Time Series Analysis Project: Zillow Home Value Index

STAT 626, Team 4

Elizabeth Storm (Theory), Allison Moore (Programming), Nicholas Klein (Computation), Jialin Lai (Writing), & Rashmi Sai Burugupali (Data Visualization)

Introduction

Buying, selling, or renting a home is a significant event for both homeowners and potential buyers/renters. Understanding the price fluctuation of the housing market is an important part of the decision-making process. Housing prices can also signal broader economic conditions, so it is of interest to economists and financial institutions. One source of information on the housing market is the Zillow Group, a company that provides a variety of technological platforms that connect renters and buyers with current homeowners and potential lenders. Over time, the Zillow Group has established an index, the Zillow Home Value Index (ZHVI), which records information about the housing market over time.

The goal of this project is to conduct a time series analysis using the ZHVI to forecast future home values with historical home prices. The results of the analysis would provide buyers and sellers with an estimate of the best time to buy or sell a home. In addition, the findings could have potential applications in economic outlooks or within the financial industry for lenders and insurance companies.

Background

Our data consists of the ZHVI for the entire United States as well as individual metro regions. This data is provided by Zillow on a monthly basis using a proprietary calculation that is designed to capture the value of a typical property across the nation or neighborhood, not just the homes that were recently sold. It is a measure of the typical home value, reflecting the typical value for homes in the 35th to 65th percentile range, summarized across all home types including single-family residence, condos and co-ops. This data goes back as far as 1996. Zillow provides the ZHVI as a raw measure or as a smoothed, seasonally adjusted measure, and this project will be focused on the raw measure.

Exploratory Data Analysis

During our initial exploratory data analysis, we found no missing values or anomalies in our data. We uncovered a few insights about our data including (a) a clear increasing trend over time, (b) annual cycles starting in January and going through August, and (c) minimal short-term volatility, as shown in Figure 1. There was initially surprise at the lack of volatility, but we believe this stems from the nature of how Zillow calculates the index value for the month, which is proprietary and involves a trimmed-mean.

The data is available back to 1996 and includes years impacted by the mortgage crisis in 2008. However, we still included the whole span of data rather than avoiding the valuable information before 2009. In our additional analysis, we did explore the impact of what years were included in the model.

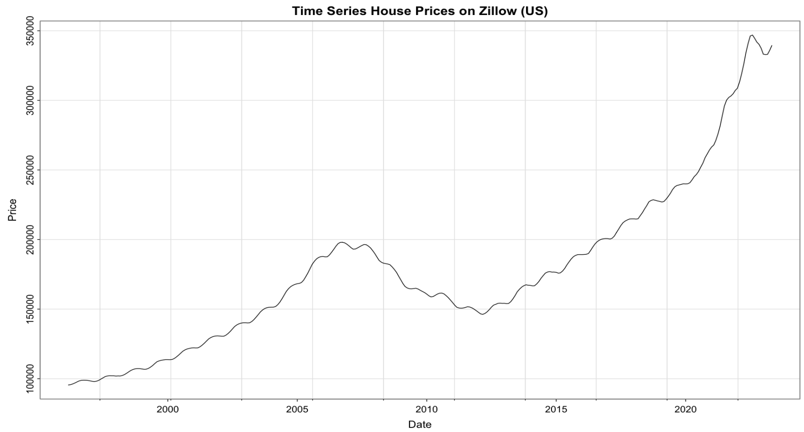

Figure 1 - Plot of ZHVI for the entire United States. There is an overall increasing trend with a noticeable decrease around 2008 due to the financial crisis.

Transform the Data to Stationarity

Further expanding on the trend and cyclicality seen in the exploratory analysis, a visual inspection of Figure 1 suggests that the time series data is not stationary, because it violates the two requirements of stationarity: The mean of the current time series data varies over time, and the variance of the time series plot also showed variation, especially after 2020. Moreover, we saw seasonality within each year, increasing from January to August, and then decreasing until December. Thus, we explored the possibility of stationarizing our current dataset with multiple approaches to detrend and deseasonalize it.

