices. Figure 22 displays the impulse response from Sutter County on Santa Clara County. The results are not statistically significant, except between the 6th and 7th periods, though this phenomenon could be the result of a consistently increasing function that happens to pass through the confidence interval at one period. This is probably as there is no change in the slope when passing through the confidence interval. That is, a change in the slope in the confidence interval could suggest significance during that period. Similarly, in Figure 23, there is no significance except possibly in period 12. Though the impulse response function passes through the confidence interval between periods 4 and 5, there is no change in slope of the response function though the interval. The shape is identical to that in Figure 22, decreasing in the beginning, only to increase though the confidence interval and remain positive. Furthermore, Santa Clara County experiences a moment of statistical significance around the 11th period, where the impulse response function decreases into the confidence interval and remains constant and positive for about one period. This could suggest an extremely delayed response, which is unusual but not implausible. Some possible economic and statistical interpretations are: 1) Shasta County acting as a delayed feedback mechanism, where a boom in a more developed county like Santa Clara takes time to express its effect on delayed markets, 2) information spillovers, where rural counties may signal anticipatory investment that takes time to set in, or 3) reflection of noise that becomes statistically significant by chance and therefore can be disregarded.

5.3 Markov Switching Model

In order to better capture the regime switches, a Markov Switching Model is implemented to build off of the previously discussed TAR model. First, an AR(1) linear model is fit, inclusive of the lagged value and time, using the second differenced data. Applying a Markov Switching model with parameters k = 2 and p = 0 with sw parameters turned to TRUE, the model and its probabilities were implemented. In Figures 30 and 31, respectively, the smoothed probabilities of the two regimes were plotted against time, with the red and purple blocks indicating the 2008 Recession and the COVID-19 Pandemic. Visually, Regime 1 dominates during the late Financial Crisis and parts of the COVID-19 Pandemic. Regime 2 dominates before the 2008 Crisis and after, during the recovery period in 2010. It also briefly dominates during the COVID-19 Pandemic. This alternating pattern between the two regimes suggest that the first regime describes times of crises, dominating during 2008 and decreasing around the pandemic, and the second regime describes stability, dominating during the times around 2008 and during the pandemic. It is important to note the inverse outcomes between the 2008 Financial Crisis and the COVID-19 pandemic. This is consistent with the housing prices in Santa Clara, which had a major decrease in 2008, while during COVID-19 actually slightly increased. This possibly explains the alternation between spikes in probabilities of the two regimes. Figure 32 displays the most likely regimes to show up during each of the periods.

6 Cross Validation

Cross validation was performed on Equations (3.1) and (3.4), the ARIMA and TAR models, respectively, as well as the VAR model. Using rolling forecast origin methods for each of the three models, the RMSE, MAE, and MAPE were analyzed. The ARIMA and VAR models performed sufficiently, while the TAR model struggled, producing 1172886 RMSE, 1140232 MAE, and 98.973 MAPE. The ARIMA and VAR models produced 3669.345 RMSE, 2761.529 MAE, and 0.343 MAPE; and 11311.1 RMSE, 7859.535 MAE, and 0.763 MAPE, respectively. For the TAR and VAR models, 120 months (or 10 years) of data was used to train the model, allowing for substantial time before testing the model.

7 Limitations

The biggest limitation I faced during this analysis was time. In order to completely capture the housing market dynamics in California, I would have needed to replicate each model for the 35 counties, something I did not have the time to do. Moreover, once the models were run for all counties, I would have performed comparative analysis across the counties. Next, after performing clustering techniques, I wanted to rerun the models using Shasta County as the focus, as k-medoids clustering listed Shasta County as the county that was most central to the dataset. Finally, I had originally included more housing variables that captured factors such as housing inventory and the time it took for houses to sell and how long they were on the market for. However, these variables were only available in limited time samples, and thus could not effectively be used in conjunction with the rest of the macroeconomic variables included in this analysis.

8 Conclusions

This paper provides an introductory analysis of California housing data using time series methods, specifically ARIMA, SARIMAX, TAR, and VAR models and linear, polynomial, fourier, and splines regression models to model the housing market from 2004 to 2024. Due to time limitations, Santa Clara County was the focus of this analysis, chosen due to the author's residency in the county. The TAR model performed the best out of the four time series specific methods. The splines regression model both performed the best and was best suited to capture regime shifts due to the two recessions. Furthermore, advanced modeling techniques were used, such as clustering and a Markov Switching Model. Finally, cross validation was performed on a few models, the ARIMA, TAR, and VAR models, with the ARIMA being the best performing model under cross validation. Results from the regressions were the most telling of the macroeconomic implications of different variables on the housing market. Most of the macroeconomic variables were statistically significant, displaying the connection between the broader U.S. economy and California's housing market. Surprisingly, dummy variables accounting for times of recession/crisis were never statistically significant, though we know from domain knowledge that the 2008 Crisis and the COVID-19 Pandemic certainly had implications for shocks in the housing market. This analysis was successful in achieving the four aforementioned goals. 1) The analysis reestablished the connection between the housing market and fluctuations in the economy, represented by shocks in the housing market due to the two regime changes. 2) The analysis visually represented the housing market recovery cycles around the 2008 Financial Crisis and the COVID-19 Pandemic. 3) The analysis utilized forecasting methods to predict the expectations of the housing market. And 4) the analysis connected macroeconomic indicators and housing prices through the use of regression analysis.

9 Appendix

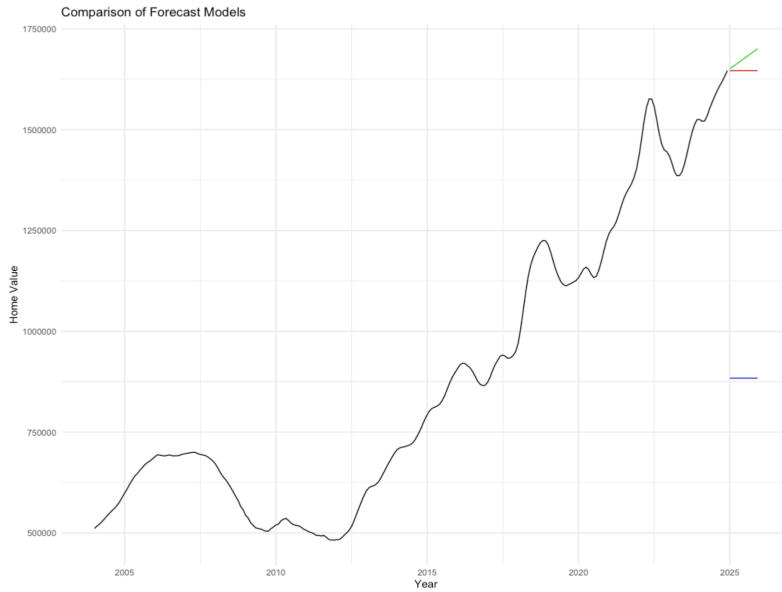

Figure 1: Naive Models

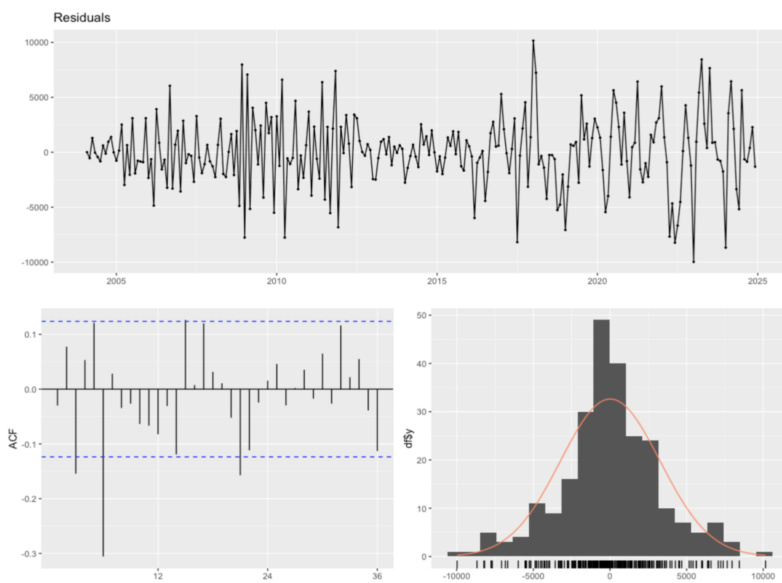

Figure 2: Residuals from ARIMA Model

```
     Estimate Std. Error z value  Pr(>|z|)
ar1  0.771131   0.062270 12.3836 < 2.2e-16 ***
ar2  0.014615   0.076957  0.1899 0.8493801
ar3 -0.538887   0.068842 -7.8278 4.963e-15 ***
ar4  0.283866   0.076940  3.6894 0.0002247 ***
ar5 -0.153255   0.062005 -2.4717 0.0134483 *
```

