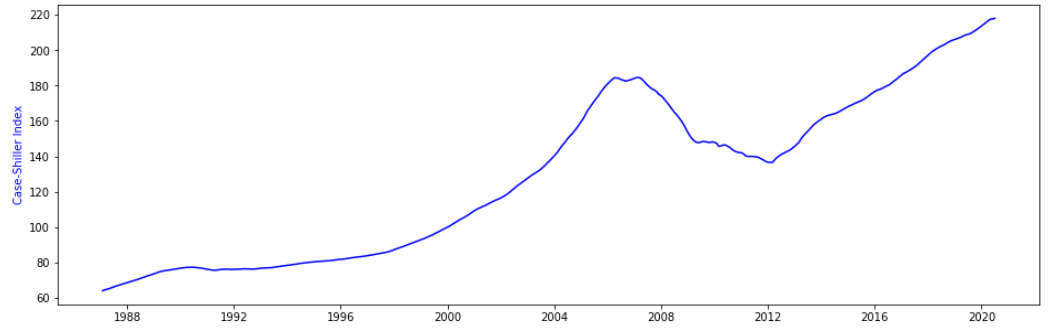
**The 1 st Submission for 601-Econ**

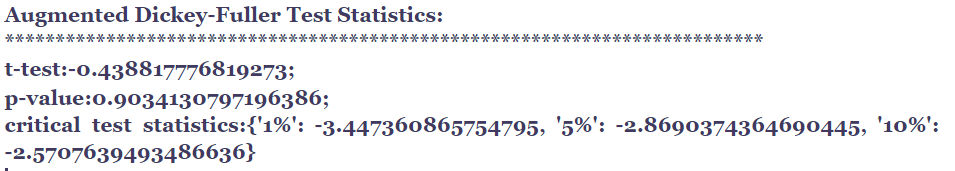
**( Xing Tao, Tanguy de La Sablieret, Abishek Puri, Ningxi Wang, Bian Wu )**

**Univariate Time Series**

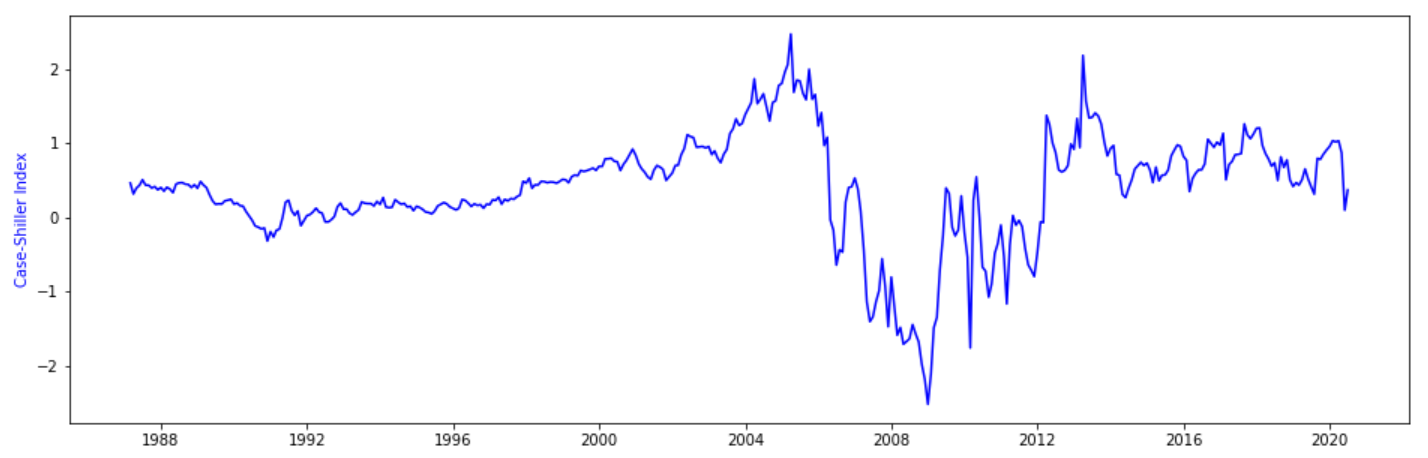
1. **Download the data, run ADF Test and do diff to make the series stationary.**

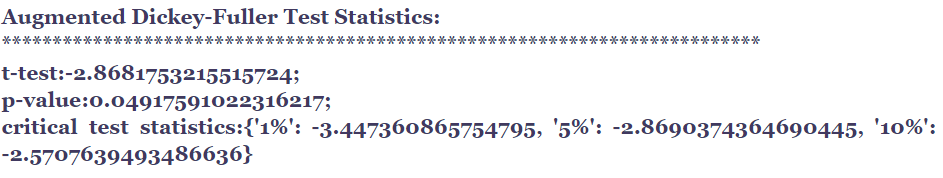
Download the data and plot the index.