There are several options available for transforming the data to achieve stationarity. We started with transforming the data manually by differencing our data, including (i) first order differencing to reduce trend, (ii) first order seasonal differencing, (iii) monthly average differencing to reduce seasonality, and (iv) a combination of the above. Second, we also considered log transformation and Box-Cox transformation. Third, we fit regression models both for trend and monthly seasonality in an attempt to produce residuals that were stationary. Alternative to the three different manual transformations listed above, we also explored Seasonal Decomposition of Time Series by Loess (STL) in R to decompose each dataset into its Seasonal, Trend, and Remaining components. Figure 2 demonstrates a summary of the methodology we applied to our data in order to find a stationary transformation.

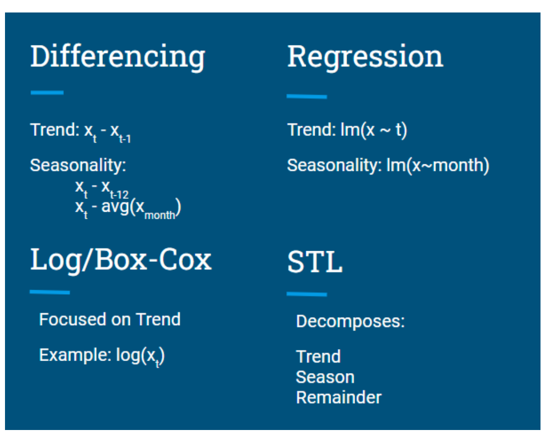

Figure 2 – Methodologies applied to the data to transform to stationarity. Although these methods were used, not all of them were maintained for the final transformation.

We evaluated the effect of transformation with visual inspection of time series plots and unit root test. To test if data shows stationarity or is a random walk model, we used the Augmented Dickey-Fuller (ADF) test and its p-value to determine stationarity. The null hypothesis of the test is non-stationarity. If the p-value is less than .05, we would reject the null hypothesis and conclude that the data is stationary. The ADF test was statistically significant (p=0.01) when data is differenced or stationarized with STL method, showing that either differencing or STL produced stationary data. Ultimately, we decided to go with the STL remainder, because the plot appeared to have less cyclicality remaining, as shown in Figure 3 below. The ADF estimate for this approach was -4.885, with p=0.01.

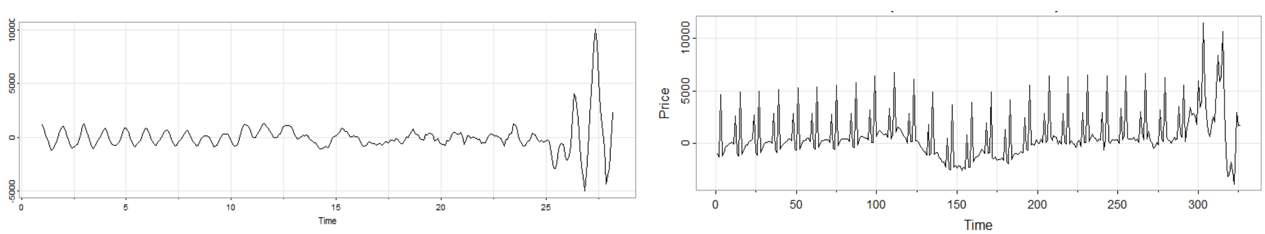

Figure 3- STL function remainder (left) and manually detrended and deseasoned (right) data. There are noticeable sharp peaks in the manual plot, indicating some remaining seasonality.

With the deseasonalized and detrended data from STL method, We also considered applying first order differencing to the remaining stationary component of the results. Figure 4 shows a comparison of Partial Autocorrelation Function (PACF) results from this method against the PACF of the remaining component from STL. We can see that by differencing the remaining component, the data is still autocorrelated, while the PACF values of the remaining component itself cuts off to zero after lag 2. Thus, we decided not to apply another round of differencing. Therefore, while the ADF test was better after differencing our STL remainder, this resulted in additional undesired autocorrelations and over-differencing.