Figure 3: Coefficients from ARIMA Model

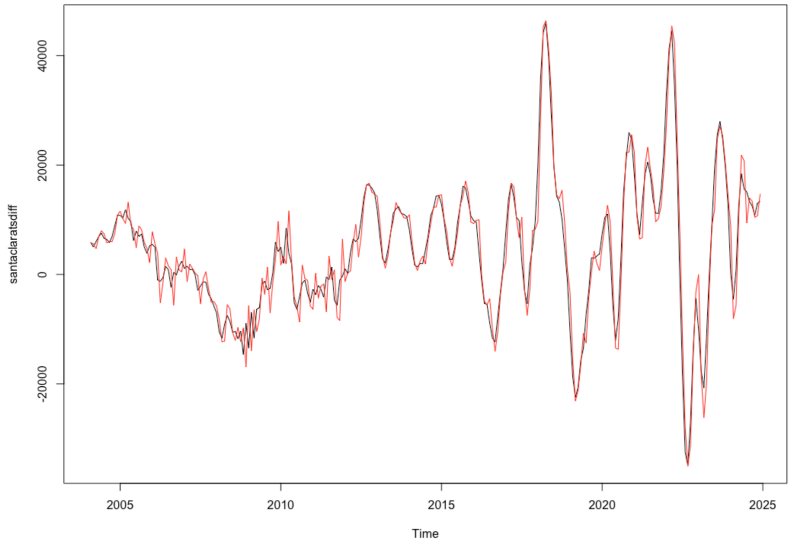

Figure 4: Overlay of ARIMA Using First Difference Data

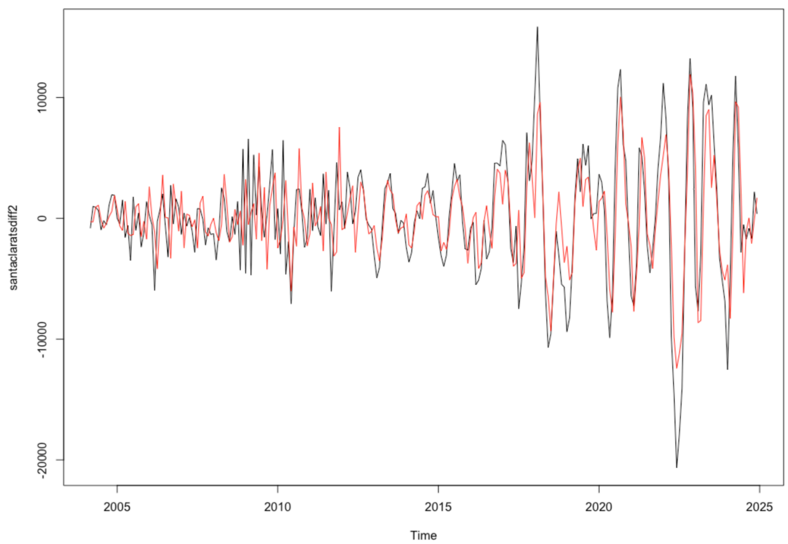

Figure 5: Overlay of ARIMA Using Second Difference Data

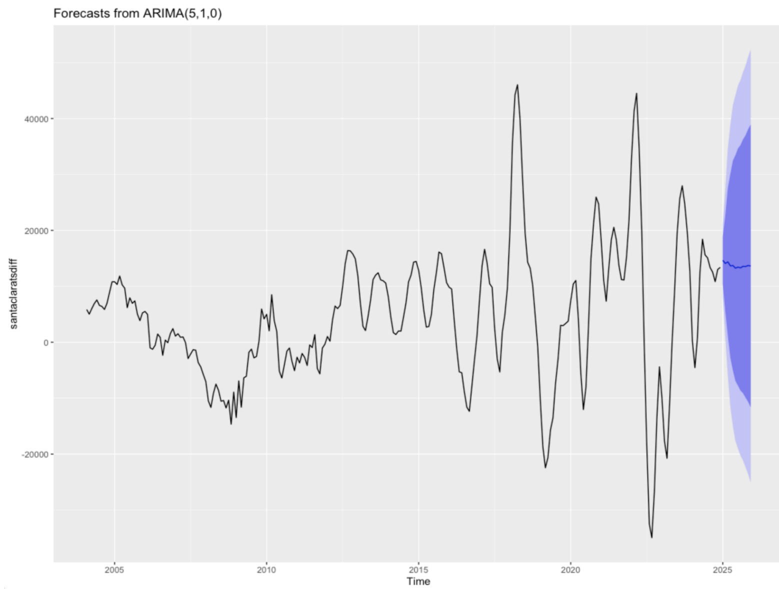

Figure 6: Forecast of ARIMA Using First Difference Data

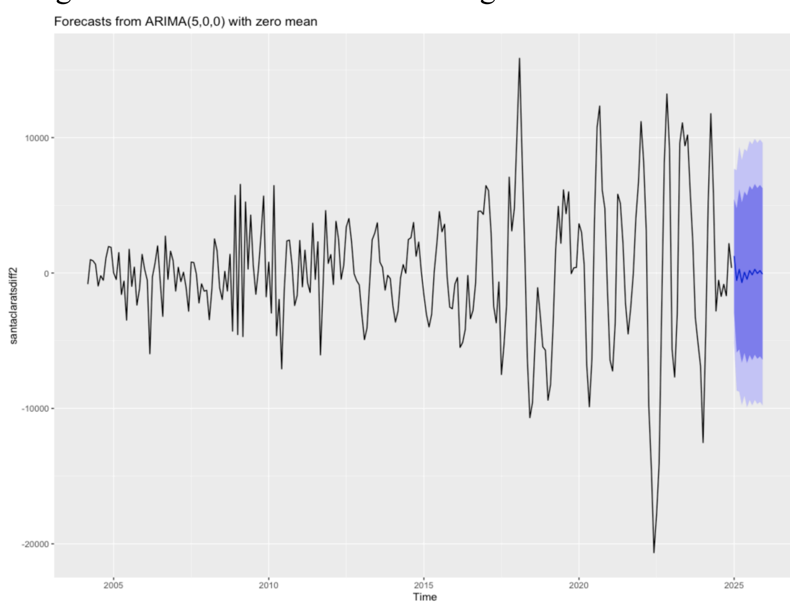

Figure 7: Forecast of ARIMA Using Second Difference Data

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	1.347e+07	8.078e+05	16.681	< 2e-16	***
bbkgdp	-7.665e+02	4.378e+02	-1.751	0.0812	.
caunemp	-1.979e+04	1.846e+03	-10.722	< 2e-16	***
cci	2.137e+03	4.645e+03	0.460	0.6459	
cpiall	-7.520e+03	9.103e+02	-8.261	9.39e-15	***
cpishelter	3.493e+03	7.053e+02	4.953	1.37e-06	***
mrates	3.722e+04	8.012e+03	4.645	5.57e-06	***
population	-6.864e+01	3.855e+00	-17.803	< 2e-16	***
date	6.032e+02	3.837e+01	15.720	< 2e-16	***

Figure 8: Regression Results from Equation (4.1)

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	2.546e+07	2.324e+06	10.955	< 2e-16	***
bbkgdp	-7.739e+02	4.376e+02	-1.769	0.078211	.
caunemp	-1.986e+04	1.846e+03	-10.761	< 2e-16	***
cci	4.429e+03	5.056e+03	0.876	0.381980	
cpiall	-7.803e+03	9.429e+02	-8.276	8.62e-15	***
cpishelter	4.595e+03	1.194e+03	3.848	0.000153	***
mrates	3.565e+04	8.123e+03	4.389	1.70e-05	***
population	-7.692e+01	8.205e+00	-9.374	< 2e-16	***
poly(date, 2)1	2.233e+07	1.671e+06	13.362	< 2e-16	***
poly(date, 2)2	-3.348e+05	2.929e+05	-1.143	0.254133	