****

****

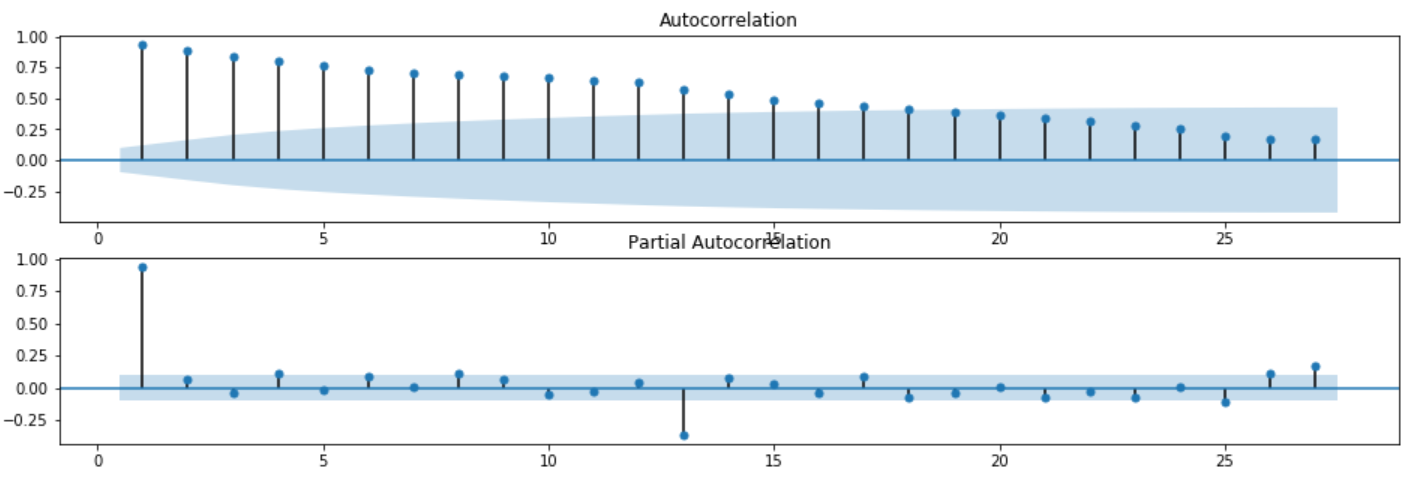
P-value > 5%, fail to reject the null hypothesis (H0), the data has a unit root and is non-stationary. So diff or log is necessary to make the series stationary. And the diffed series with lag 1 is stationary with p-value =0.049

****

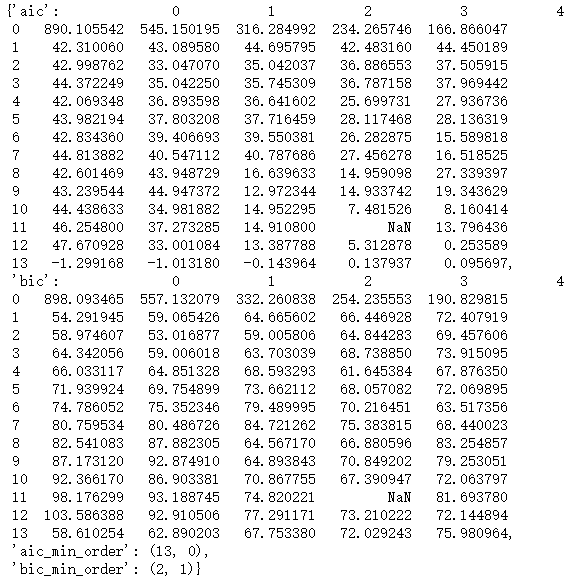
****

1. **Decide the value of p and q.**

Draw the ACF and PACF plot.