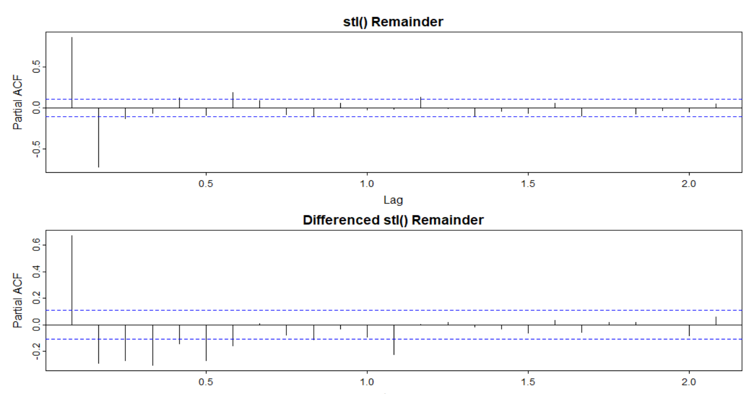

Figure 4- The Partial ACF for the STL remainder (top) and the differenced STL (bottom) data.

Figure 5 provides a visual of how the original data was broken down into trend and seasonal components using the STL function. With all other transformation methods excluded, we moved forward our analysis with the remainder data as our best resulting dataset.

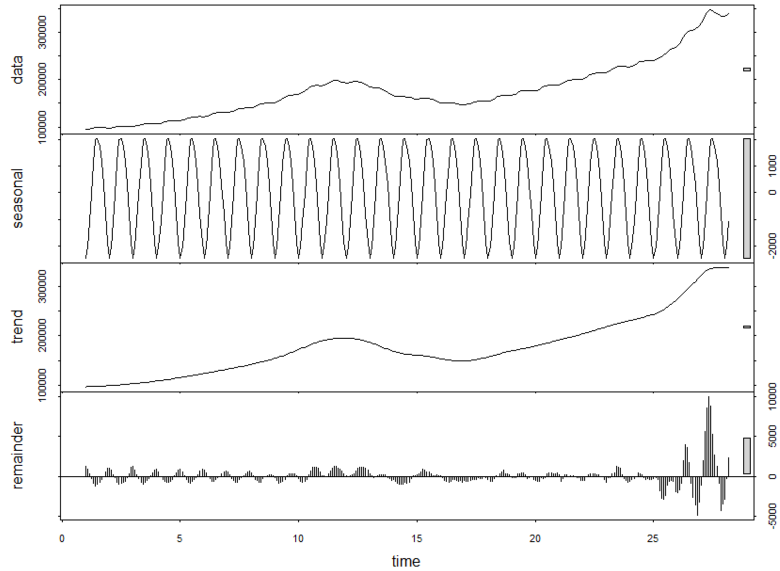

Figure 5- The decomposition of the original data using the STL function in R. The remainder, which was used throughout our project, is shown in the bottom panel of the figure.

Autocovariance and Model Selection

Moving forward with our transformed data, we calculated the ACF and PACF to gain further insight. Figure 6 shows the ACF and PACF correlogram plots from the transformed data. From the ACF plot, we see the ACF values tail off to zero with some cycles. Meanwhile, the PACF plot demonstrates a clear cut off after lag 2, after which the PACF values all fall within the confidence interval of zero. Therefore, we hypothesized that an AR (2) model may fit the current transformed data.

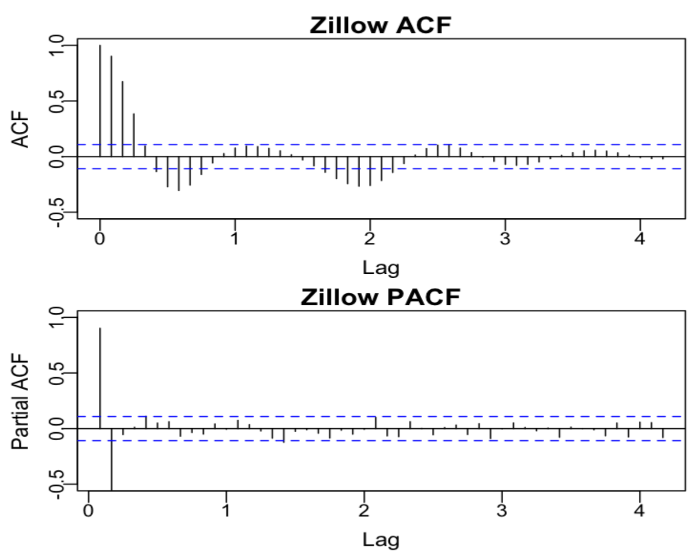

Figure 6- The ACF and Partial ACF of our transformed data. The PACF is indicative of an AR(2) model due to the cut off after lag 2.