Figure 9: Regression Results from Equation (4.2)

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	3937543.68	763735.27	5.156	5.28e-07	***
bbkgdp	-217.34	629.62	-0.345	0.7303	
caunemp	-31060.53	2458.23	-12.635	< 2e-16	***
cci	-9739.08	6593.75	-1.477	0.1410	
cpiall	-3456.64	1259.08	-2.745	0.0065	**
cpishelter	11509.90	697.41	16.504	< 2e-16	***
mrates	-51026.10	8266.32	-6.173	2.84e-09	***
population	-12.79	2.20	-5.814	1.94e-08	***
cos(2 * pi * date/12)	-1169.59	6396.71	-0.183	0.8551	
sin(2 * pi * date/12)	-62.32	6372.19	-0.010	0.9922	
cos(4 * pi * date/12)	-4321.29	6637.12	-0.651	0.5156	
sin(4 * pi * date/12)	-1940.21	6400.63	-0.303	0.7621	

Figure 10: Regression Results from Equation (4.3)

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	2.354e+07	3.735e+06	6.304	1.38e-09	***
bbkgdp	-8.380e+02	4.029e+02	-2.080	0.0386	*
caunemp	-1.478e+04	3.602e+03	-4.104	5.58e-05	***
cci	-4.590e+03	6.260e+03	-0.733	0.4641	
cpiall	-1.276e+04	2.369e+03	-5.385	1.73e-07	***
cpishelter	1.886e+02	1.161e+03	0.162	0.8711	
mrates	4.417e+04	8.108e+03	5.447	1.27e-07	***
population	-6.973e+01	1.044e+01	-6.679	1.67e-10	***
bs(date, knots = c(13878, 14425, 18321, 18382), degree = 1)1	1.209e+06	1.801e+05	6.714	1.36e-10	***
bs(date, knots = c(13878, 14425, 18321, 18382), degree = 1)2	1.445e+06	2.149e+05	6.725	1.29e-10	***
bs(date, knots = c(13878, 14425, 18321, 18382), degree = 1)3	4.374e+06	6.363e+05	6.874	5.39e-11	***
bs(date, knots = c(13878, 14425, 18321, 18382), degree = 1)4	4.377e+06	5.903e+05	7.414	2.11e-12	***
bs(date, knots = c(13878, 14425, 18321, 18382), degree = 1)5	6.048e+06	8.942e+05	6.763	1.03e-10	***

Figure 11: Regression Results from Equation (4.4)

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	1.506e+07	6.369e+06	2.364	0.018864	*
bbkgdp	-5.922e+02	3.593e+02	-1.648	0.100591	
caunemp	-2.592e+04	1.955e+03	-13.257	< 2e-16	***
cci	-1.014e+04	4.807e+03	-2.110	0.035941	*
cpiall	-1.016e+04	1.775e+03	-5.721	3.16e-08	***
cpishelter	-1.202e+04	2.081e+03	-5.779	2.35e-08	***
mrates	5.570e+04	7.835e+03	7.109	1.35e-11	***
population	-3.151e+01	2.192e+01	-1.438	0.151780	
bs(date, knots = c(13878, 14425, 18321, 18382), degree = 2)1	6.232e+05	1.161e+05	5.366	1.91e-07	***
bs(date, knots = c(13878, 14425, 18321, 18382), degree = 2)2	1.192e+06	2.867e+05	4.157	4.49e-05	***
bs(date, knots = c(13878, 14425, 18321, 18382), degree = 2)3	2.194e+06	6.582e+05	3.334	0.000994	***
bs(date, knots = c(13878, 14425, 18321, 18382), degree = 2)4	4.102e+06	8.536e+05	4.806	2.72e-06	***
bs(date, knots = c(13878, 14425, 18321, 18382), degree = 2)5	4.885e+06	8.874e+05	5.505	9.55e-08	***
bs(date, knots = c(13878, 14425, 18321, 18382), degree = 2)6	6.255e+06	1.071e+06	5.841	1.69e-08	***

Figure 12: Regression Results from Equation (4.5)

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	2.044e+07	4.707e+06	4.343	2.08e-05	***
bbkgdp	-6.206e+02	3.184e+02	-1.949	0.052455	.
caunemp	-9.189e+03	2.716e+03	-3.384	0.000837	***
cci	-7.193e+03	4.248e+03	-1.693	0.091741	.
cpiall	-2.409e+03	1.914e+03	-1.259	0.209347	
cpishelter	-1.370e+04	2.143e+03	-6.392	8.62e-10	***
mrates	4.471e+04	7.201e+03	6.209	2.37e-09	***
population	-5.471e+01	1.686e+01	-3.245	0.001345	**
bs(date, knots = c(13878, 14425, 18321, 18382), degree = 3)1	3.294e+05	6.831e+04	4.822	2.54e-06	***
bs(date, knots = c(13878, 14425, 18321, 18382), degree = 3)2	1.120e+06	1.339e+05	8.360	5.38e-15	***
bs(date, knots = c(13878, 14425, 18321, 18382), degree = 3)3	1.647e+06	3.823e+05	4.308	2.41e-05	***
bs(date, knots = c(13878, 14425, 18321, 18382), degree = 3)4	3.526e+06	5.267e+05	6.694	1.56e-10	***
bs(date, knots = c(13878, 14425, 18321, 18382), degree = 3)5	4.997e+06	6.269e+05	7.971	6.65e-14	***
bs(date, knots = c(13878, 14425, 18321, 18382), degree = 3)6	5.707e+06	6.548e+05	8.716	5.10e-16	***
bs(date, knots = c(13878, 14425, 18321, 18382), degree = 3)7	6.724e+06	7.420e+05	9.062	< 2e-16	***

Figure 13: Regression Results from Equation (4.6)

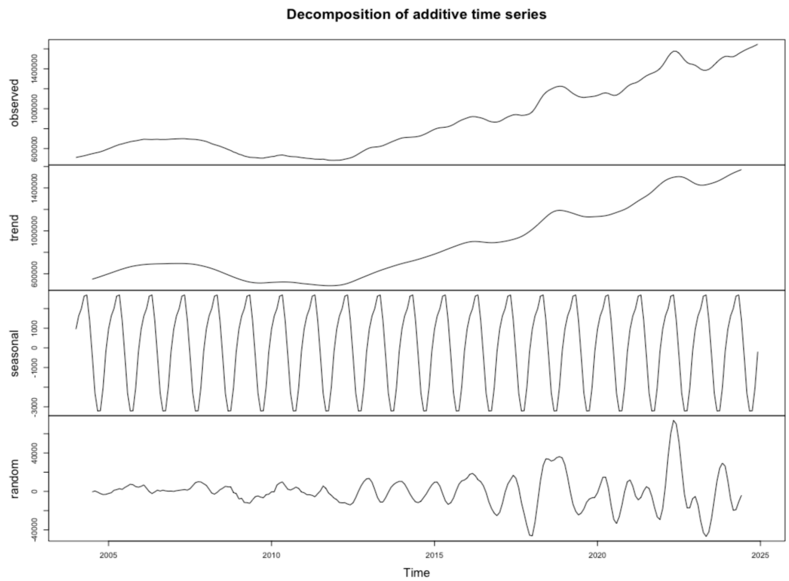

Figure 14: Decomposition of Santa Clara County Data

	Estimate	Std. Error	z value	Pr(>|z|)	
ar1	1.8919e+00	2.8443e-02	66.5157	< 2.2e-16	***
ar2	-8.9519e-01	2.8517e-02	-31.3911	< 2.2e-16	***
ma1	4.8528e-01	4.1563e-02	11.6756	< 2.2e-16	***
sma1	-1.2365e-01	7.6019e-02	-1.6266	0.1038248	
intercept	-5.0913e+06	1.4457e+06	-3.5216	0.0004289	***
bbkgdp	2.0535e+01	4.4231e+01	0.4643	0.6424484	
caunemp	-1.0750e+02	3.0854e+02	-0.3484	0.7275172	
cci	-8.9161e+03	3.2376e+03	-2.7539	0.0058882	**
cpiall	2.3292e+02	3.0961e+02	0.7523	0.4518819	
cpishelter	-1.7259e+02	3.5812e+02	-0.4819	0.6298456	
mrates	-1.1619e+03	9.4959e+02	-1.2236	0.2211174	
population	2.1590e+01	4.5255e+00	4.7707	1.836e-06	***