****

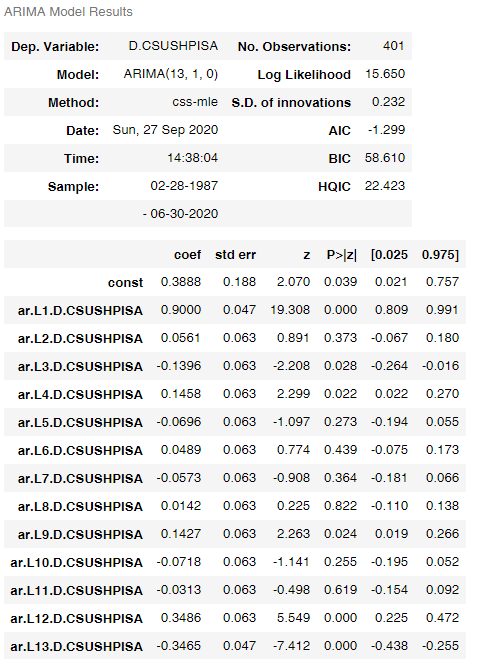
ACF shows a gradual decay, while PACF has a spark at lag 1, 13, 25, etc. It’s not reliable to predict autoregressive lag or moving average lag. And we use a grid searching method to find the best p and q.

****

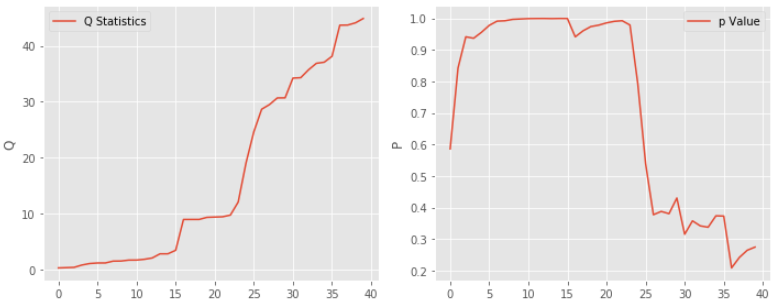
The grid searching result shows that AIC and BIC do not agree with each other. So we have two optional models ARIMA(13,1,0) and ARIMA(2,1,1).

1. **Fit the data with the two models and choose a better model.**

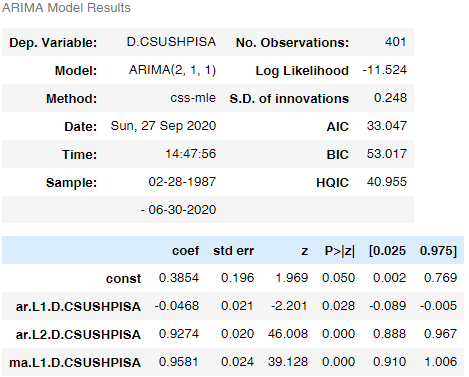
**3.1 Fit the data using ARIMA(13,1,0) model.**

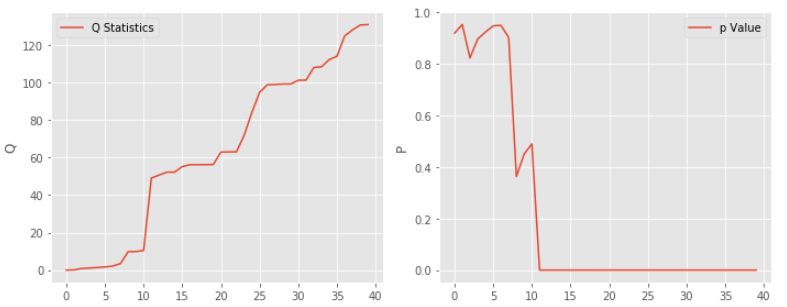


Ljung-Box Q-test can more formally assess autocorrelation. We run the Ljung-Box Q-test to show whether the residual of ARIMA(13,1,0) behaves like white noise. The Q Statistics and P Value show that we can not reject the hypothesis that the series is not white noise. So the residual behaves likewhite noise we do not need to further analyse the residual using volatility models including ARCH and GARCH.