Leveraging the lag plots in Figure 7, we see the strong correlation at lags 1 and 2 which supports moving forward with an AR(2) model. The correlation at lag 1 is 0.9 and at lag 2 is 0.67. Additionally, the loess fits within these plots, as demonstrated by the red line, show no nonlinearities.

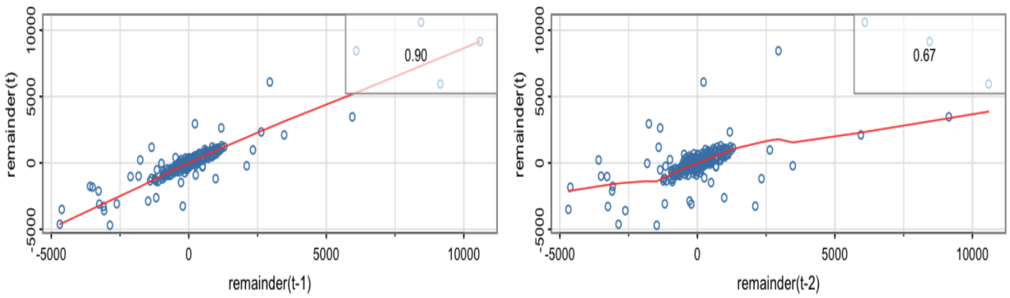

Figure 7- The lag plots associated with our transformed data for lags 1 and 2.These two lags provided the largest correlations.

Model Fitting

Based on the ACF and PACF plots, an AR(2) model was fit to the transformed data. The following formula shows the parameter estimation of the AR(2) model.

$$\hat{x}_t = 1.5801_{(0.0363)} x_{t-1} - 0.752_{(0.0364)} x_{t\_2} + w_t$$

Both autoregressive coefficients were statistically significant. The estimate of the variance of white noise, σ^2, was 146,318, and the AIC is 14.766. The AIC was more relevant in fit comparison in our further analysis.

.

We were aware that for an AR model, we needed to check the roots of the equation to examine the causality, which indicates that the values of our models would not depend on the future values, but just on present and previous values. Solving the roots of our AR model results in two complex roots: Z =1.0506 ± 0.4754i. With the complex roots, we were able to calculate the modulus, which was 1.15, showing that the roots were outside the unit circle of 1. Thus, we concluded that the model was causal.

Model Diagnostics

Reviewing the model diagnostic plots in Figure 8, the standardized residuals nearly resembled white noise; however, there appeared to be heteroscedasticity in the residuals. The residual ACF plot indicated the residuals were uncorrelated since they were mostly within the CI bounds. The QQ plot indicated that the residuals were mostly normally distributed, except for the tails of the plot. The Ljung-Box plot indicated the residuals were not statistically significantly different from white noise, given that for most of the lag values plotted, the p-value was large. Thus, the current AR(2) model showed an adequate fit to our stationary data.

In response to the residual heteroscedasticity, we also explored GARCH to strive for a potential better fit. Details are given in the section of “Additional Analyses” below. However, we still decided on the current AR(2) model for forecasting, due to the simplicity of the model. Other analytic methods, as shown in the Additional Analyses section, were also explored but we favored AR(2) over all other models.