Figure 15: Coefficients from SARIMAX Model

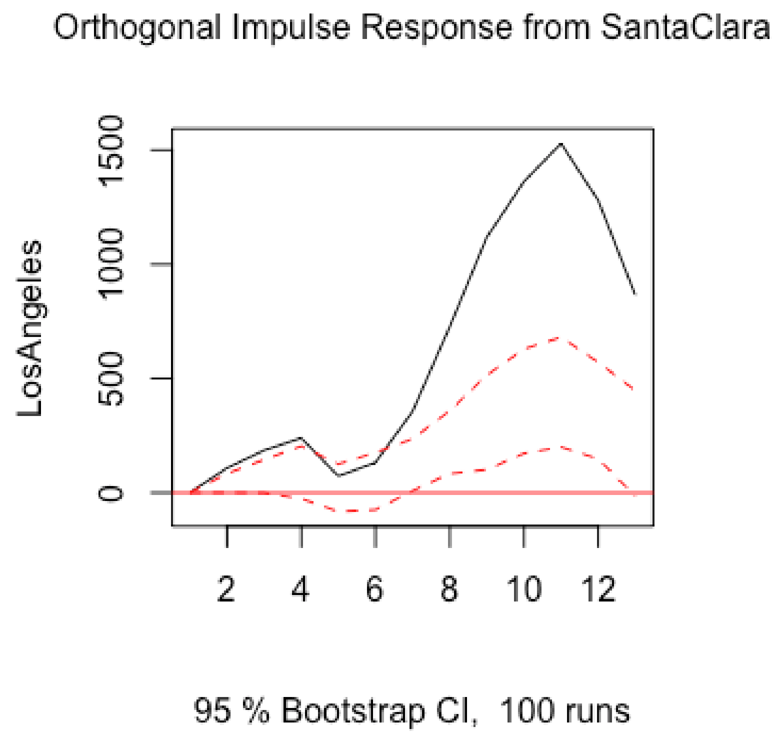

Figure 16: Impulse Response from Santa Clara on Los Angeles

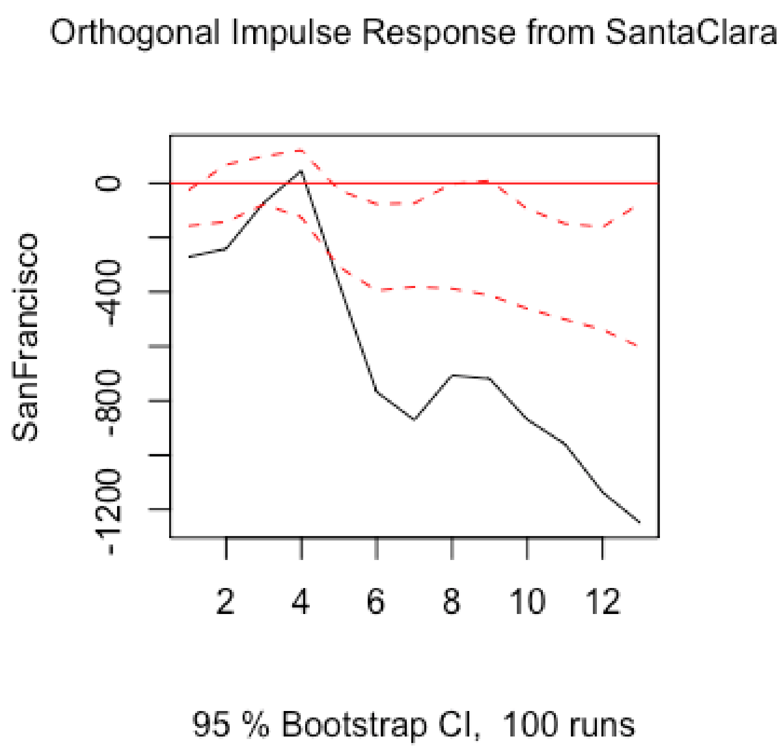

Figure 17: Impulse Response from Santa Clara on San Francisco

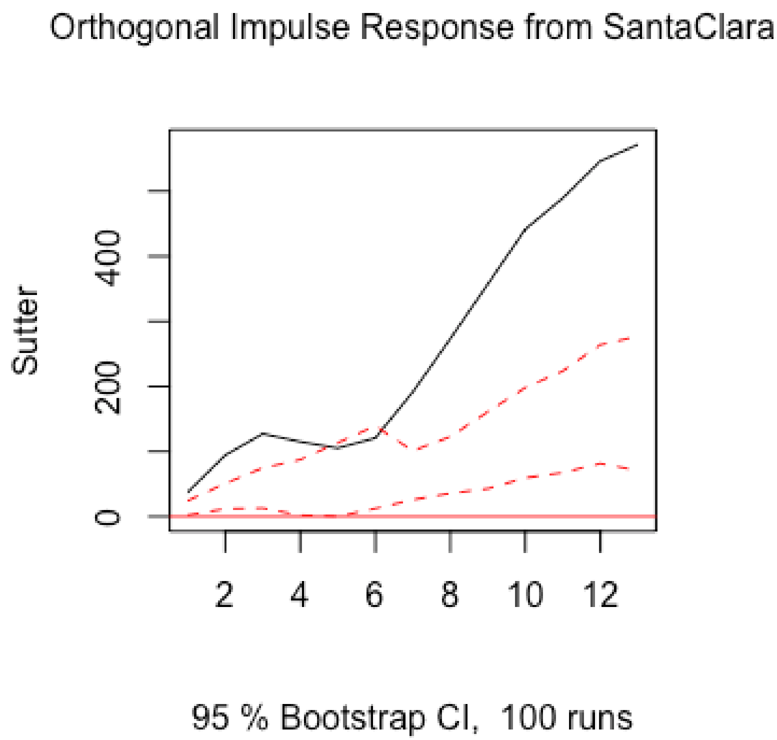

Figure 18: Impulse Response from Santa Clara on Sutter

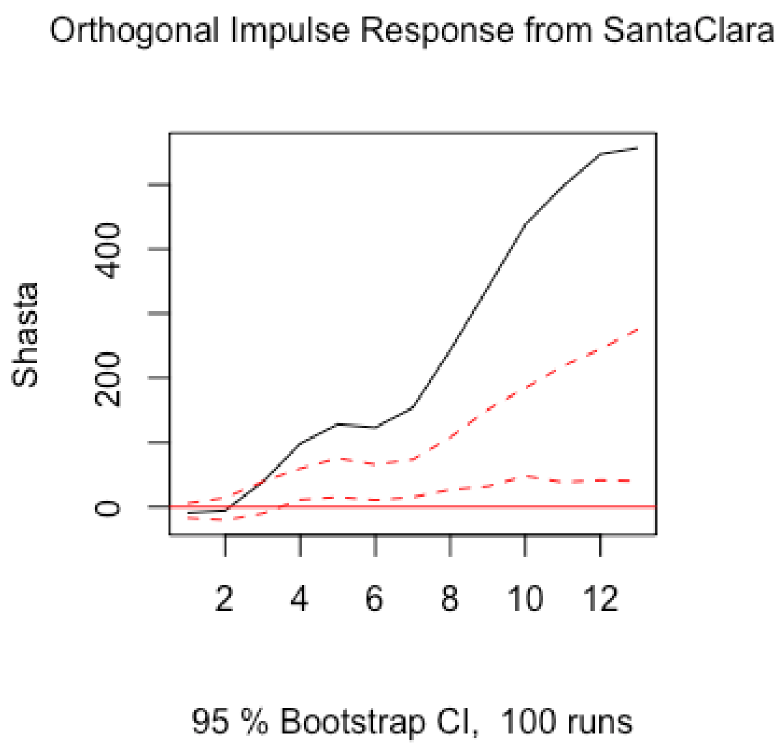

Figure 19: Impulse Response from Santa Clara on Shasta

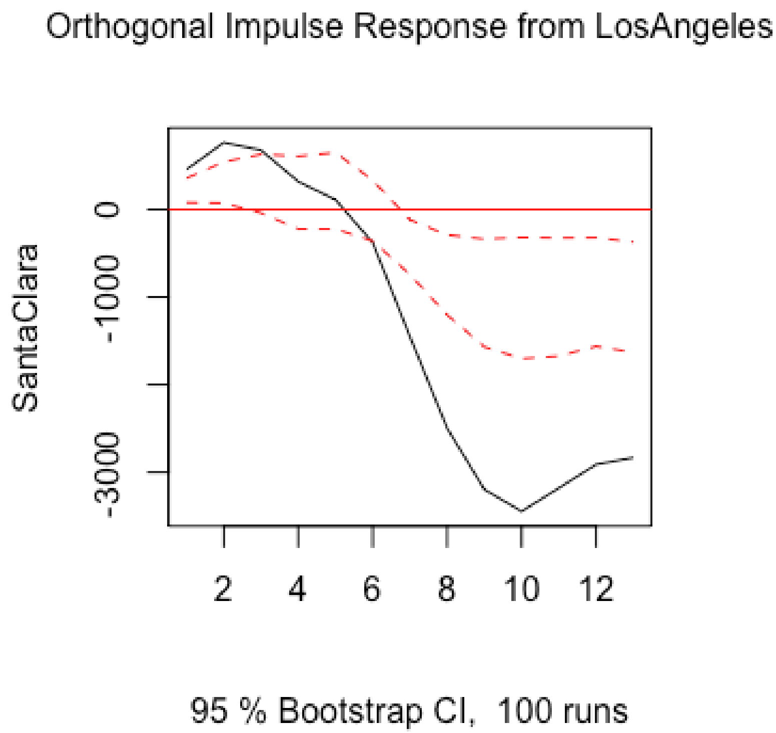

Figure 20: Impulse Response from Los Angeles on Santa Clara

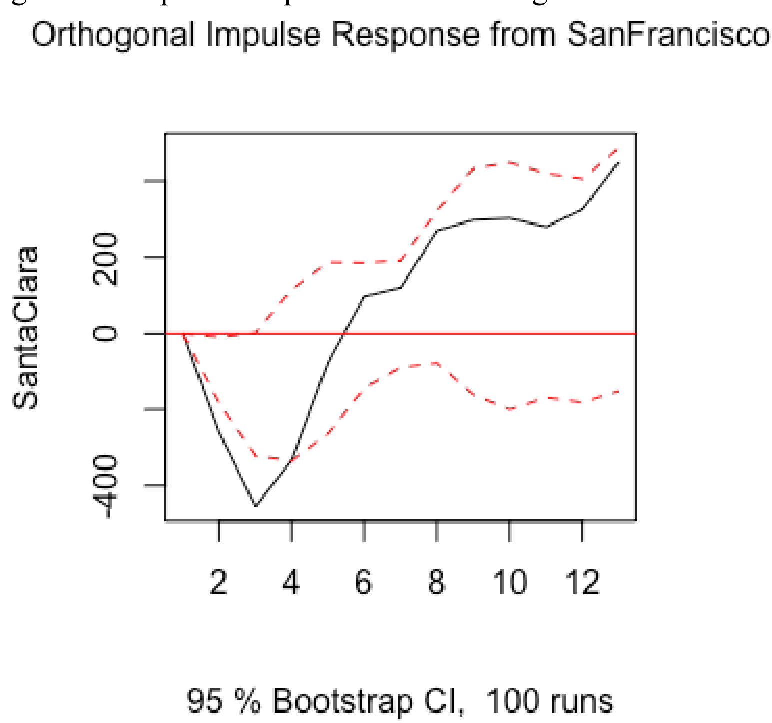

Figure 21: Impulse Response from San Francisco on Santa Clara

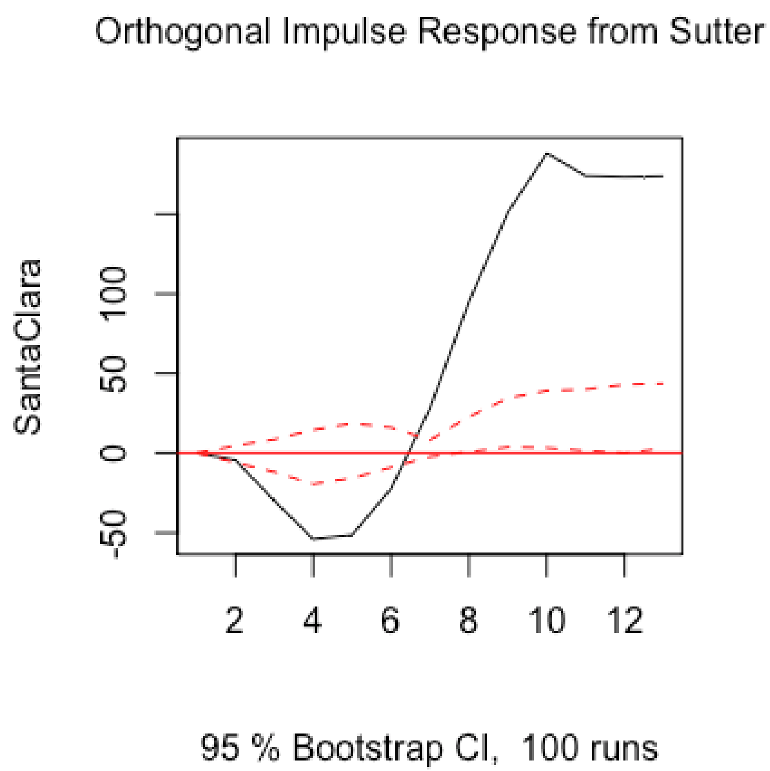

Figure 22: Impulse Response from Sutter on Santa Clara

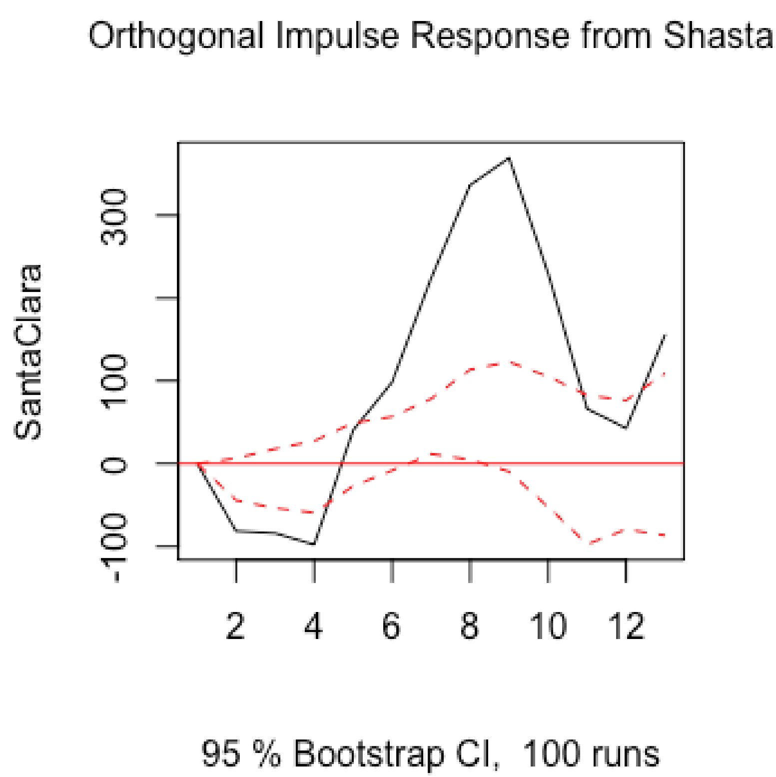

Figure 23: Impulse Response from Shasta on Santa Clara

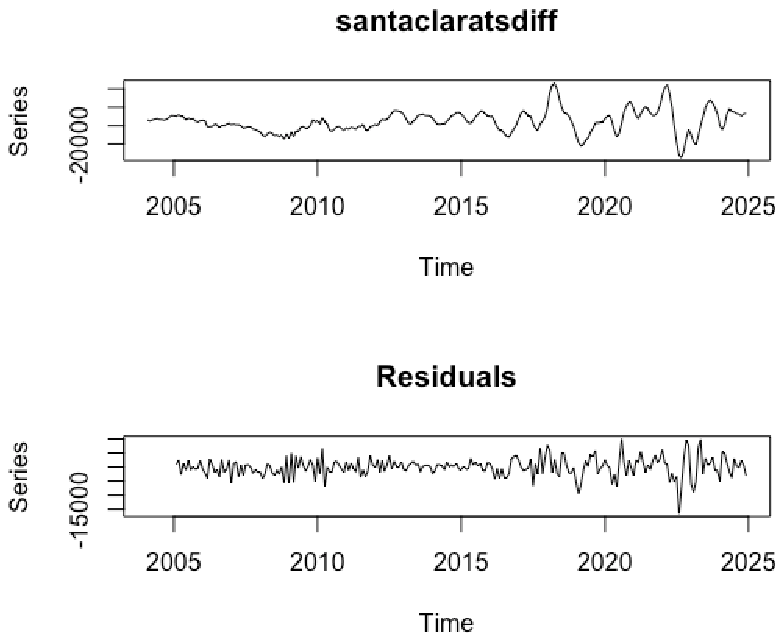

Figure 24: TAR Model & Residuals

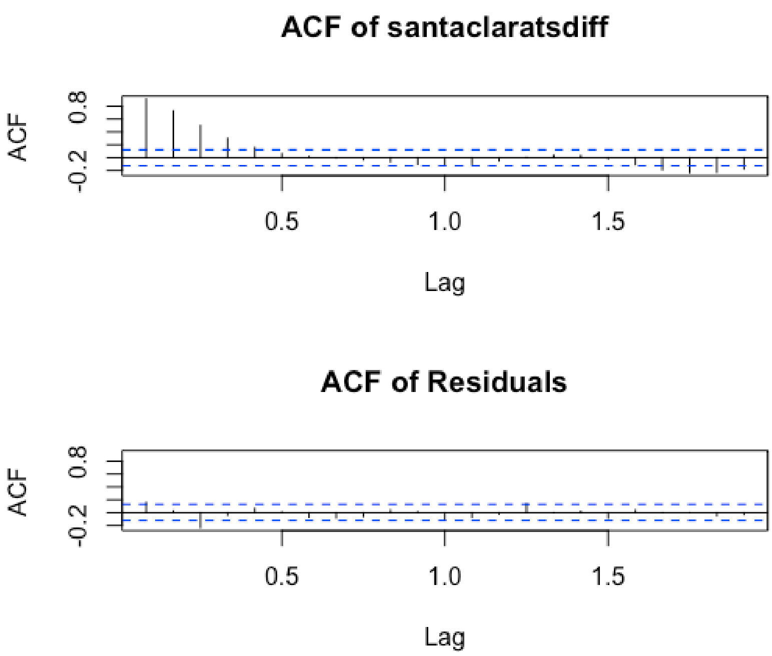