**3.2 Fit the data using ARIMA(2,1,1)model and analyse the residual.**

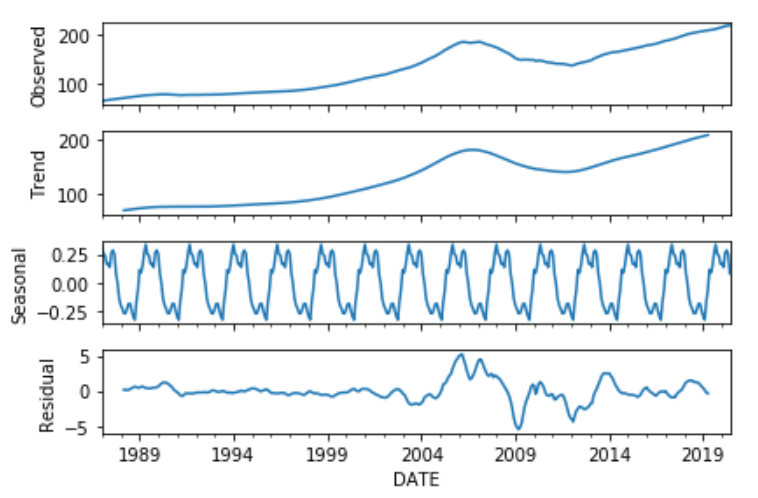




Ljung-Box Q-test shows that the p value drops to around zero after lag 12. The residual series shows autocorrelation with a lag of 12. ARIMA(2,1,1) is not suitable to fit the data and make predictions, and we need to consider the **SARIMAX** model.

**3.3 Fit the SARIMAX** **model**

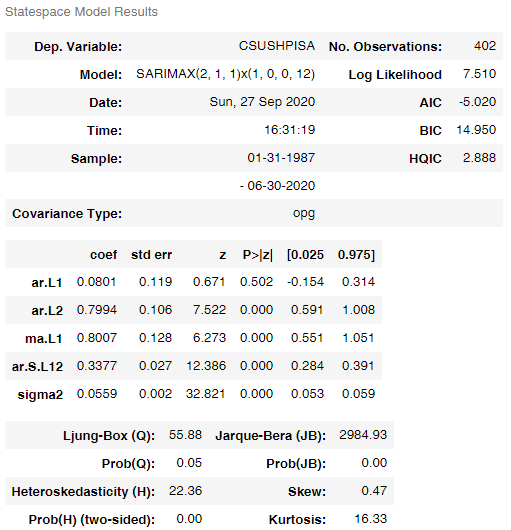
Decompose the original data to three parts: Trend, Seasonal and Residual.



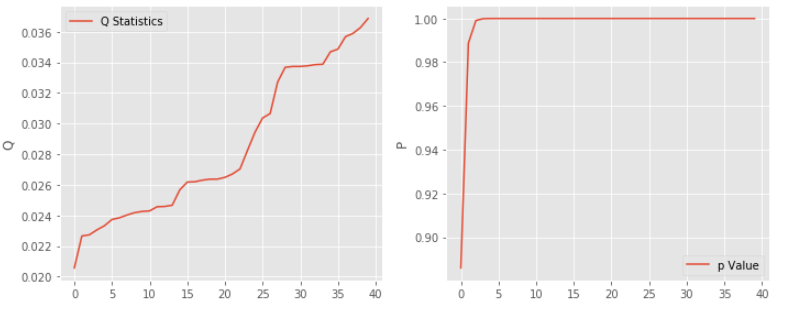
We run ADF Test and the result shows that the seasonal series is stationary and then we use a grid search to find the suggested parameters for seasonal\_order(1, 0, 0, 12) in **SARIMAX** .



The parameters for **SARIMAX** are order=(2,1,1), seasonal\_order=(1, 0, 0, 12).