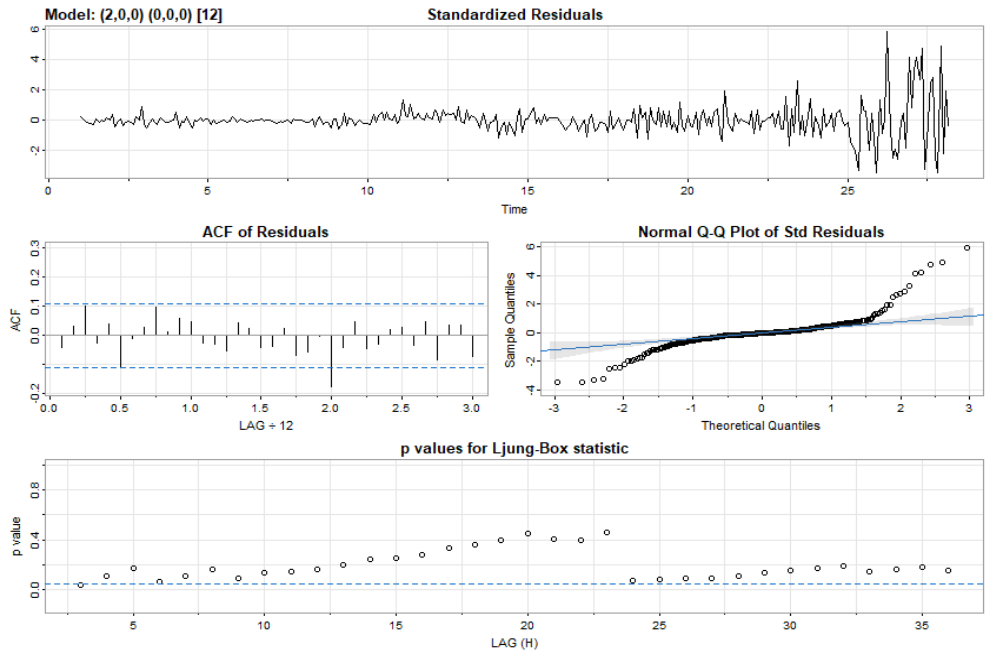

Figure 8- Diagnostics from the initial AR(2) fit. All diagnostics indicate our residuals are not statistically different from white noise, indicating a good fit.

Forecasting

Forecasting is a key goal of our time series analysis, so that we can understand the expected future behavior of housing prices and provide indications for decision-making. Forecasting our model in R resulted in the plot shown in Figure 9 below, showing a forecast for the next 24 months.

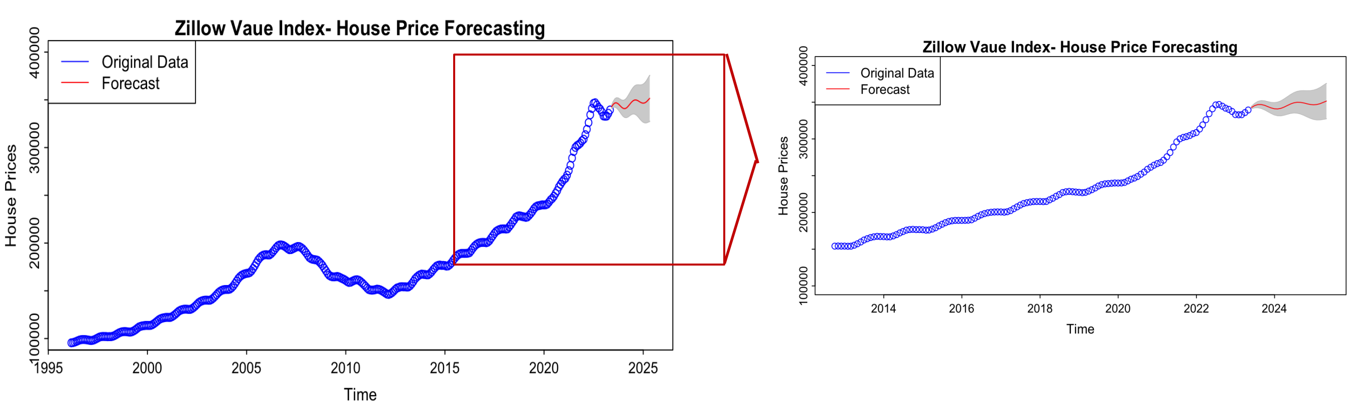

Figure 9- AR(2) fit and forecast (left) and a zoomed in depiction of the forecast (right). As expected, the confidence interval increases over time.

Since we started our modeling initially using data from the STL decomposition, we took the appropriate steps to add seasonality and trend back to the forecasting mode. As Figure 9 shows, the resulting forecast appeared to follow a similar behavior to the prior data, continuing the upward trend with seasonal cycles. The confidence interval was fairly small, but gradually increased over time, which is expected. The right-hand side of Figure 9 shows a zoomed in view of 2014 onward.