Figure 25: ACF of TAR Model & Residuals

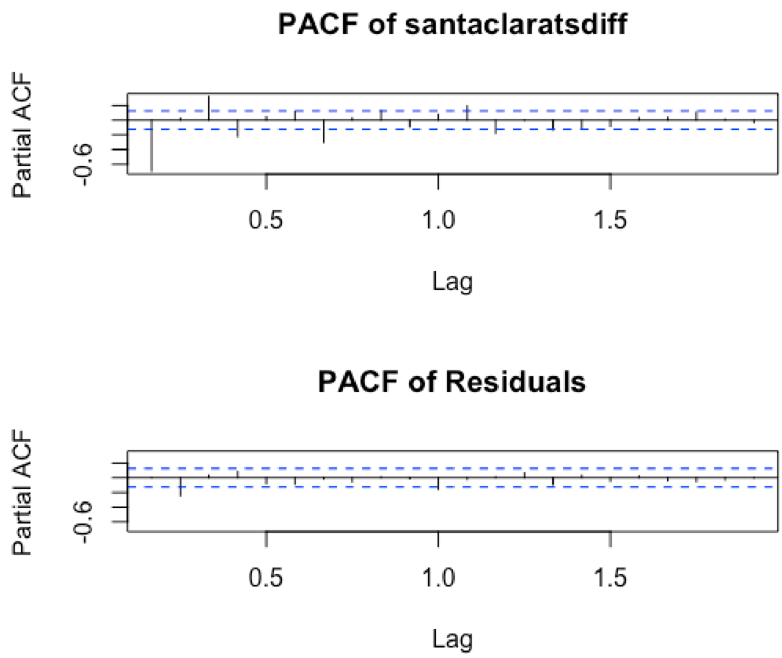

Figure 26: PACF of TAR Model & Residuals

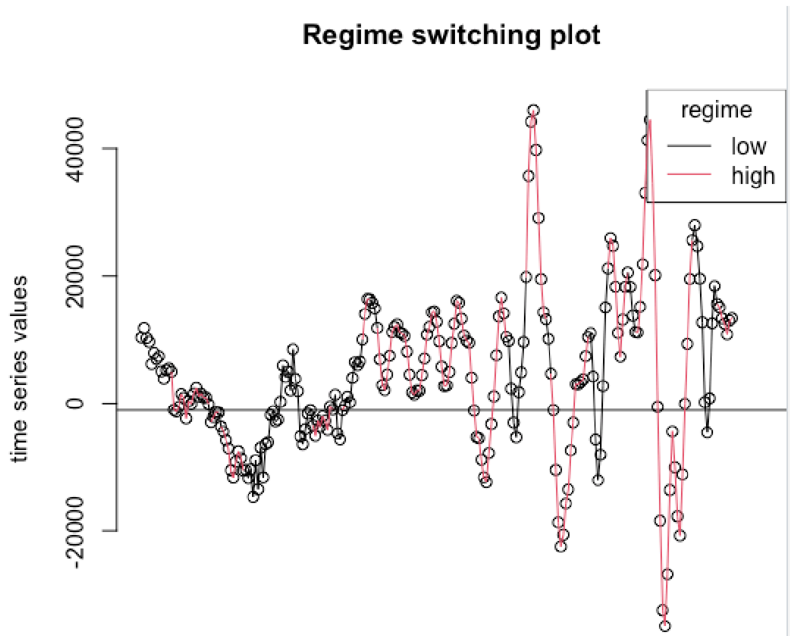

Figure 27: Regime Switching Plot of TAR Model

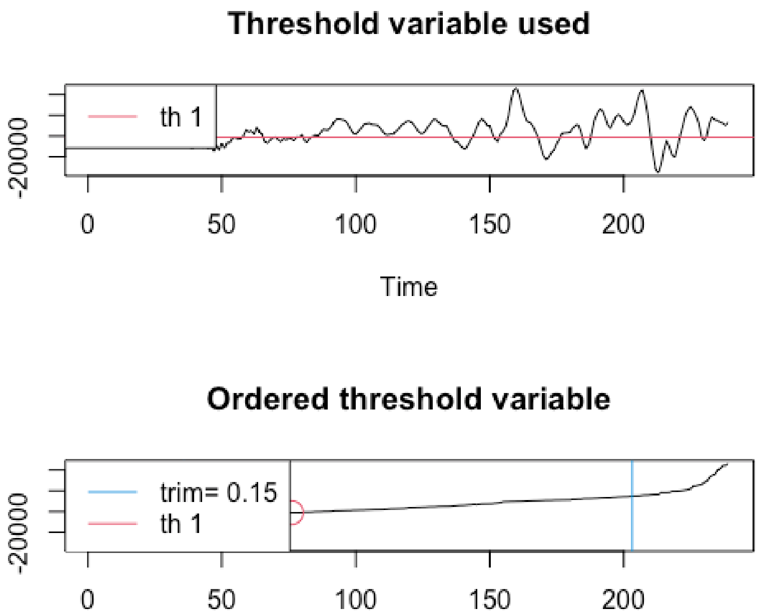

Figure 28: Threshold Variable Plot of TAR Model

```
Non linear autoregressive model

SETAR model ( 2 regimes)
Coefficients:
Low regime:
     const.L       phiL.1
-306.0158369    0.8551282

High regime:
     const.H       phiH.1       phiH.2       phiH.3       phiH.4       phiH.5       phiH.6       phiH.7       phiH.8       phiH.9      phiH.10
441.36449113   1.80481322  -0.73205653  -0.88968759   1.29694639  -0.64074984  -0.44575152   1.01829767  -0.59965940  -0.08375870   0.26761221
     phiH.11      phiH.12
 -0.15695762   0.07896308

Threshold:
-Variable: Z(t) = + (1) X(t)+ (0)X(t-1)+ (0)X(t-2)+ (0)X(t-3)+ (0)X(t-4)+ (0)X(t-5)+ (0)X(t-6)+ (0)X(t-7)+ (0)X(t-8)+ (0)X(t-9)+ (0)X(t-10)+ (0)X(t-11)