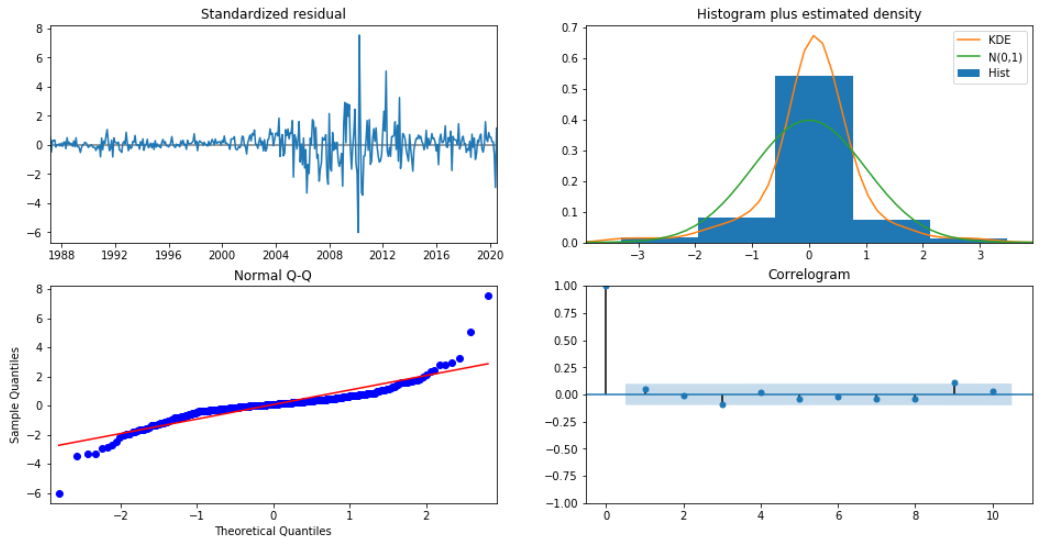


The residual of **SARIMAX** behaves like white noise, so we do not need to analyse the residual using volatility models including ARCH and GARCH.



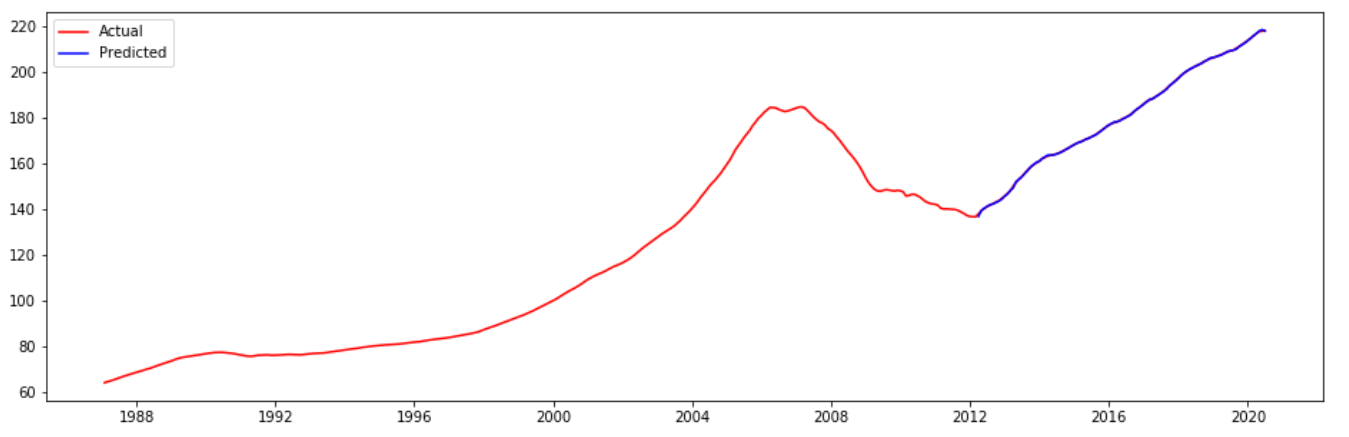
**3.4 Conclusion**

Although both ARIMA(13,1,0) and SARIMA(order=(2,1,1), seasonal\_order=(1, 0, 0, 12)) can fit the data and with white-noise-like residual, we believe that SARIMAX (order=(2,1,1), seasonal\_order=(1, 0, 0, 12)) is a better choice due to less parameters and better performance in AIC and BIC.



1. **Forecast**

Compare the predicted data with the actual data during the last 100 month.

****

**Reference:**

Nazemi, Nader, Introduction to Time Series Forecasting With Python(v 1.9), 2020.

Tsay,Ruey S., Analysis of Financial Time Series(Third Edition),The University of Chicago,2010