Additional Analyses

As we continued to conduct our time series analysis on the Zillow Home Value index, we wanted to explore other modeling techniques. First, we re-examined the plot of standardized residuals in Figure 8, which demonstrates increasing volatility in the data. Thus, the first additional model we applied to our data was ARCH/GARCH. Figure 10 shows the ACF and PACF plots for the squared residuals, which showed signs of autocorrelation. The ACF plot clearly tailed off and the PACF appeared to tail off as well although the pattern was less clear. We moved forward with fitting an AR(2) - GARCH(1,1) model. The resulting diagnostic plots showed a reasonable fit. The AIC and BIC statistics were also a little lower than the AR(2) model. The Ljung-Box statistics also point to residuals not being statistically different from white noise.

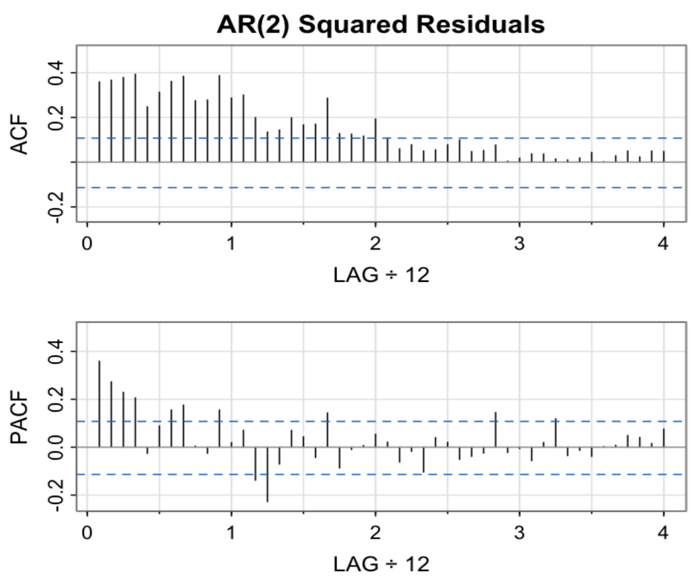

Figure 10- The ACF and Partial ACF of the transformed data residuals, squared. There is substantial correlation within the squared residuals indicating an ARCH/GARCH may be appropriate for our data.

There was evidence of improvement using an AR(2) - GARCH(1,1) model; however, we still decided on the AR(2) model over the AR(2)-GARCH (1,1) model because of the additional complexity from the GARCH model. The additional complexity could be hard to be communicated to a client/stakeholder.

Next, we explored Facebook's Prophet package to fit our data. We started our analysis using the STL remainder data that we had transformed to stationarity. The Prophet model performed poorly on our stationary data; it didn't adequately capture the fluctuations around the financial crisis nor did it adapt to the heteroscedasticity exhibited towards 2023. However, using Prophet on our raw data yielded an excellent fit, which was nearly perfect before, during, and immediately after the financial crisis until it began to diverge from the actual values around 2020, as shown in Figure 11. According to our research, Prophet works well with time series data that has strong seasonal effects, especially those where the trend would benefit from being modeled with a piecewise function (Taylor and Letham 2017).

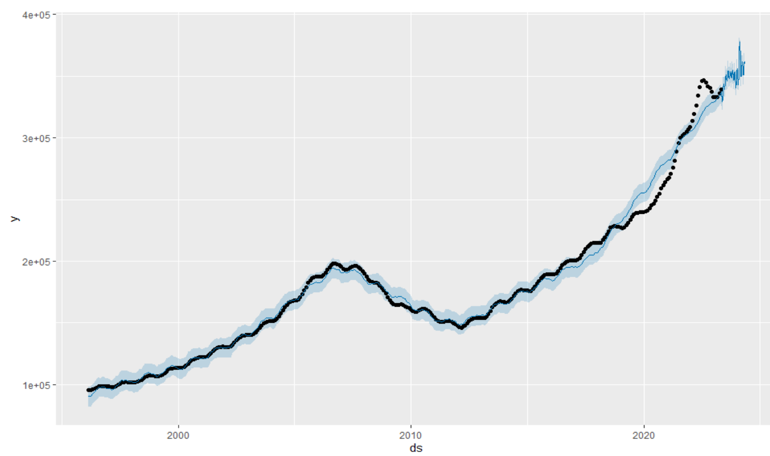

Figure 11- The fit and forecast of the untransformed ZHVI data using Facebook's Prophet package.