-Value: -1043
Proportion of points in low regime: 31.8%       High regime: 68.2%

Residuals:
      Min        1Q    Median        3Q       Max
-16428.14  -1800.72    143.02   1943.36   9917.94

Fit:
residuals variance = 10700665,  AIC = 4095, MAPE = 172.2%

Coefficient(s):

           Estimate  Std. Error  t value  Pr(>|t|)
const.L -306.015837  610.373888  -0.5014 0.6165865
phiL.1     0.855128    0.055405  15.4342 < 2.2e-16 ***
const.H  441.364491  423.608219   1.0419 0.2985165
phiH.1     1.804813    0.098830  18.2619 < 2.2e-16 ***
phiH.2    -0.732057    0.203748  -3.5930 0.0003979 ***
phiH.3    -0.889688    0.221282  -4.0206 7.814e-05 ***
phiH.4     1.296946    0.238761   5.4320 1.385e-07 ***
phiH.5    -0.640750    0.258781  -2.4760 0.0139878 *
phiH.6    -0.445752    0.234760  -1.8988 0.0588161 .
phiH.7     1.018298    0.227193   4.4821 1.153e-05 ***
phiH.8    -0.599659    0.234944  -2.5523 0.0113312 *
phiH.9    -0.083759    0.209252  -0.4003 0.6893154
phiH.10    0.267612    0.198058   1.3512 0.1779308
phiH.11   -0.156958    0.187041  -0.8392 0.4022278
phiH.12    0.078963    0.094534   0.8353 0.4044004
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Threshold
Variable: Z(t) = + (1) X(t) + (0) X(t-1)+ (0) X(t-2)+ (0) X(t-3)+ (0) X(t-4)+ (0) X(t-5)+ (0) X(t-6)+ (0) X(t-7)+ (0) X(t-8)+ (0) X(t-9)+ (0) X(t-10)+ (0) X(t-11)

Value: -1043
```

Figure 29: Summary of TAR Model

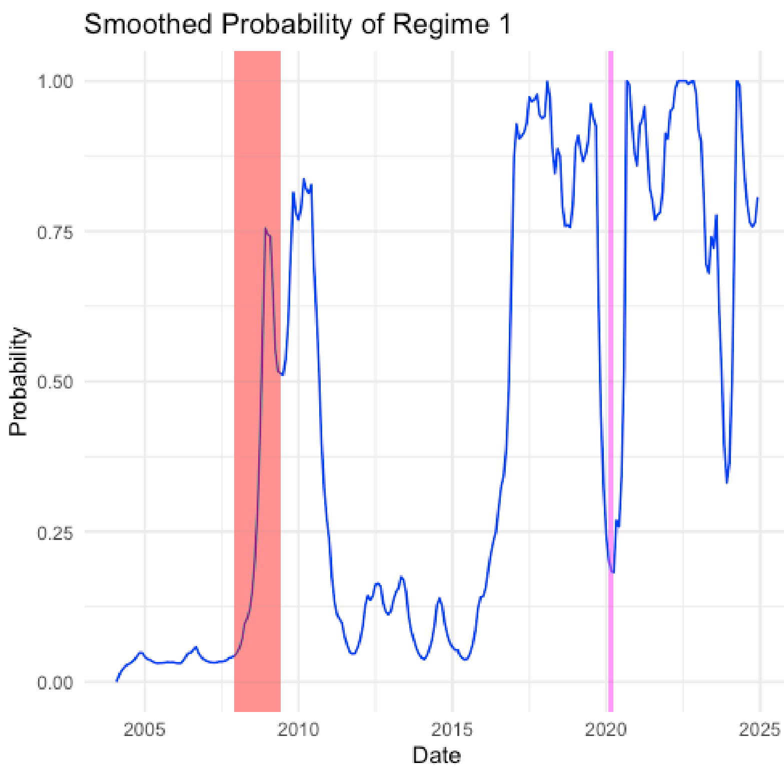

Figure 30: Smoothed Probability of Regime 1
Note: red indicates 2008 Recession, purple indicates COVID-19 Pandemic

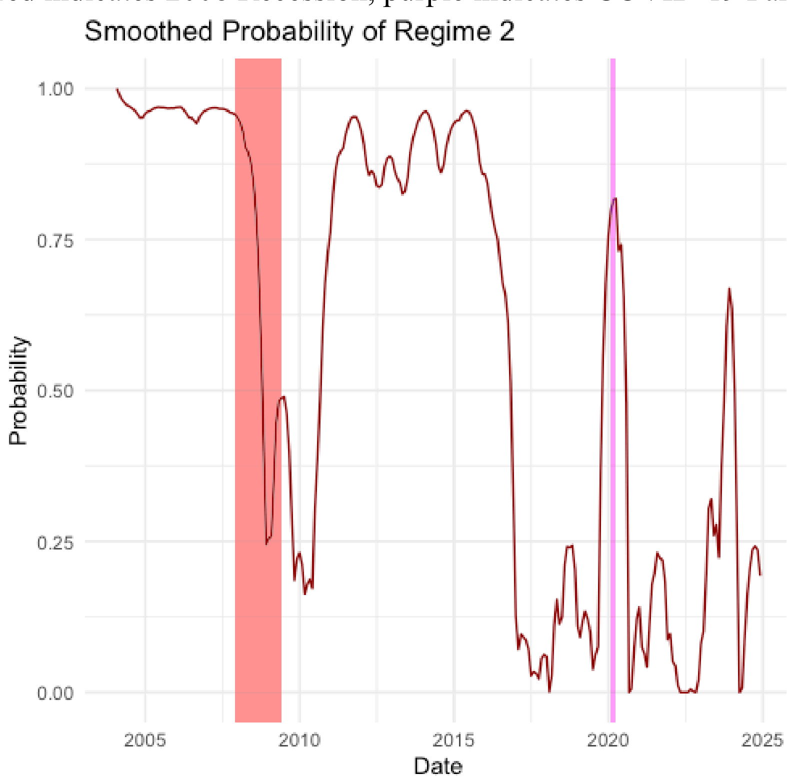

Figure 31: Smoothed Probability of Regime 2

Note: red indicates 2008 Recession, purple indicates COVID-19 Pandemic

	Most_Likely_Regime	
Period	1	2
COVID-19	0	2
Financial Crisis	8	11
Other	99	131

Figure 32: Table of Most Likely Regimes Around 2008 Recession & COVID-19 Pandemic

