

Proposal for the time series analysis

The goal of the following time series analysis is to explore the bubble effect which occurred in 2000 and forecast the future trend of house prices from the given the S&P/Case-Shiller Home Price Index. The dataset includes 393 proportions of housing prices compared to the price in January 2000 from January 1, 1986 to September 1, 2019. It has 0.6375 for a minimum proportion and 2.112 for a maximum value. If the housing price in January 2000 was \$500,000, the maximum proportion of 2.112 in September 2019 implies that the price was increased 212.203 percent which were $\$500,000 \times 2.122 = \$1,601,015$. There was a housing bubble in 2000 that the housing prices increased dramatically and dropped extremely in the second phase. Therefore, the data were not stationary, and they required differentiating.

I will use 2 different ways to find the optimal time-series model to predict the future housing price. The first method is using ACF and PACF values to find the best ARIMA model and the other method is using python library 'statsmodel' that automatically trains and tests the model and predicts the future values. I will use R for the first method and python for the other one. Afterward a QQplot and residual plot will help to evaluate the model.