Although the fit appeared to be fairly good for our raw data, the residuals did not demonstrate white noise behavior. There was substantial correlation present in both the residual ACF and the residual PACF. The divergence from normality in the tails in the residual QQ Plot were also concerning and the Ljung-Box statistics were statistically significant for most lag values.

Lastly, we considered a piecewise model. As initially laid out in our data exploration, we had to consider throughout this project how the subprime mortgage crisis, starting in 2008, disrupted our modeling and ability to forecast the future. When looking at a piecewise model, there were three clear breakpoints (i.e., 1996-2009, 2009-2012, 2012-2020) we identified where the trend swung from positive to negative and then back to positive. For the piecewise model, we only focused on fitting models to the period from 1996 to 2009 (older period) and 2012 to 2022 (recent period), and discarded the period from 2009 to 2012, which was less of an interest to us to forecast the future values.

Figure 12 shows the forecasted values of the next 24 months for period 1996-2009 and 2012-2020. Forecasted results for both periods resulted in a similar model fit, of a SARIMA(1,1,0)(0,1,1) with s=12. In particular, the autoregressive coefficients for the non-seasonal AR(1) were similar for both periods. Both models resulted in forecasts that fit well to our actual experience, as shown in Figure 12 below. These models probably could be further explored, but this satisfied the needs of our project.

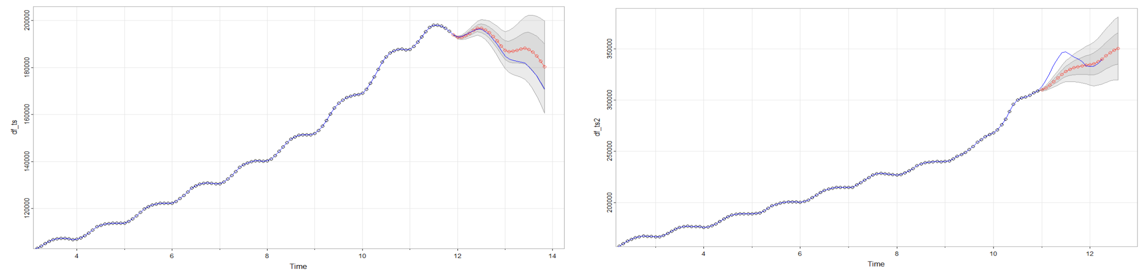

Figure 12- The fit and forecast using a manual piecewise function for 1996-2009 (left) and 2012-2020 (right).

Ultimately, this exercise in piecewise modeling brings to light that consideration should be given to what years are included in the model given past shocks such as the subprime mortgage crisis. As we recognized during our exploratory data analysis, this is an important part of time series analysis.

Conclusion

One of the biggest challenges during this research was the transformation of our data to stationarity. There were several combinations of transformations that resulted in stationarity, and different orders of operations that resulted in different ADF values. We used those ADF values in combination with the resulting correlograms to determine where to stop and if additional dependencies had been created by our transformation process. We speculated that the reason why STL was the optimal option for transformation is that the trend is fit with a non-parametric loess model that fully captures the different trends across time periods.

In conclusion, we recommend the simple solution of AR(2) for the ZHVI data. Although the GARCH model performed slightly better, it is more complex and, thus, less interpretable when explaining the model to a client. We believe that maintaining some parsimony in modeling these types of data is essential for retaining client engagement and for providing understandable interpretations of the model.

Bibliography

Our Services. Zillow Group. https://www.zillowgroup.com/about-us/our-services/
Zillow. Housing Data - Zillow Research. Zillow Research. Published 2011. https://www.zillow.com/research/data/

Olsen S. Zillow Home Value Index Methodology, 2023 Revision: What's Changed? Zillow. Published February 11, 2023. Accessed June 11, 2023. https://www.zillow.com/research/methodology-neural-zhvi-32128/

Taylor SJ, Letham B. Forecasting at Scale, 2017, *PeerJ Preprints* 5:e3190v2. Accessed July 28, 2023.https://doi.org/10.7287/peerj.preprints.3190v2.