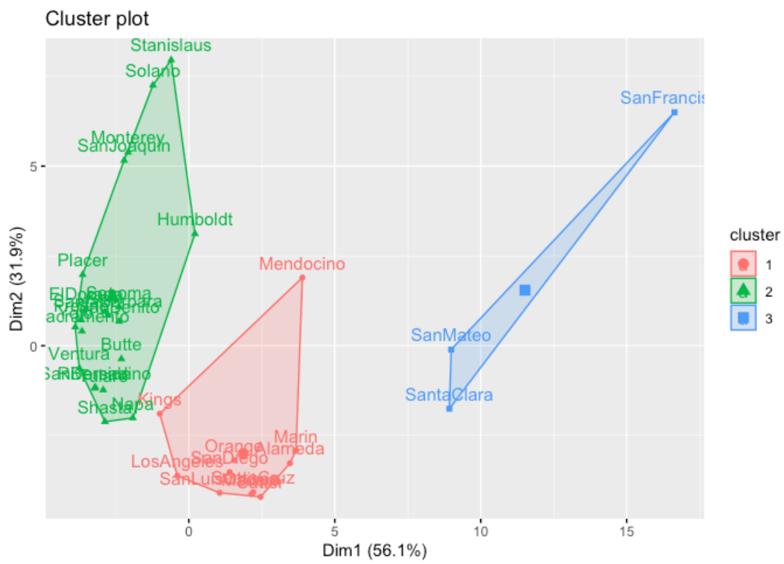

Figure 33: k-Means Clustering of Counties

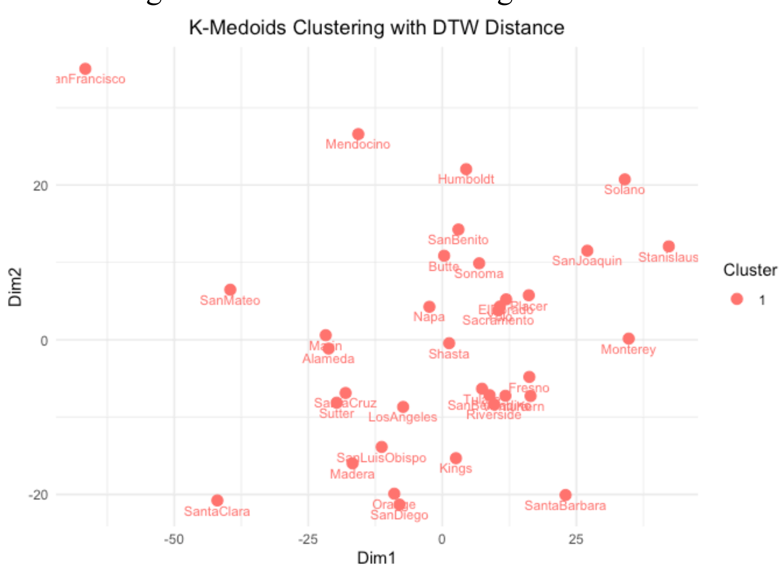

Figure 34: k-Medoids Clustering of Counties

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	1.347e+07	8.173e+05	16.478	< 2e-16	***
bbkgdp	-1.129e+03	6.195e+02	-1.823	0.069568	.
caunemp	-1.835e+04	2.122e+03	-8.644	7.66e-16	***
cci	7.820e+03	6.733e+03	1.161	0.246626	
cpiall	-7.288e+03	1.022e+03	-7.129	1.17e-11	***
cpishelter	3.105e+03	8.319e+02	3.732	0.000237	***
mrates	3.925e+04	8.224e+03	4.773	3.15e-06	***
population	-7.164e+01	4.441e+00	-16.131	< 2e-16	***
date	6.296e+02	4.290e+01	14.677	< 2e-16	***
hcrisis	2.245e+04	1.944e+04	1.155	0.249330	
covid	-3.883e+04	4.341e+04	-0.895	0.371888	

Figure 35: Regression Results from Equation (4.7)

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	2.500e+07	2.413e+06	10.357	< 2e-16	***
bbkgdp	-1.149e+03	6.211e+02	-1.850	0.0655	.
caunemp	-1.871e+04	2.198e+03	-8.513	1.87e-15	***
cci	7.781e+03	6.741e+03	1.154	0.2495	
cpiall	-7.580e+03	1.119e+03	-6.773	9.63e-11	***
cpishelter	3.920e+03	1.511e+03	2.594	0.0101	*
mrates	3.771e+04	8.573e+03	4.399	1.64e-05	***
population	-7.614e+01	8.265e+00	-9.212	< 2e-16	***
poly(date, 2)1	2.263e+07	1.696e+06	13.341	< 2e-16	***
poly(date, 2)2	-2.093e+05	3.238e+05	-0.646	0.5186	
hcrisis	1.659e+04	2.146e+04	0.773	0.4403	
covid	-3.970e+04	4.348e+04	-0.913	0.3621	

Figure 36: Regression Results from Equation (4.8)

References

Balemi, N., Füss, R., & Weigand, A. (2021). COVID-19's impact on real estate markets: review and outlook. *Financial Markets and Portfolio Management*, 35, 495-513. URL: https://doi.org/10.1007/s11408-021-00384-6

Bardhan, A., & Walker, R. (2011). California shrugged: fountainhead of the Great Recession. *Cambridge Journal of Regions, Economy and Society*, 4(3), 303-322. URL: https://academic.oup.com/cjres/article/4/3/303/333498

BLS (2025). All items in U.S. city average, all urban consumers, seasonally adjusted. Retrieved from https://data.bls.gov/dataViewer/view/timeseries/CUSR0000SA0

BLS (2025). Shelter in U.S. city average, all urban consumers, not seasonally adjusted. Retrieved from https://data.bls.gov/dataViewer/view/timeseries/CUUR0000SAH1

BLS (2025). Unemployment Rate: California (S). Retrieved from https://data.bls.gov/dataViewer/view/timeseries/LASST060000000000003

CA.gov (2023). Registration by County [Report of Registration as of October 3, 2023]. Retrieved from https://elections.cdn.sos.ca.gov/ror/154day-presprim-2024/county.pdf

Duca, J. V., Muellbauer, J., & Murphy, A. (2010). Housing markets and the financial crisis of 2007–2009: Lessons for the future. *Journal of Financial Stability*, 6(4), 203-217. URL: https://www.sciencedirect.com/science/article/pii/S1572308910000343

Ferreira, F., & Gyourko, J. (2012). Heterogeneity in Neighborhood-Level Price Growth in the United States, 1993-2009. *American Economic Review: Papers & Proceedings 2012*, 102(3), 134-140. URL: https://pubs.aeaweb.org/doi/pdfplus/10.1257/aer.102.3.134

Freddie Mac, FRED (2025). 30-Year Fixed Rate Mortgage Average in the United States. Retrieved from https://fred.stlouisfed.org/series/MORTGAGE30US

Himmelberg, C., Mayer, C., & Sinai, T. (2005). Assessing High House Prices: Bubbles, Fundamentals and Misperceptions. *Journal of Economic Perspectives*, 19(4), 67-92. URL: https://pubs.aeaweb.org/doi/pdfplus/10.1257/089533005775196769

Indiana University, FRED (2025). Brave-Butters-Kelley Real Gross Domestic Product. Retrieved from https://fred.stlouisfed.org/series/BBKMGDP

Little Hoover Commision (2021). COVID's Impact on California Housing [Issue Brief]. Retrieved from https://lhc.ca.gov/wp-content/uploads/Reports/258/Report258.pdf

Nobari, T. Z., Anderson C. E., & Whaley S. E. (2022). The COVID-19 Pandemic Contributed to Disparities in Housing-Cost Burden Among WIC-Participating Households in the Most Populous County in California. *J Racial Ethn Health Disparities*, 10(1), 100-109. URL: https://pmc.ncbi.nlm.nih.gov/articles/PMC8739018/

OECD (2025). Consumer confidence index. Retrieved from https://www.oecd.org/en/data/indicators/consumer-confidence-index-cci.html

Peterson, A. M. (2024). Shift in house price estimates during COVID-19 reveals effect of crisis on collective speculation. *EPJ Data Science*, 13, 47. URL: https://doi.org/10.1140/epjds/s13688-024-00488-9

Ragan, C. (). The U.S. Housing Collapse and the Financial Crisis of 2007-2008. *McGill University*. URL: https://www.mcgill.ca/economics/files/economics/the_u.s._housing_collapse_and_the_financial_crisis_of_2007-08.pdf

Rohe, W. M., & Lindblad, M. (2013). Reexamining the Social Benefits of Homeownership after the Housing Crisis. *Joint Center for Housing Studies, Harvard University*. URL: https://www.jchs.harvard.edu/sites/default/files/media/imp/hbtl-04.pdf

Shellae Versey, H. (2021). The Impending Eviction Cliff: Housing Insecurity During COVID-19. *American Journal of Public Health*, 111(8), 1423-1427. URL: https://ajph.aphapublications.org/doi/full/10.2105/AJPH.2021.306353?role=tab

U.S. Bureau of Economic Analysis, FRED (2025). Population. Retrieved from https://fred.stlouisfed.org/series/POPTHM

Wang, B. (2021). How Does COVID-19 Affect House Prices? A Cross-City Analysis. *Journal of Risk and Financial Management*, 14(2), 47. URL: https://www.mdpi.com/1911-8074/14/2/47

Zillow (2025). Zillow Home Value Index. Retrieved from https://www.zillow.com/research/data/

Zillow (2025). For-Sale Inventory. Retrieved from https://www.zillow.com/research/data/

Zillow (2025). New Listings. Retrieved from https://www.zillow.com/research/data/

Zillow (2025). Sales Count. Retrieved from https://www.zillow.com/research/data/

Zillow (2025). Days to Pending. Retrieved from https://www.zillow.com/research/data/

Zillow (2025). New Construction Sales Count. Retrieved from https://www.zillow.com/research/data/